

## Improved Skill of Northern Hemisphere Winter Surface Temperature Predictions Based on Land–Atmosphere Fall Anomalies

JUDAH COHEN

*Atmospheric and Environmental Research, Inc., Lexington, Massachusetts*

CHRISTOPHER FLETCHER

*Department of Physics, University of Toronto, Toronto, Ontario, Canada*

(Manuscript received 20 June 2006, in final form 6 December 2006)

### ABSTRACT

A statistical forecast model, referred to as the snow-cast (sCast) model, has been developed using observed October mean snow cover and sea level pressure anomalies to predict upcoming winter land surface temperatures for the extratropical Northern Hemisphere. In operational forecasts since 1999, snow cover has been used for seven winters, and sea level pressure anomalies for three winters. Presented are skill scores for these seven real-time forecasts and also for 33 winter hindcasts (1972/73–2004/05). The model demonstrates positive skill over much of the eastern United States and northern Eurasia—regions that have eluded skillful predictions among the existing major seasonal forecast centers. Comparison with three leading dynamical forecast systems shows that the statistical model produces superior skill for the same regions. Despite the increasing complexity of the dynamical models, they continue to derive their forecast skill predominantly from tropical atmosphere–ocean coupling, in particular from ENSO. Therefore, in the Northern Hemisphere extratropics, away from the influence of ENSO, the sCast model is expected to outperform the dynamical models into the foreseeable future.

### 1. Introduction

It is estimated that about one-third, or 3–4 trillion dollars (NOAA 2002; Dutton 2002), of the U.S. 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 (ENSO)] with surface temperatures and precipitation patterns across the globe. However, predictive skill for temperature forecasts outside of the Tropics, including the United States, 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 winter-time climate variability, in particular in the extratropics, where the dominance of ENSO is more tenuous. Better understanding of the dominant mode of Northern Hemisphere (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 United States and Europe, regions where temperature 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 winter NAO/AO has been linked with prior sea surface temperature (SST) variability, sea ice variability, stratospheric forcing, subpolar air temperature, and aerosols (Rodwell et al. 1999; Mysak and Venegas 1998; Bald-

---

*Corresponding author address:* Judah Cohen, Atmospheric and Environmental Research, Inc., 131 Hartwell Avenue, Lexington, MA 02421.

E-mail: jcohen@aer.com

win and Dunkerton 1999; Fletcher and Saunders 2006; Perlwitz and Graf 1995). Linking the NAO/AO to slowly varying boundary conditions could therefore provide predictability; however, recent articles on the subject have emphasized the lack of understanding of the underlying dynamics driving NAO/AO variability and consequently its poor predictability (Hurrell et al. 2001).

During the period in which it has been extensively monitored, snow cover has exhibited similar trends to the NH climate cycles, peaking in the late 1970s and collapsing to record minimum values in the late 1980s and early 1990s, followed most recently by an increasing trend. However over the entire satellite period no significant trend in snow cover was found; this is in contrast to NH land surface temperatures that have been in a monotonic upward trend over the same period (Cohen and Barlow 2005). 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), and 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 is strongest.

In the remainder of the article we will present a simple statistical model, which has been developed making use of observed Eurasian snow cover and sea level pressure (SLP) anomalies for winter climate prediction of extratropical NH surface temperatures. We will demonstrate that the skill of this model, tested both in real time and in hindcasts, consistently outperforms winter forecasts from the major governmental forecast centers. This paper aims 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. Data and methods

### a. Data

All atmospheric data are taken from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis

(Kalnay et al. 1996) for the years 1948 through 2006. For snow cover data we used the weekly and monthly datasets, produced by the National Oceanic and Atmospheric Administration (NOAA), which cover the period from 1972 through 2005 (Robinson et al. 1993). These large-scale observations of the spatial extent of NH continental snow cover are primarily based on visible-band satellite imagery; reliable satellite-derived estimates of snow cover extent have only been continuously available since 1972.

We henceforth restrict our focus to the winter AO mode of variability, rather than the NAO, because our study concerns seasonal prediction for the entire NH and not only the North Atlantic sector. Our version of the AO index is derived from NCEP–NCAR reanalysis data and is the first principal component of gridded SLP poleward of 20°N for the winters 1972/73–2004/05, as defined by Thompson and Wallace (1998). The surface temperature trend is computed as the linear trend for the individual grid points for the winters 1972/73–2004/05 using a least squares regression fit.

Ensemble-mean seasonal hindcast data from general circulation model (GCM) simulations are analyzed for comparison. We assess the hindcast skill of the following three GCMs or ensemble of GCMs: 1) the Climate Prediction Center (CPC) Climate Forecast System (CFS) retrospective hindcast project for the period 1983–2004 (Saha et al. 2006), 2) the Canadian Historical Forecasting Project (HFP) for the period 1972–93 (Derome et al. 2001), and 3) four of the seven GCMs from the European Centre for Medium-Range Weather Forecasts (ECMWF) Development of a European Multimodel Ensemble System for Seasonal to Interannual Prediction (DEMETER) project from which data were available for the period 1972–2001 (Palmer et al. 2004). The original DEMETER seasonal hindcast system employed a correction to remove systematic biases from the output of each GCM prior to computing the multimodel ensemble climatology (e.g., Palmer et al. 2000). However, in this study we did not perform this correction, which could have the effect of reducing the apparent hindcast skill of the DEMETER system.

### b. Derivation of SLP/snow index

The ENSO phenomenon is the universal lynchpin of seasonal forecasts (Barnston et al. 1994; van Oldenborgh et al. 2005a; Saha et al. 2006). The modern age of seasonal forecasting is considered to have been born in the winter of 1997/98 when the U.S. government successfully forecasted temperatures and precipitation across the United States. However repeat success has remained elusive. Plotted in Fig. 1 is the correlation of the Niño-3.4 index and NH extratropical surface tem-

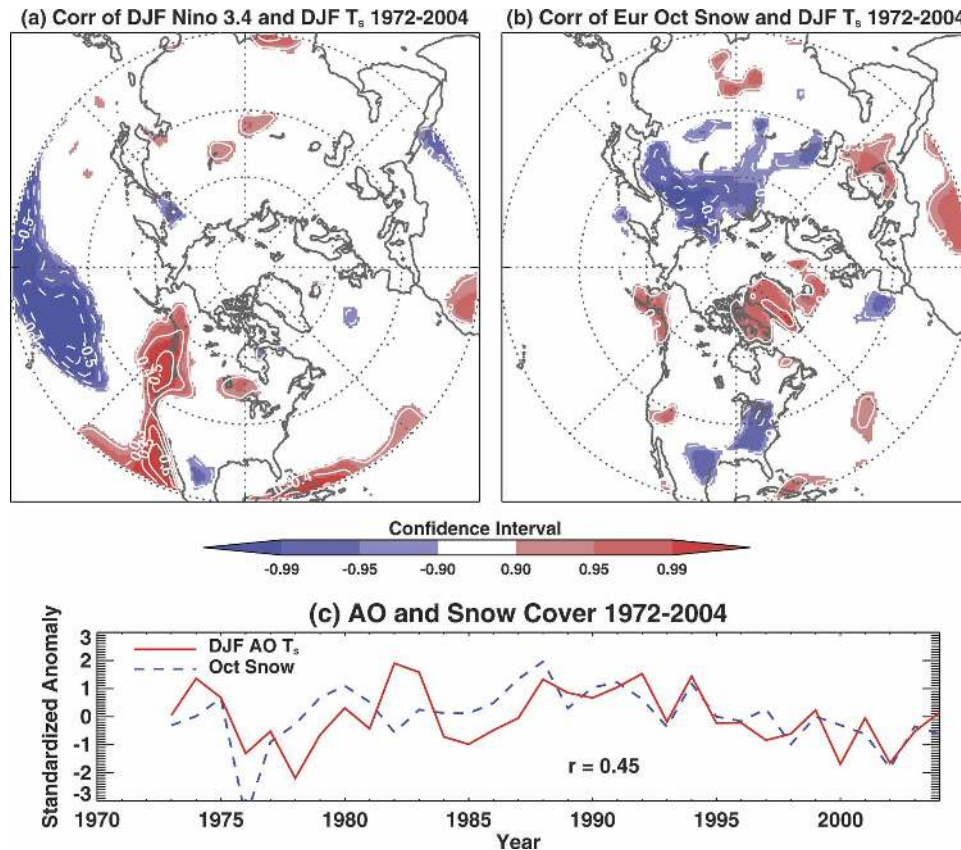


FIG. 1. (a) Correlation of DJF Niño-3.4 index and DJF NH surface temperatures. (b) Correlation of October Eurasian snow cover and DJF NH surface temperatures. Contour intervals are  $\pm 0.30, 0.40, 0.50, 0.60,$  and  $0.70$ , and the light, dark, and darkest color shading represents correlations that exceed 90%, 95%, and 99% confidence intervals, respectively, based on the Student's  $t$  test. (c) Time series of the first EOF of DJF NH surface temperatures (poleward of  $20^{\circ}\text{N}$ ) and October Eurasian snow cover for 1972–2005; also included is the correlation value ( $r$ ) between the two time series.

peratures. Little of the NH landmasses are highly correlated with ENSO with the exception of the immediate west coast of North America and the Canadian prairies. Based on this figure, accurate temperature forecasts derived from ENSO would be the exception rather than the rule given the scarcity of significant correlations. However the figure only reflects the linear relationship between ENSO and surface temperatures. It is plausible that nonlinear linkages between ENSO and remote temperatures exist that could improve forecast skill beyond what is evident from Fig. 1, such as in a dynamical model.

The fact that ENSO offers only limited predictability for the extratropics motivates the need to find other, more reliable, predictors to complement the use of ENSO for climate prediction outside of the Tropics. We now outline the development of a statistical forecast model for the extratropics, based on Eurasian snow cover, atmospheric precursors, and the AO index. We

performed an empirical orthogonal function (EOF) analysis on observed surface temperatures ( $T_s$ ) from the NCEP–NCAR reanalysis data for years 1972–2004; the dominant mode of variability for December, January, and February (DJF) accounts for 18% of the total variance (not shown; pattern closely resembles Fig. 2a). The first mode is often referred to as a quadrupole 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 United States and two same-signed anomaly centers across the Mediterranean and North Africa and northeastern Canada and Greenland. So, for example, when the AO is negative, anomalous high pressure over the continents advects a cold flow of air over northern Eurasia and the eastern United States, while North Africa, the Mediterranean,

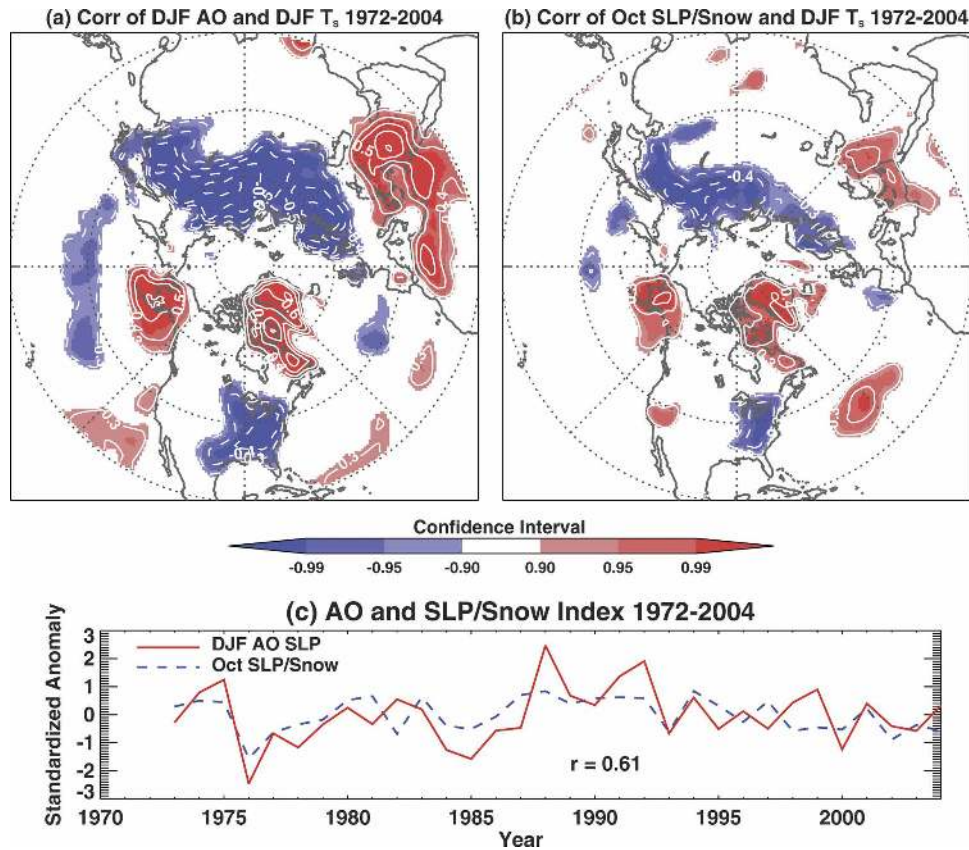


FIG. 2. (a) Correlation of DJF AO (based on SLP poleward of  $20^{\circ}\text{N}$ ) index and DJF NH surface temperatures. (b) Correlation of October SLP/snow index and DJF NH surface temperatures. Contour and shading are the same as in Fig. 1. (c) Time series of the first EOF of DJF NH SLP (poleward of  $20^{\circ}\text{N}$ ) and October SLP/snow index for 1972–2005; also included is the correlation value between the two time series.

northeastern Canada, and Greenland are warmed by an anomalous southerly flow of air. Comparison of correlation maps between the winter AO and winter Niño-3.4 indices and 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. Therefore, the motivation is to construct a predictive index that is highly correlated with the observed winter AO in order to exploit the demonstrated coupled variability between the AO index and NH winter  $T_s$  to achieve improved winter climate prediction.

From the weekly snow cover 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 October Eurasian snow cover with the time series of the first EOF of  $T_s$  (Fig. 1c). The two time series are correlated at a value of

0.45, a value statistically significant at the 99% confidence interval even though it explains less than a quarter of the variance. Also shown in Fig. 1 is the correlation of the Eurasian October snow cover time series with the gridpoint 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 AO. Present is the quadrupole pattern, albeit weaker, with a one-signed anomaly in the eastern United States and across northern Asia and an 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. Nonetheless, using snow cover alone still provides greater skill than ENSO indices for surface temperatures in the extratropical NH. However, the skill from using snow cover alone is still probably insufficient for reliable climate forecasts.



Cohen et al. (2001) were the first to attempt to increase the correlation value of fall Eurasian snow cover with the winter AO by combining it with observed simultaneous SLP. They derived a time series from observed October snow cover and SLP anomalies from a fixed grid point 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.

A single index could then be derived that linearly combined both the observed dominant SLP anomaly in October and the observed October Eurasian snow cover extent anomaly that was highly correlated with the winter AO index. This index is referred to as the SLP/snow index. The weighting for the two variables, SLP and snow cover extent, is determined by the multiple regression of the October SLP and snow anomalies with the observed winter AO index. This yields the equation

$$SS_i = \alpha \frac{SNOW_i}{\sigma_{sn}} + \beta \frac{SLP_i}{\sigma_{slp}} + \lambda,$$

where  $SS$  is the SLP/snow index,  $SNOW$  is the observed October Eurasian snow cover extent anomaly in  $10^6 \text{ km}^2$ ,  $\sigma_{sn}$  is the standard deviation of October Eurasian snow cover extent,  $SLP$  is the observed October SLP anomaly in northern Eurasia in hPa,  $\sigma_{slp}$  is the standard deviation of October SLP, and  $i$  is the year.

For example, in the hindcasts shown later in the paper (see Figs. 7–9), the value of  $\alpha = 0.25$ ,  $\beta = 0.40$ , and  $\lambda = 0.0$ , as determined by the multiple regression. The correlation value between the October SLP/snow index and the winter AO is 0.9. However, it should be noted that this index is based on forecaster interpretation as the SLP anomaly chosen for the index is derived from an analysis of hemispheric temperature anomalies, Eliassen–Palm flux anomalies, and forecaster experience [techniques are described in Cohen et al. (2002) and Cohen (2003)].

The SLP/snow index is the basis of a simple statistical model employed in real-time winter forecasts for the United States, Europe, and Asia. The current forecast model uses October snow cover and SLP anomalies, plus the recent trend in DJF surface temperatures, as predictors for surface temperatures. This set of predictors has been used operationally for the past three winter forecasts. However, in some of the earlier real-time

forecasts, summer snow cover and ENSO were also used as predictors in the model; the motivation for the different predictors will be presented in section 3. In the remainder of this paper, we will refer to this statistical model as the snow-cast model or sCast model for short.

### c. Forecast/hindcast 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,

$$RMSS = \frac{100}{n} \sum_{i=1}^n \left[ 1 - \sqrt{\frac{(\hat{T}_i - T_i)^2}{(\bar{T} - T_i)^2}} \right],$$

where  $\hat{T}$  is the model predicted temperature,  $T$  is the observed temperature,  $\bar{T}$  is the climatological value of observed temperature,  $i$  is the year, and  $n$  is the total number of years.

Climatology in this study is the long-term mean for the assessment period 1972/73–2004/05; we have found this climatology to provide the best unbiased climatology with respect to the hindcast period. Alternative climatologies could be used such as a fixed 30-yr prior climatology or a rolling prior climatology. The sensitivity of the skill values to the chosen climatology was examined by recomputing the model skill score with a rolling 30-yr prior climatology. This was found to inflate RMSS (not shown) because the rolling prior climatology introduces a cold bias as temperatures have been trending upward since the 1950s, when the rolling climatology begins. This cold bias in the climatology increases the mean-square errors of the climatological hindcasts, thus inflating the RMSS for the hindcast model with respect to climatology.

The mean-square skill score is the preferred skill measure of the World Meteorological Organization (WMO) 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-square error) and hindcasts, we use the closely related RMSS.

The statistical significance of the anomaly correlation is estimated using the 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 significant 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 grid point, and the average gridpoint root-mean-square errors (RMSEs) in the domain (Wilks 1995). Both of these metrics are calculated using data area-weighted by the square root of the cosine of latitude.

### 3. Evaluation of model skill

The focus of the remainder of the paper is to verify the skill of the real-time winter (DJF) forecasts for the past seven years and cross-validated hindcasts. Hindcasts have been produced for the winters 1972/73–2004/05, which complement the full set of available hindcasts from the three dynamical models described in section 2a.

#### 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 10th business day of the month for the following three months (i.e., the forecast for DJF 2005/06 was issued on or before 15 November 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 extent, and the SLP/snow index.

The model has evolved over time in response to the latest research results; hindcasts and forecaster experience and the inputs or predictors have not remained constant. Following Cohen and Entekhabi (1999), the initial predictor employed in the forecast model for the winter of 1999/2000 was October snow cover extent. The first forecast also included trend, since hindcasts showed that including recent trend improved skill over using October snow cover alone. The forecasts for the winters 2000/01 through 2002/03 used the alternating predictors of July and October snow cover extent determined by atmospheric conditions in the fall, following Cohen and Saito (2003). The combined index of July and October snow cover was found to be more

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

Winter	Predictors
1999/2000	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

skillful than October snow and trend. In 2002 El Niño was included as a predictor given the moderate El Niño predicted for that season. For the last three years of the forecast, the October SLP/snow index was used as the main predictor, following Cohen et al. (2002). Hindcasts showed that including the trend contributed some additional skill for predicting winter surface temperatures, especially in the western United States. Though the skill of the SLP/snow is comparable or even less than that of the July/October snow index, July snow cover extent is experiencing a strong decreasing trend, which may be inflating the skill derived from this index. However, October snow cover is experiencing no such decreasing trend. Table 1 lists the predictors used for each real-time winter forecast. The current version of the forecast model linearly combines the October SLP/snow index and recent temperature trend. ENSO is not included as a predictor, as hindcasts showed that it did not improve forecast skill. The hindcasts produced for this study are therefore fixed with just these two predictors.

In Fig. 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. Evident are positive correlations in all four centers of the quadrupole region of temperature variability associated with the AO. Especially high correlations of greater than 0.6 are observed in the eastern United States, along the west coast of North America, northern Europe, and eastern Siberia. However, when plotting the RMSS, only the eastern United States 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. We have found that, over the United States, the correct anomaly sign is predicted in 59% of cases.

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

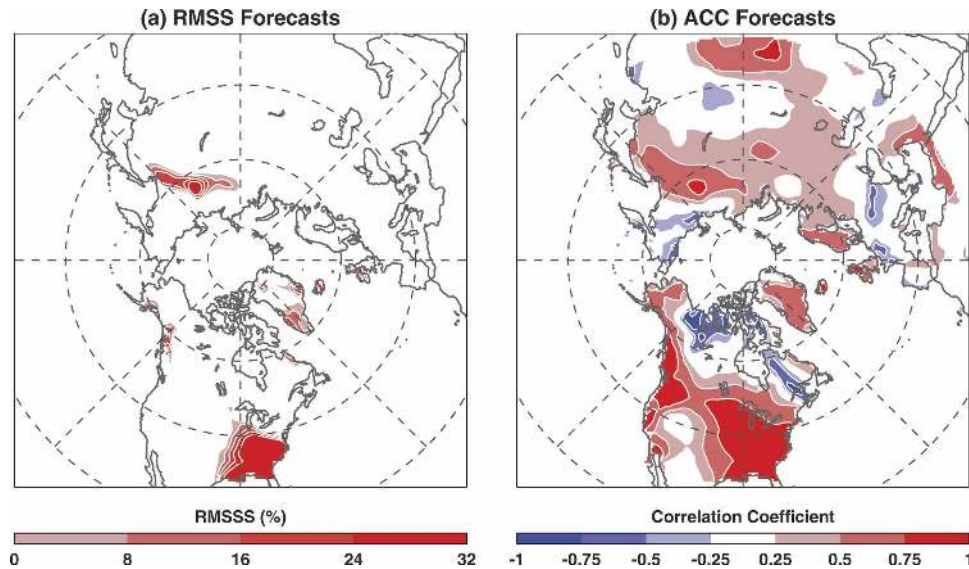


FIG. 3. (a) Root-mean skill score values between real-time predicted and observed DJF NH surface temperatures anomalies for the winters 1999/2000–2005/06 (only positive values shown). (b) Anomaly correlation values between real-time predicted and observed DJF NH surface temperatures anomalies for the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

two skill measures shown in Fig. 3 display the skill achieved at individual grid points, but not the cumulative skill among many grid points or the accuracy of the predicted large-scale pattern of temperature anomalies. Therefore, in Fig. 4 we plot the pattern correlations between the predicted and the observed temperature anomalies for three specific regions and for all seven real-time forecasts. The pattern correlations for the sCast model are shown in darker gray. For comparison we have also included a forecast based on a standard climatology in lighter gray. Correlations for the U.S. real-time forecasts have always been positive, and all years except for 2001/02 equal or exceed 0.6. The highest correlation was achieved in the winter of 2004/05 with a value approaching 0.9. In comparison, a forecast based on recent climatology yields negative scores for the pattern correlations. Also shown are the area-weighted RMSEs for the same regions. In addition to the pattern correlations that are consistently high, the model gives RMSEs that are consistently low, with most years averaging close to  $1^{\circ}\text{C}$  of error or less and on average less than the RMSEs using climatology. Again the winter forecast of 2004/05 scored best among the real-time forecasts.

Included in Fig. 4 are the pattern correlations and the area-weighted RMSEs for Europe and the NH. The model scored positive correlations for all years except one, 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. Again the model handily scores

higher than a forecast based on a standard climatology. Compared to the U.S. forecasts, the correlation values for Europe and the NH are generally lower. The magnitude of the average RMSE is greater than those for the United States and on average close to those derived from using climatology, with values in the range of  $1^{\circ}$ – $2^{\circ}\text{C}$ . Interestingly, the model performed best for the United States in the winter of 2004/05 but poorest for Europe and the NH that same year. The model was most consistent in the winter of 2002/03, performing well in all three regions. This coincides with the second largest observed value of October Eurasian snow cover extent in 2002.

In Table 2 we list the time-mean and standard deviation over the seven real-time forecasts for the pattern correlation and the RMSEs for the United States, Europe, and the NH. All three regions show positive pattern correlations, though the value for the United States is double that of Europe and the NH. Similarly the mean RMSE is  $\sim 1.5^{\circ}\text{C}$  for Europe and the NH but  $\sim 0.9^{\circ}\text{C}$  for the United States.

#### b. Cross-validated hindcasts

##### 1) HINDCASTS WITH THE sCAST MODEL

Cross-validated hindcasts were produced using the SLP/snow index, described in section 2, for the winters of 1972/73 through 2004/05. This predictor index is derived using the same method as that employed in the operational real-time forecasts, which means that it

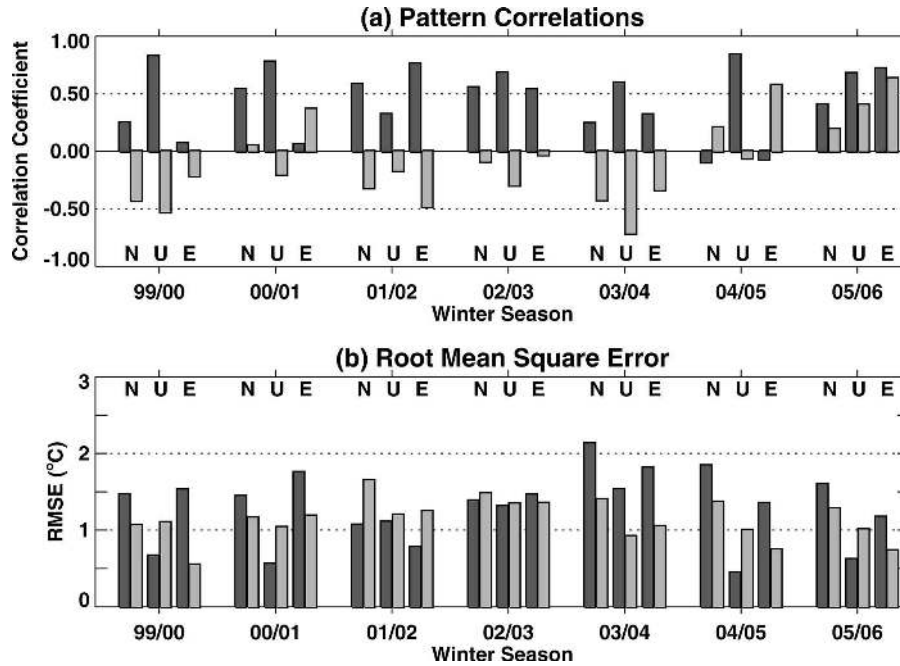


FIG. 4. Real-time DJF surface temperature forecast skill for the sCast system (darker gray) and a climatological forecast using the prior 30-yr mean (lighter gray) averaged over the extratropical Northern Hemisphere (N), United States (U), and Europe (E). (a) Area-weighted pattern correlations and (b) area-weighted RMSE between forecasted and observed temperatures for 1999–2005.

contains subjectivity introduced by forecaster interpretation. In our cross-validation scheme, the observed value of the SLP/snow index and the observed trend are removed when the regression is computed for the temperature forecast. Reliable satellite-derived estimates of snow cover extent have only been available since 1972. In Fig. 5 we plot the 33-yr mean 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. As seen in Figs. 5a,b, the model has positive skill in the eastern United States, parts of southern Canada, the North American Arctic, northern Eurasia, and the Mediterranean region. The regions where the sCast model demonstrates positive skill closely overlap the

quadrupole regions most strongly influenced by variability in the AO and are highlighted in Fig. 2. Therefore the model appears to successfully capture a significant fraction of the temperature variability associated with annual variations in the winter AO as did the real-time forecasts. And by comparing Figs. 5a,b with Fig. 3, the skill demonstrated by the hindcasts is broadly consistent with the skill demonstrated by the model in the real-time forecasts.

In Table 2 we include the mean value and the standard deviation for both the pattern correlation and the RMSE for the United States, 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 United

TABLE 2. The area-weighted pattern correlations and RMSEs (°C) for forecasts and hindcasts of the United States, Europe, and the Northern Hemisphere. Standard deviations for pattern correlations and RMSEs are included in parentheses.

Region	Forecasts (1999–2005)		Hindcasts (1972–2004)			
	Pattern correlations	RMSE	sCast		Simplified sCast	
			Pattern correlations	RMSE	Pattern correlations	RMSE
United States	0.68 (0.17)	0.89 (0.42)	0.33 (0.46)	1.02 (0.76)	0.19 (0.46)	1.12 (0.81)
Europe	0.33 (0.32)	1.41 (0.36)	0.45 (0.32)	1.18 (0.90)	0.24 (0.43)	1.29 (0.98)
Northern Hemisphere	0.35 (0.25)	1.56 (0.34)	0.33 (0.25)	1.33 (1.02)	0.20 (0.32)	1.40 (1.08)



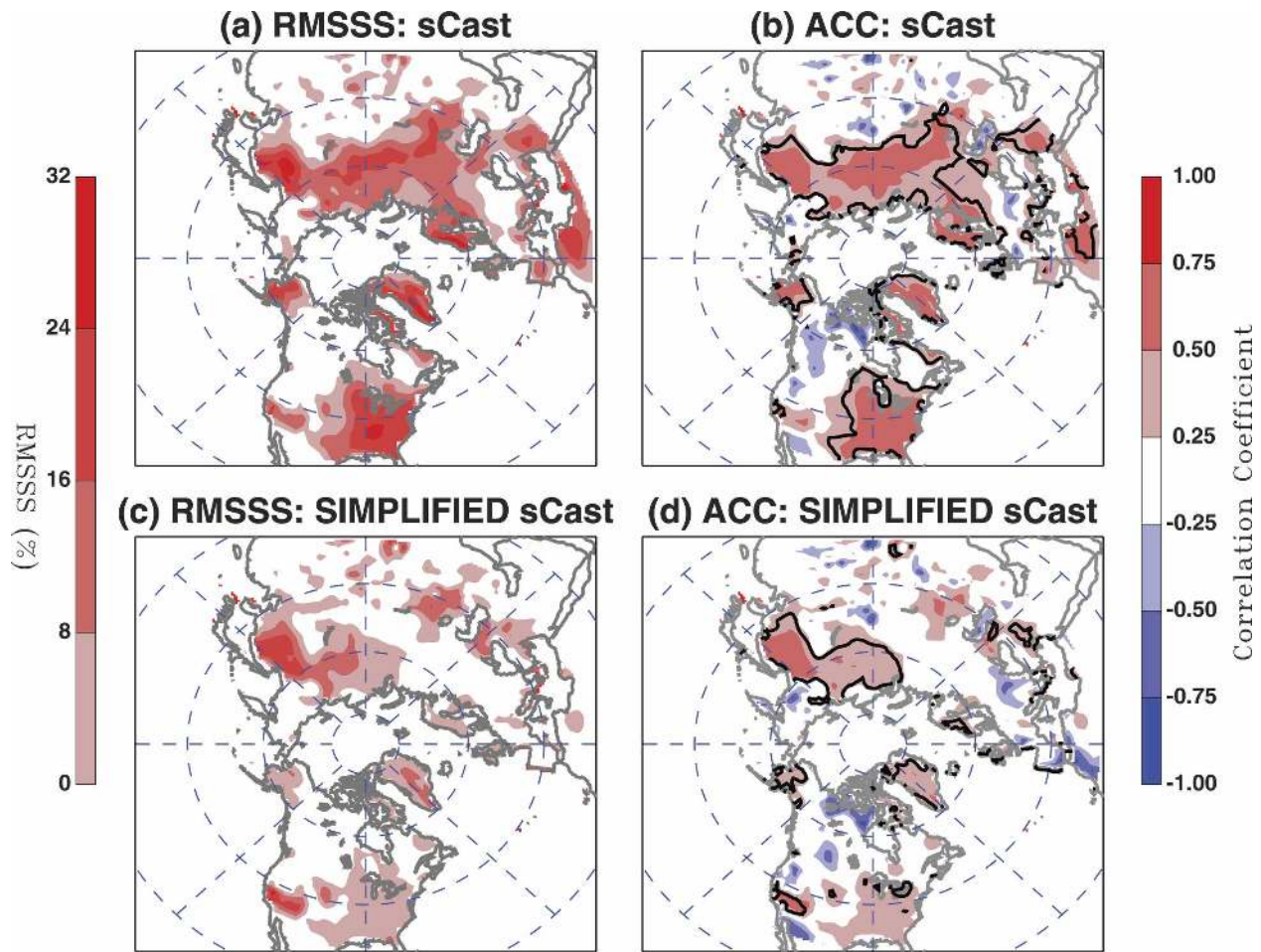


FIG. 5. (a) Root-mean skill score values between hindcasts of sCast model and observed DJF NH surface temperatures anomalies for the winters 1972/73–2004/05 (only positive values shown). (b) Anomaly correlation values between hindcasts of sCast model and observed DJF NH surface temperatures anomalies for the winters 1972/73–2004/05. (c) Same as in (a) and (d) same as in (b) except for the simplified sCast model. See text for differences between sCast and simplified sCast models. Those values exceeding 90%, 95%, and 99% confidence intervals based on the Student's  $t$  test are contoured by thin, thick, and thickest lines, respectively.

States is lower and equal to that of the NH. The average RMSE for all three regions is between  $1.0^{\circ}$  and  $1.4^{\circ}\text{C}$  with the U.S. region scoring the lowest RMSE values. The pattern correlations for the hindcasts are mostly consistent with the forecasts. The favorable comparison between the hindcasts and the forecasts suggests that the positive skill of the forecasts is sustainable over the long term.

One limitation of the SLP/snow index is its dependence on forecaster interpretation; the SLP anomaly chosen for the index is derived from an analysis of hemispheric atmospheric anomalies and forecaster experience. Therefore, it is not appropriate to compare the hindcasts of the operational sCast model with hindcasts from dynamical models, which are independent of forecaster bias. To address this problem, we have derived an alternate SLP/snow index, which produced a

unique and reproducible value for all years of the hindcasts and is independent of forecaster bias. Henceforth, this alternate model is referred to as the simplified sCast model; its development and hindcast verification are described below.

## 2) HINDCASTS WITH THE SIMPLIFIED sCAST MODEL

We have constructed an alternate SLP/snow index that is independent of forecaster interpretation in its derivation. Gridded monthly mean October SLP anomalies for northern Eurasia are analyzed over the domain  $50^{\circ}$ – $80^{\circ}\text{N}$ ,  $0^{\circ}$ – $180^{\circ}$ . 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 con-

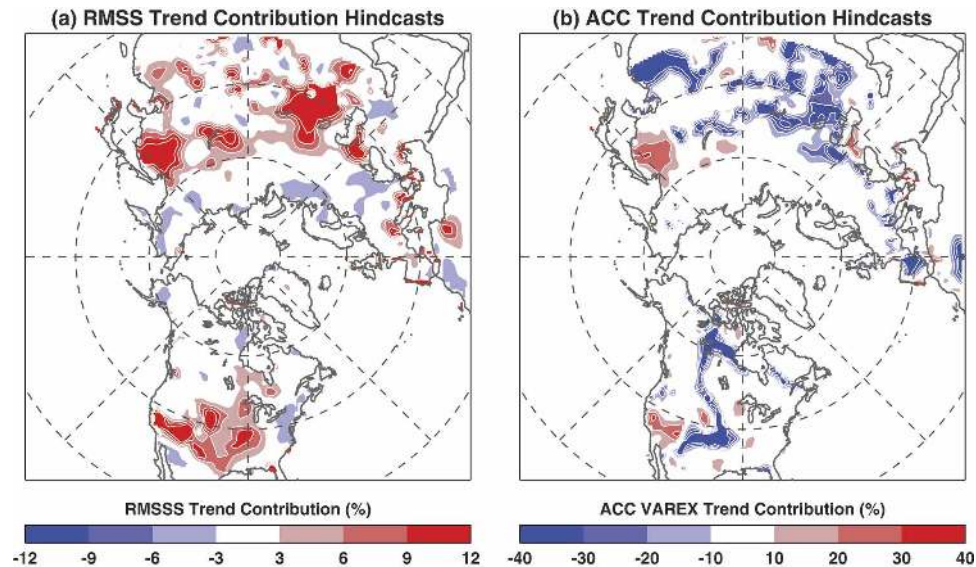


FIG. 6. Relative contribution of trend to (a) root-mean skill score values between hindcasts and observed DJF NH surface temperatures anomalies and (b) anomaly correlation values between hindcasts and observed DJF NH surface temperatures anomalies for the winters of 1972/73–2004/05.

temporaneous 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. Normal is the mean October value, defined as 9.7 million km<sup>2</sup> as determined by NOAA's Climate Prediction Center. For the hindcasts, this algorithm produced a unique solution for the SLP anomaly and therefore a unique value of the SLP/snow index for each year of the hindcast.

In Fig. 2 we plot the SLP/snow index and the correlation value with the winter AO, which is equal to 0.61. Also shown in Fig. 2 is the correlation of the October SLP/snow index with NH  $T_s$ . Again the quadrupole pattern of temperature variability is noted with the same-signed anomaly across northern Eurasia and the eastern United States and opposite-signed anomalies in northeastern Canada, Greenland, North Africa, and the Mediterranean. In comparison to the correlation map for October snow cover alone (Fig. 1b), the correlations are higher and parts of Europe are now included in the region of significant correlations. The plot closely resembles that of the AO correlated with  $T_s$ .

Shown in Figs. 5c,d are the RMSS and anomaly correlations from the hindcasts of the simplified sCast model. The model demonstrates positive skill in the same regions as the operational sCast model and where the SLP/snow index is significantly correlated with NH surface temperatures, though the values are more modest. As seen from Table 2, the simplified sCast model shows positive pattern correlations for the United States, Europe, and the NH, with values more closely

resembling the hindcasts of the operational sCast model rather than the forecasts. The average RMSE for all three regions is between 1.2° and 1.5°C, with the U.S. region once again scoring the lowest RMSE values. Though the skill for the hindcasts using the simplified sCast is more modest, it still represents a large improvement over other operational forecast models in the mid- to high latitudes of the NH, as will be shown in section 3d.

### c. Trend versus SLP/snow index

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 (Cohen and Barlow 2005). 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 Fig. 6 we plot the relative contribution of trend to the overall model skill for the extratropical NH. Based on the anomaly correlation, the trend contributes no discernible skill to the forecast. However, for the RMSS, the trend contributes positive skill in regions where the SLP/snow index is not significantly correlated with  $T_s$ , parts of the western United States, and central Eurasia. In fact over some of the regions where the model has its highest skill, such as in the eastern United States and northern Europe, the

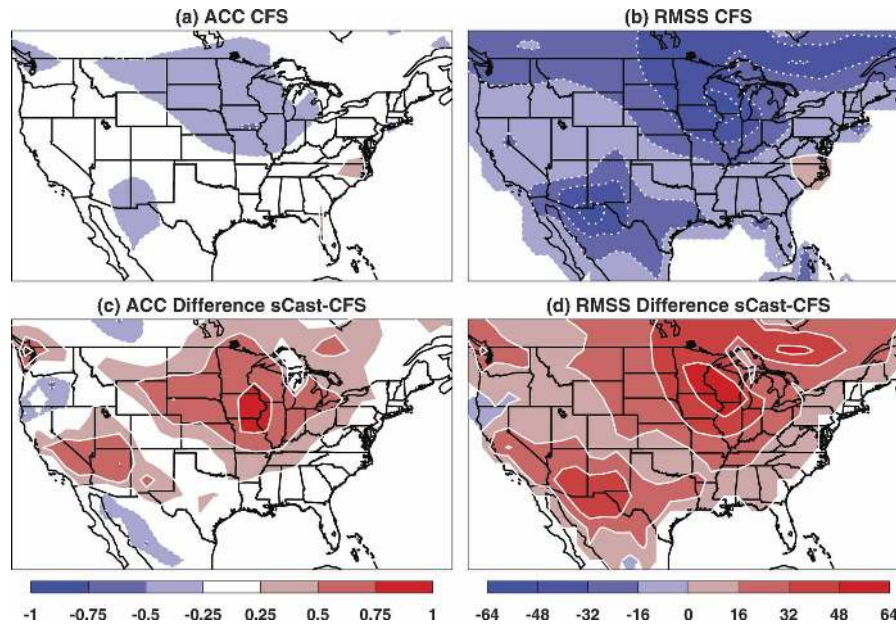


FIG. 7. (a) Anomaly correlation values between hindcasts and observed DJF U.S. surface temperature anomalies for NOAA's CFS GCM and (b) root-mean skill score values between hindcasts and observed DJF U.S. surface temperature anomalies for the winters of 1981/82–2003/04. Difference between the simplified sCast model and CFS GCM for overlapping hindcast winters of 1981/82–2003/04 for (c) anomaly correlations and (d) root-mean skill score values. In (c) and (d), red shading indicates where simplified sCast model has greater skill and blue shading indicates where CFS model has greater skill. RMSS values are in percent.

trend contributes negative skill. The skill from trend appears to extend the spatial extent, or complement, the skill derived from the SLP/snow index rather than contribute significantly to or overlap those regions where the SLP/snow index has skill.

#### d. Comparison with dynamical models

We now compare the hindcast skill achieved using the simplified sCast model with that of three GCMs employed for seasonal forecasting at leading world forecast centers (listed in section 2a). These highly complex dynamical models represent many of the major processes in the ocean–land–atmosphere climate system; however, nearly all of their seasonal forecast skill can be attributed to variability in ENSO (van Oldenborgh et al. 2005a,b; Quan et al. 2006; Saha et al. 2006). As discussed in section 2b, ENSO variability offers only limited atmospheric predictability away from the Tropics. Therefore, given the completely different emphasis of the sCast model, it is a worthwhile exercise to compare between the skill derived from hindcasts performed using sCast and the dynamical forecast systems.

In Figs. 7a,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 Figs. 7c,d, we plot the difference in skill between the simplified sCast model and the CFS model; blue shading indicates that the CFS model has greater skill and red shading that the simplified sCast model has greater skill. The superior skill demonstrated by the simplified sCast model is especially large in the eastern United States, a region not well correlated with ENSO variability but highly correlated with AO variability. For both skill metrics, the simplified sCast model demonstrates greater skill for most of the United States, especially when comparing the RMSS. In Fig. 8 we show the same plot as Fig. 7, but now for the entire extratropics of the NH. The superior skill of the simplified sCast model is not limited to the United States but is widespread across the NH extratropics, including most of Europe and Asia.

In Fig. 9 we plot the difference between the Canadian seasonal forecast GCM and the simplified sCast model in the top panels and the difference between the DEMETER ensemble of seasonal forecast models and the simplified sCast model in the bottom panels for both the anomaly correlation (left) and the RMSS



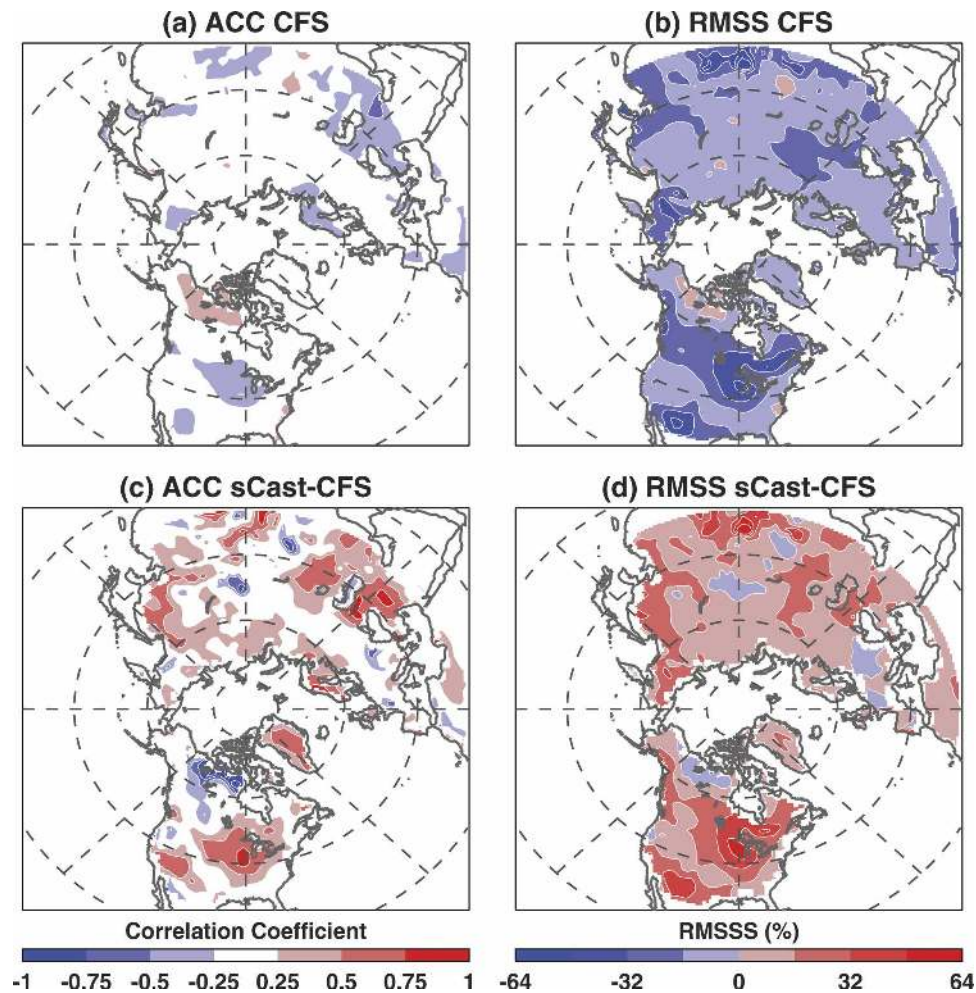


FIG. 8. Same as in Fig. 7, but now for the entire extratropics of the NH.

(right) for the extratropical NH. Again, the simplified sCast compares favorably with the two dynamical forecast systems. The simplified sCast model performed with higher skill across the eastern United States and northern Eurasia. The skill from the DEMETER ensemble hindcasts is greater than those from the CFS and HFP models. The simplified sCast model only performed marginally better when compared to DEMETER across the United States 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 simplified sCast model (not shown).

#### 4. Discussion and conclusions

Seasonal forecasts have traditionally relied on variability in SSTs in general and ENSO variability in par-

ticular to generate skillful forecasts. Though the ENSO pattern of variability is one of the dominant global patterns of variability, its greatest impacts are focused in the Tropics, with a much more damped signal in the extratropics. The AO is also one of the dominant patterns of variability in the NH but most of its related variability is focused in the extratropics, with less of a signal in the Tropics. The sCast model, a statistical model, attempts to predict the sign and magnitude of the upcoming winter AO, which is then utilized as a predictor of DJF surface temperatures.

The early sCast climate forecast model used observed Siberian snow cover extent and recent temperature trends to predict winter surface temperatures. However, statistical analysis demonstrates that by combining the anomaly in October Eurasian snow cover extent with a second variable, the anomaly in October monthly mean SLP, the hindcast skill of extratropical NH DJF surface temperatures is improved. The corre-



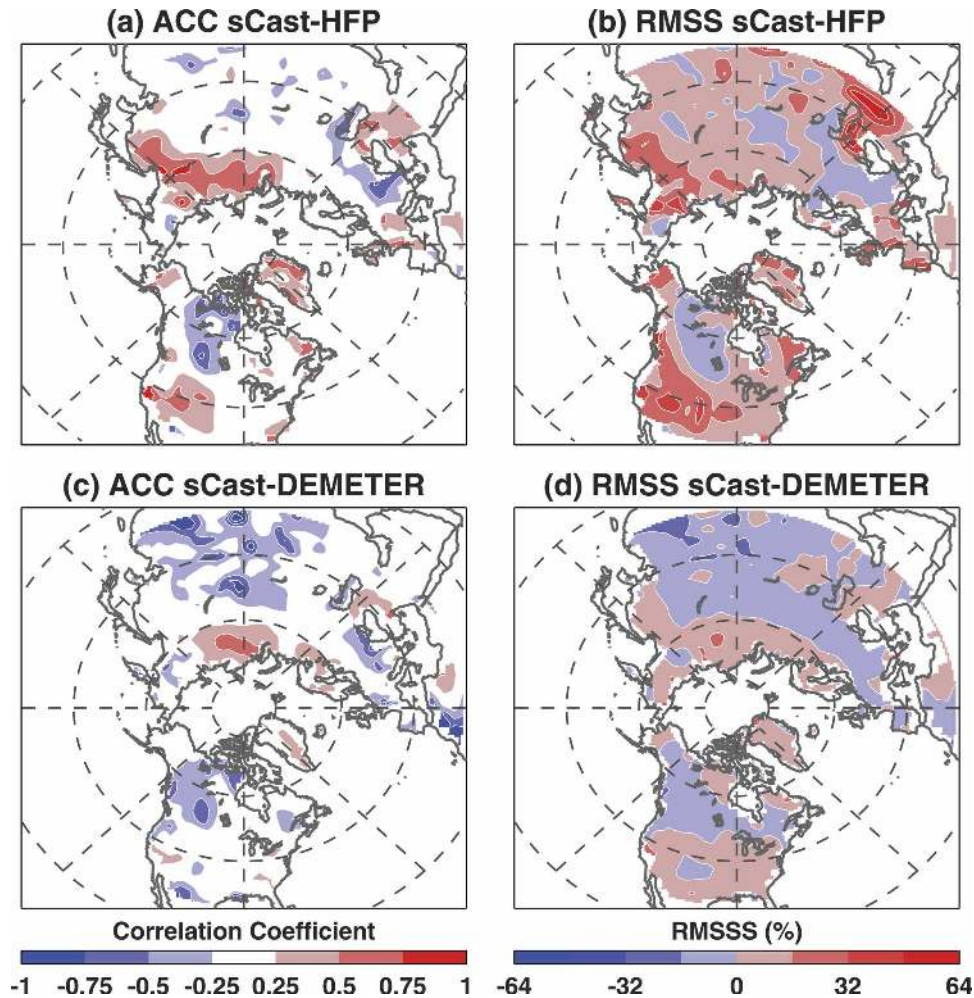


FIG. 9. Difference between the simplified sCast and Canadian HFP GCM for overlapping hindcast winters of 1972/73–1992/93 for (a) anomaly correlations and (b) root-mean skill score values. Red shading indicates where the simplified sCast model has greater skill and blue shading indicates where HFP GCM has greater skill. (c) Same as in (a) and (d) same as in (b), except for the DEMETER ensemble of GCMs and for the winters of 1972/73–2000/01.

lation of this combined October SLP/snow index with DJF surface temperatures yields a region of statistically significant correlation comparable in extent, though slightly reduced, to the contemporaneous correlation of the DJF AO with DJF surface temperatures. Favorable comparison of the October SLP/snow index with the DJF AO is an important advance for climate prediction given that the DJF AO is the single variable with the highest observed correlations with extratropical DJF NH surface temperatures (Thompson and Wallace 1998).

Real-time forecasts using the sCast statistical model demonstrate positive skill for parts of the eastern and western United States, northern Eurasia, and the Mediterranean and Middle East, as measured by the RMSS and anomaly correlations. Similar skill was found for

33 yr of hindcasts for the sCast model, providing confidence that the skill of the model is sustainable over the long term. For comparison with hindcasts from three major operational forecast centers, we derived an alternate model, which we refer to as the simplified sCast model. For the simplified sCast model we chose a more simple and objective index where the highest skill was sacrificed in order to facilitate reproducibility. Hindcast skill was found to be higher for the operational sCast model than the simplified sCast model, especially for the United States, which can be attributed to the difference in the SLP/snow index employed by these two models. Nonetheless, the more modest skill demonstrated by the simplified sCast model for the eastern United States, Europe, and Asia is still 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 simplified sCast model and the dynamical models of CPC, Environment Canada, and ECMWF demonstrate the superior skill of the sCast model in these same regions, often by a wide margin. Furthermore, the higher forecast and hindcast skill of the operational sCast model compared with the simplified sCast model suggests that forecast skill can potentially be further improved by the knowledge and experience of a forecaster.

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 still strongly dependent on ocean-atmosphere coupling associated with ENSO for much of their skill. However, given the impact regions of ENSO, there is little reason to believe that these models are capable of the accuracy required to produce a forecast of benefit to society in the important regions of the eastern United States 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 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 R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126.
- , and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114.
- , and Coauthors, 1999: Review of skill of CPC's long-lead seasonal U.S. predictions since 1995. *Proc. 24th Annual Climate Diagnostics and Prediction Workshop*, Tucson, AZ, NOAA/Climate Prediction Center, 13–16.
- Bojariu, R., and L. Gimeno, 2003: The role of snow cover fluctuations in multiannual NAO persistence. *Geophys. Res. Lett.*, **30**, 1156, doi:10.1029/2002GL015651.
- Cohen, J., 1994: Snow cover and climate. *Weather*, **49**, 150–156.
- , 2003: Introducing sub-seasonal spatial and temporal resolution to winter climate prediction. *Geophys. Res. Lett.*, **30**, 1018, doi:10.1029/2002GL016066.
- , and D. Entekhabi, 1999: Eurasian snow cover variability and Northern Hemisphere climate predictability. *Geophys. Res. Lett.*, **26**, 345–348.
- , and K. Saito, 2002: A test for annular modes. *J. Climate*, **15**, 2537–2546.
- , and —, 2003: Eurasian snow cover, more skillful in predicting U.S. winter climate than the NAO/AO? *Geophys. Res. Lett.*, **30**, 2190, doi:10.1029/2003GL018053.
- , and M. Barlow, 2005: The NAO, the AO, and global warming: How closely related? *J. Climate*, **18**, 4498–4513.
- , K. Saito, and D. Entekhabi, 2001: The role of the Siberian high in Northern Hemisphere climate variability. *Geophys. Res. Lett.*, **28**, 299–302.
- , D. Salstein, and K. Saito, 2002: A dynamical framework to understand and predict the major Northern Hemisphere mode. *Geophys. Res. Lett.*, **29**, 1412, doi: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. Meteor. Soc.*, **83**, 1303–1311.
- Fletcher, C. G., and M. A. Saunders, 2006: Winter North Atlantic Oscillation hindcast skill 1900–2001. *J. Climate*, **19**, 5762–5776.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679.
- , 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. Meteor. 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.
- NOAA, 2002: NOAA economic statistics. Office of Policy and Strategic Planning, U.S. Department of Commerce, 26 pp.
- 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.
- , and Coauthors, 2004: Development of a European multi-model ensemble system for seasonal-to-interannual prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872.
- 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.
- Quan, X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of U.S. seasonal forecast skill. *J. Climate*, **19**, 3279–3293.
- Robinson, D. A., K. F. Dewey, and R. R. Heim Jr., 1993: Global snow cover monitoring: 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*, **19**, 3483–3517.

- 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**, 1302, doi:10.1029/2002GL016341.
- Saunders, M. A., B. Qian, and B. Lloyd-Hughes, 2003: Summer snow extent heralding of the winter North Atlantic Oscillation. *Geophys. Res. Lett.*, **30**, 1378, doi:10.1029/2002GL016832.
- 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.
- , and —, 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, 2005a: Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *J. Climate*, **18**, 3240–3249.
- , —, —, —, and —, 2005b: Evaluation of atmospheric fields from the ECMWF seasonal forecasts over a 15-year period. *J. Climate*, **18**, 3250–3269.
- 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 Rep. 485, Vol. 1, 21 pp.