
CHRISTOPHER G. FLETCHER* AND MARK A. SAUNDERS

Department of Space and Climate Physics, Benfield Hazard Research Centre, University College London, Surrey, United Kingdom

(Manuscript received 7 January 2005, in final form 13 March 2006)

ABSTRACT

Recent proposed seasonal hindcast skill estimates for the winter North Atlantic Oscillation (NAO) are derived from different lagged predictors, NAO indices, skill assessment periods, and skill validation methodologies. This creates confusion concerning what is the best-lagged predictor of the winter NAO. To rectify this situation, a standardized comparison of NAO cross-validated hindcast skill is performed against three NAO indices over three extended periods (1900–2001, 1950–2001, and 1972–2001). The lagged predictors comprise four previously published predictors involving anomalies in North Atlantic sea surface temperature (SST), Northern Hemisphere (NH) snow cover, and an additional predictor, an index of NH subpolar summer air temperature ($T_{SP}$). Significant ($p < 0.05$) NAO hindcast skill is found with May SST 1900–2001, summer/autumn SST 1950–2001, and warm season snow cover 1972–2001. However, the highest and most significant hindcast skill for all periods and all NAO indices is achieved with $T_{SP}$. Hindcast skill is nonstationary using all predictors and is highest during 1972–2001 with a $T_{SP}$ correlation skill of 0.59 and a mean-squared skill score of 35%. Observational evidence is presented to support a dynamical link between summer $T_{SP}$ and the winter NAO. Summer $T_{SP}$ is associated with a contemporaneous midlatitude zonal wind anomaly. This leads a pattern of North Atlantic SST that persists through autumn. Autumn SSTs may force a direct thermal NAO response or initiate a response via a third variable. These findings suggest that the NH subpolar regions may provide additional winter NAO lagged predictability alongside the midlatitudes and the Tropics.

1. Introduction

The North Atlantic Oscillation (NAO) is the dominant mode of boreal winter climate variability over the North Atlantic (Walker and Bliss 1932; Barnston and Livezey 1987). The NAO is strongly linked to patterns of winter temperature, precipitation, and storminess over the whole North Atlantic sector (Hurrell 1995; Trigo et al. 2002). Accurate and timely forecasts of the winter NAO are therefore an important challenge for seasonal forecasters and increasing research is focusing on improving the quality and lead time of NAO predictions.

Empirical winter NAO forecast models are formulated using observations of slowly varying boundary variables such as sea surface temperature (SST) and snow cover. The persistence of these boundary conditions and their forcing of the atmosphere can make seasonal atmospheric conditions predictable at a later time (Goddard et al. 2001). Winter NAO forecasts have also recently been produced using general circulation models (GCMs; e.g., Palmer et al. 2004). However, only one or two GCMs produce statistically significant forecast skill at levels comparable to empirical models, and only when assessed over a relatively short time period (1980–2001).

Recent empirical studies have uncovered links between North Atlantic SST anomalies in the preceding late spring/summer/autumn and the upcoming winter NAO. Rodwell and Folland (2002) highlighted a pattern of May North Atlantic SST that is related significantly to the upcoming winter mean 500-hPa geopotential height field, achieving a hindcast correlation skill ($r_s$) of $r_s = 0.45$ for the period 1948–97. Saunders and Qian (2002) found that two modes of late summer/early autumn North Atlantic SST variability 1950–2001 were skillful in predicting a range of upcoming winter NAO
indices with $r_s$ between $r_s = 0.47$ and $r_s = 0.63$. A horse-shoe-like North Atlantic summer SST pattern has also been shown to explain up to 16% of the early winter atmospheric variance 1958–97 (Dréville et al. 2001; Czaja and Frankignoul 2002; Cassou et al. 2004).

Another lagged NAO predictor is the monthly mean areal extent of snow cover over different regions of the Northern Hemisphere (NH). Reliable satellite observations of snow cover are available only since 1972 (Robinson et al. 1993). Thus, the role of anomalous snow cover in influencing atmospheric circulation and climate has received less attention than that of SST anomalies. Saunders et al. (2003) used NH snow cover to predict a range of NAO indices and found that the June–July mean produced the most significant hindcast skill with $r_s = 0.61$ for 1972–2001. A significant correlation of $r = -0.56$ is highlighted by Bojariu and Gimeno (2003) between April and October mean Eurasian (EU) snow cover and the upcoming winter NAO index 1973–98. These studies suggest a recent significant link between NH climate in the warm season and that in the upcoming winter. Cohen and Entekhabi (1999) highlighted the significant relationship between EU snow cover during fall and the upcoming winter Arctic Oscillation (AO) index. However, these links were not significant when assessed against the NAO index.

The above discussion highlights the current confusing situation in the literature where several authors claim to have uncovered significant predictive links and skill for the winter NAO using different prior climatic predictors. Furthermore, these authors all reference these links and skill to different NAO indices and/or employ different predictand time periods and/or use different skill assessment methodologies and measures. In this study we compare the most skillful recent empirical hindcast models using the same hindcast procedure for all predictors. This provides a standardized platform on which to evaluate and compare the skill arising from the chosen NAO predictors. Hindcasts are made for three different winter NAO indices over three extended time periods out to 100 yr and skill is assessed using two skill measures. Where possible, hindcast skill sensitivity to choice of predictor observational dataset is also examined by using two datasets for each analysis. These methods will allow us to quantify which of the lagged predictors drawn from the recent literature provides the best hindcast skill for the upcoming winter NAO.

The paper is structured as follows. Section 2 outlines the chosen assessment periods, predictand NAO indices, predictors and the different observational datasets used to compute them. Here we also describe our hindcast procedure and the method for evaluating hindcast skill and its statistical significance. Results from the hindcast comparison are presented in section 3, where stationarity of the predictive relationships is also discussed. Section 4 provides an interpretation of the key findings and examines the physical mechanisms linking the most skillful lagged predictor to the winter NAO. Section 5 discusses the implications of our work for seasonal forecasting in general and a summary and conclusions are presented in section 6.

2. Data and methods

Henceforth, “winter” denotes the three-month December–February (DJF) seasonal mean and “winter NAO” will be referred to as $\text{NAO}_{\text{DJF}}$. Previous studies of $\text{NAO}_{\text{DJF}}$ hindcast skill have examined data since 1950. Here, we assess hindcast skill, where possible, over the entire twentieth century. Specifically, $\text{NAO}_{\text{DJF}}$ hindcast skill is assessed for the three periods 1900–2001, 1950–2001, and 1972–2001, with the 1972–2001 period corresponding to the interval of reliable satellite snow cover data (Robinson et al. 1993). These three assessment periods also allow the stationarity in $\text{NAO}_{\text{DJF}}$ hindcast skill to be examined. Figure 1 shows the time series of the three $\text{NAO}_{\text{DJF}}$ indices employed as predictands in this study. The first index is the standardized difference in mean sea level pressure (MSLP) between Reykjavik, Iceland, and Gibraltar maintained by the Climatic Research Unit (CRU) at the University of East Anglia (henceforth the CRU NAO$_{\text{DJF}}$ index; Jones et al. 1997). The second is defined as the standardized MSLP difference between Stokkisholmur, Iceland, and Ponta Delgada, Azores, provided by the National Center for Atmospheric Research (NCAR) Climate Analysis Section (henceforth the Hurrell NAO$_{\text{DJF}}$ index). The third index is the time series of 500-hPa North Atlantic geopotential height coefficients from a singular value decomposition (SVD) analysis with May North Atlantic SST (henceforth the Z500$_{\text{DJF}}$ index; Rodwell and Folland 2002).

a. Previously published $\text{NAO}_{\text{DJF}}$ predictors

The four previously published $\text{NAO}_{\text{DJF}}$ predictors examined in this study are shown in Table 1 and are taken from results published since 2001. They were selected for comparison because they are believed to encompass the highest currently observed seasonal predictive skill for the winter NAO.

The first $\text{NAO}_{\text{DJF}}$ predictor [henceforth the May SST (SVD)] is the time series of May North Atlantic (10°–80°N) SST derived from an SVD analysis with DJF
mean 500-hPa geopotential height over the North Atlantic sector (Rodwell and Folland 2002). The SVD procedure was performed in “cross-validated” mode, with the resulting SST expansion coefficients called the “predicted” NAO series and the corresponding geopotential height coefficients the “observed” NAO series. The correlation skill was taken as the Pearson correlation between these two time series.

Our implementation of the SVD analysis differs from that employed by Rodwell and Folland (2002). First, in accordance with other analyses in this study, we employ a 5-yr block elimination to the SVD procedure. This means that we remove each year in turn and additionally 2 yr on either side. Second, we remove sea ice from the SST fields prior to performing the SVD (sea ice was retained in the original analysis). Third, we employ different data periods. Fourth, we compute skill from a detrended data analysis. These differences in implementation mean that our results are expected to differ from those of Rodwell and Folland (2002). However, as a preliminary check, we first reproduced the cross-validated hindcast correlation skill ($r = 0.45$) using their original implementation. This predictor maximizes covariance with the Z500$_{DJF}$ index; however, we also assess its hindcast skill against all NAO indices.

The second predictor is the second principal component (PC2) of the June–October (JJASO) mean North Atlantic ($0°$–$65°$N) SST. This is slightly the stronger of two lagged JJASO SST predictive modes employed by Saunders and Qian (2002). The PC2 loading pattern (see their Fig. 1a) features a ring of SST anomalies around an opposite-signed center off southeast Newfoundland.

The third and fourth predictors employ the monthly mean areal extent of snow cover, with observations considered reliable since 1972. The third predictor is snow cover over Eurasia for the period April–October (A–O; Bojariu and Gimeno 2003). The fourth predictor

**TABLE 1. Published NAO$_{DJF}$ lagged predictors examined in this paper.**

<table>
<thead>
<tr>
<th>Lagged predictor</th>
<th>Assessment period</th>
<th>Lead time (months)</th>
<th>Domain area</th>
<th>NAO$_{DJF}$ index</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>JJASO SST (PC2)</td>
<td>1950–2001</td>
<td>1</td>
<td>0°–65°N, 100°W–0°</td>
<td>CRU/CPC*/PC1</td>
<td>Saunders and Qian (2002)</td>
</tr>
<tr>
<td>JJ NH snow cover</td>
<td>1972–2002</td>
<td>4</td>
<td>n/a</td>
<td>CRU/CPC/PC1</td>
<td>Saunders et al. (2003)</td>
</tr>
</tbody>
</table>

* U.S. Climate Prediction Center index.
is the June–July (JJ) mean snow cover for the entire NH (Saunders et al. 2003).

b. An additional NAO DJF predictor

Alongside the four previously published predictors outlined above, we also compare an index of JJ mean NH subpolar (60°–70°N) 2-m air temperature. This index and its link to the upcoming winter NAO was introduced by Saunders et al. (2003). The index is defined as

\[ T_{\text{SUBPOLAR}} = \frac{\text{NA} + \text{EU}}{2} - \text{SG}, \]  

where North America (NA), Eurasia (EU), and southern Greenland (SG) refer to subpolar (60°–70°N) 2-m air temperatures over these areas (Fig. 2). These centers correspond to where gridded JJ air temperature is correlated most significantly with contemporaneous NH snow cover (|r| ~ 0.5 1972–2001).

Gridded 2-m air temperature data back to 1900 are available from the CRU Air Temperature Anomalies version 2 (CRUTEM2) dataset (Jones and Moberg 2003) for NA, EU, and SG. We therefore calculate the \( T_{\text{SUBPOLAR}} \) index (\( T_{\text{SP}} \)) for the period 1900–2001 and employ it as an NAO\(_{DJF}\) predictor. The summer period is selected because during the summer months the relationship peaks between the \( T_{\text{SP}} \) index and the upcoming winter NAO (Saunders et al. 2003). Figure 3 shows that the link is strongest in JJ using both the CRUTEM2 and the (National Centers for Environmental Prediction) NCEP–NCAR reanalysis data (Kallay et al. 1996) for the 50- and 30-yr assessment periods. However, for the 100-yr period the link is significant over an extended summer period May–September (MJJAS). Therefore, we also evaluate the NAO\(_{DJF}\) predictive skill of the longer MJJAS mean \( T_{\text{SP}} \).

c. Predictor datasets

The May SST (SVD) assessment back to 1900 is performed using MSLP data because pre-1948 geopotential height data are not available. This analysis is performed over the shorter period 1900–98. This is the period of extended MSLP data available from the United Kingdom Met Office (UKMO) Hadley Centre Sea Level Pressure (HadSLP) dataset, which is an update of the Met Office’s Global Mean Sea Level Pressure (GMSLP2.1f; Basnett and Parker 1997). Geopotential heights at 500 hPa (1950–2001) are taken from the NCEP–NCAR reanalysis (Kallay et al. 1996). The NH gridded surface air temperatures 1900–2001 are from the University of East Anglia CRUTEM2 (5° × 5° latitude–longitude grid) dataset (Jones and Moberg 2003) and for the periods since 1950 are also from the NCEP–NCAR reanalysis.

Two global gridded SST datasets are employed in this study. For the period 1900–2001, we use the Met Office Hadley Centre Coupled Sea Ice and SST (HadISST) dataset (Rayner et al. 2003) and for the period 1950–2001 SSTs are also taken from NCEP–NCAR reanalysis. All SST data are interpolated onto a 2.5° × 2.5° latitude–longitude grid for computational efficiency. Snow cover data for EU and NH are satellite-derived estimates of monthly mean areal extent (Robinson et al. 1993).

Throughout our analysis we employ linear detrended time series corrected for autocorrelation to compute regressions and hindcast skill. This approach minimizes the influence of time series trends and multiyear-to-decadal signal variability on the strength and significance of the deduced hindcast skill. The use of raw (not detrended) time series corrected for autocorrelation gives in most cases hindcast skills that are of similar magnitude and significance to those achieved with the detrended data analysis. For reference, Table 2 shows the linear trends in NAO\(_{DJF}\) and the three predictors for each assessment period. Over the 100-yr period the trends are negative or close to zero. However, since...
1950 all three NAO\textsubscript{DJF} indices and JJ \(T_{SP}\) exhibit positive trends of \(\sim 0.2\)–\(0.3\) standard deviations per decade. Also shown is the percentage of NAO\textsubscript{DJF} hindcast correlation skill that is attributable to linear trend, which is calculated as

\[
\Delta_{\text{Skill}} = \left(1 - \frac{r_s^2}{r^2}\right) \times 100\%,
\]

\(\text{where } r_s\) is the correlation skill from the detrended data analysis and \(r_s^2\) is the correlation skill obtained using raw (not detrended) time series. For the 1900–2001 period the inclusion of trends decreases the hindcast skill by 4%–200%; for the 1950–2001 period it increases skill by 14%–46% (for May SST there is a 30%–50% decrease); for 1972–2001 it decreases skill by 0%–40% (for May SST there is a 40%–60% increase). These differences reflect the decadal and longer time-scale variability in the NAO\textsubscript{DJF} and predictor indices (see also section 3b).

**d. Hindcast methodology**

The predictive skill of the selected NAO\textsubscript{DJF} predictors is computed for each assessment period using cross-validated (Michaelsen 1987) hindcasts with block elimination. Except for the May SST (SVD) predictor, linear empirical models for the NAO\textsubscript{DJF} index \(\hat{y}\) in a given year \(t\) using predictor \(x\) take the following form:

\[
\hat{y}_t = \beta_0 + \beta_1 x_t + e_t,
\]

\(\text{where the coefficients } \beta_0\) and \(\beta_1\) are determined by an ordinary least squares regression and \(e_t\) is the residual. A 5-yr block elimination is employed, which means that for a given predictand year \(t\), data for the years \(t - 2, t -1, t, t + 1,\) and \(t + 2\) are removed from the regression. Furthermore, predictors derived from principal

---

**Table 2.** Linear trend in NAO\textsubscript{DJF} indices and predictors in standardized units per decade (columns 3, 4, 6, and 8). The percentage of NAO\textsubscript{DJF} hindcast correlation skill attributable to linear trend (\(\Delta_{\text{Skill}}\), columns 5, 7, and 9). See text for calculation of \(\Delta_{\text{Skill}}\).

<table>
<thead>
<tr>
<th>Assessment period</th>
<th>NAO\textsubscript{DJF} index</th>
<th>May SST (SVD)</th>
<th>JJASO SST (PC2)</th>
<th>JJ (T_{SP})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trend</td>
<td>(\Delta_{\text{Skill}})</td>
<td>Trend</td>
<td>(\Delta_{\text{Skill}})</td>
</tr>
<tr>
<td>1900–2001</td>
<td>CRU NAO\textsubscript{DJF}</td>
<td>-0.07</td>
<td>0.03</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>Hurrell NAO\textsubscript{DJF}</td>
<td>-0.02</td>
<td>—</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>Z500\textsubscript{DJF}</td>
<td>-0.05</td>
<td>—</td>
<td>-4</td>
</tr>
<tr>
<td>1950–2001</td>
<td>CRU NAO\textsubscript{DJF}</td>
<td>0.21</td>
<td>0.33</td>
<td>-40</td>
</tr>
<tr>
<td></td>
<td>Hurrell NAO\textsubscript{DJF}</td>
<td>0.21</td>
<td>—</td>
<td>-29</td>
</tr>
<tr>
<td></td>
<td>Z500\textsubscript{DJF}</td>
<td>0.27</td>
<td>—</td>
<td>-50</td>
</tr>
<tr>
<td>1972–2001</td>
<td>CRU NAO\textsubscript{DJF}</td>
<td>0.23</td>
<td>0.52</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Hurrell NAO\textsubscript{DJF}</td>
<td>0.13</td>
<td>—</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Z500\textsubscript{DJF}</td>
<td>0.23</td>
<td>—</td>
<td>60</td>
</tr>
</tbody>
</table>
components (PCs) are also subjected to block elimination in their formulation. For each predictor, a time series of hindcast NAO\textsubscript{DJF} values is calculated whose skill is verified against the corresponding observed NAO\textsubscript{DJF} series.

ee. Skill assessment

In validating the hindcast skill from each predictor, we employ the Pearson product-moment correlation skill \( r_s \) and the mean-squared skill score (MSSS) against a simple prediction of climatology. Climatology in this study is the long-term mean for each assessment period because the NAO\textsubscript{DJF} is known to be nonstationary during the twentieth century (Hurrell and van Loon 1997). This nonstationarity means that no single 30-yr climatology period is representative of the twentieth century as a whole. However, our conclusions are not sensitive to the choice of climatology.

The statistical significance of each skill value is estimated using a randomized resampling method (Manly 1997). Our null hypothesis \( (H_0) \) is that there is no link between a given predictor and the NAO\textsubscript{DJF} and therefore any observed skill is achieved by random chance. The observed and hindcast NAO\textsubscript{DJF} time series are both randomly shuffled (with replacement) to create two new synthetic time series drawn from the same populations as the original time series. The length of the two synthetic time series is reduced to equal the “effective” number of degrees of freedom between the original time series, taking into account serial correlation (Davis 1976). Skill values are calculated for the two randomized time series and this shuffling/resampling process is repeated in Monte Carlo fashion 25,000 times. The statistical significance is determined by the number of cases where the skill value from the random data exceeds the original observed skill value. The a priori threshold of statistical significance is set at 0.05, which represents the probability that \( H_0 \) is incorrectly rejected. Our test is one tailed as only positive values of \( r_s \) and MSSS will lead to a rejection of \( H_0 \) (Wilks 1995).

3. NAO hindcast skill

a. Comparison by assessment period

Table 3 shows the NAO\textsubscript{DJF} cross-validated hindcast skill (and significance) for each predictor against the three NAO\textsubscript{DJF} indices over all three assessment time periods. For the 100-yr period, significant hindcast skill is found using May SST (SVD) and both \( T_{SP} \) index predictors. The long summer [May–September (MJJAS)] mean produces the highest skill (a 6%–9% improvement over a prediction of climatology) and the JJ mean results in a 2%–5% improvement. Values of \( r_s \) range from 0.26 to 0.31 for MJJAS, which represent 7%–10% of NAO\textsubscript{DJF} variance over 100 yr. The JIASO SST (PC2) predictors shows positive, but not significant, skill over this period. There is little variability in predictive skill by predictor with NAO index.

For the 50-yr assessment period, statistically significant skill is achieved using the two SST predictors and both \( T_{SP} \) indices. However, the May SST (SVD) predictor is significant only when calculated using NCEP–NCAR SST and not using HadSST. In contrast to the 1900–2001 period, it is the JJ \( T_{SP} \) rather than the MJJAS \( T_{SP} \) that is most skillful for all three NAO indices. The former has maximum \( r_s = 0.48 \) and MSSS = 22%, whereas for the MJJAS index the maximum skill values are \( r_s = 0.41 \) and MSSS = 16%. The CRUTEM2 \( T_{SP} \) indices show skill 20%–30% lower in this assessment period than the NCEP–NCAR \( T_{SP} \) indices. An ensemble mean \( T_{SP} \) index is calculated using the mean of the CRUTEM2 and NCEP–NCAR data. This produces skill levels close to those achieved using NCEP–NCAR alone. In terms of skill variation with NAO\textsubscript{DJF} index, all predictors except May SST (SVD) perform best against Z500\textsubscript{DJF} for this assessment period.

Highest hindcast skill in the 1972–2001 period is achieved using the JJ \( T_{SP} \) index and the two snow cover predictors. The JJ \( T_{SP} \) index shows the highest skill values (maximum \( r_s = 0.59 \) and MSSS = 35%) this skill showing little sensitivity to NAO\textsubscript{DJF} index or choice of temperature dataset. The JJ NH index is slightly less significant than the two snow cover predictors (maximum \( r_s = 0.53 \) and MSSS = 28%) against all three NAO\textsubscript{DJF} indices. As in the 50-yr assessment, the link between \( T_{SP} \) and the NAO\textsubscript{DJF} indices is stronger for the JJ period than for the long summer (MJJAS) period. The combined CRUTEM2–NCEP–NCAR mean JJ \( T_{SP} \) index shows skill slightly higher than that achieved with each single dataset JJ \( T_{SP} \) index. For MJJAS \( T_{SP} \) skill is similar to that achieved using single dataset indices for CRU and Hurrell NAO\textsubscript{DJF} but higher against Z500\textsubscript{DJF}. The A–O mean Eurasian snow cover produces significant skill against all three NAO indices, although it displays sensitivity to NAO\textsubscript{DJF} index. Neither of the SST-based predictors produce significant skill for the 30-yr assessment period (not shown).

b. Stationarity

The differences observed in NAO\textsubscript{DJF} predictive skill from summer \( T_{SP} \) between the 1900–2001 assessment period and the 1972–2001 period suggest that the relationship between \( T_{SP} \) and the NAO\textsubscript{DJF} is nonstationary.
### Table 3. Cross-validated hindcast skill for three assessment periods and three NAO indices. Dataset abbreviations: Had = HadISST, CRUT = CRUTEM2, NCEP = NCEP–NCAR reanalysis, NCEP–CRUT = combined CRUTEM2–NCEP–NCAR mean, Rut = Rutgers. Here $r$ is the Pearson correlation skill value and MSSS is the percentage improvement in mean-squared error over climatology. Here $p$ is the probability (values in brackets) that the observed skill was obtained by random chance. Boldface type denotes significant ($p < 0.05$) skill values as determined by a Monte Carlo resampling test.

<table>
<thead>
<tr>
<th>Period</th>
<th>Lagged predictor</th>
<th>Dataset</th>
<th>CRU NAO$_{DIFF}$</th>
<th>Hurrell NAO$_{DIFF}$</th>
<th>Z500$_{DIFF}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$r$ ($p$)</td>
<td>MSSS ($p$)</td>
<td>$r$ ($p$)</td>
</tr>
<tr>
<td>May SST (SVD)*</td>
<td>Had</td>
<td>0.23 (0.03)</td>
<td>4 (0.03)</td>
<td>0.27 (0.01)</td>
<td>6 (&lt;0.01)$</td>
</tr>
<tr>
<td>JJASO SST (PC2)</td>
<td>Had</td>
<td>0.13 (0.11)</td>
<td>1 (0.08)</td>
<td>0.15 (0.09)</td>
<td>1 (0.07)</td>
</tr>
<tr>
<td>1900–2001 JJ T$_{SP}$</td>
<td>CRUT</td>
<td>0.18 (0.05)</td>
<td>2 (0.04)</td>
<td>0.24 (0.02)</td>
<td>5 (0.02)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>CRUT</td>
<td>0.26 (0.01)</td>
<td>6 (0.01)</td>
<td>0.31 (&lt;0.01)</td>
<td>9 (&lt;0.01)</td>
</tr>
<tr>
<td>May SST (SVD)</td>
<td>Had</td>
<td>0.15 (0.15)</td>
<td>0 (—)</td>
<td>0.22 (0.07)</td>
<td>3 (0.08)</td>
</tr>
<tr>
<td>May SST (SVD)</td>
<td>NCEP</td>
<td>0.24 (0.05)</td>
<td>4 (0.07)</td>
<td>0.31 (0.02)</td>
<td>8 (0.03)</td>
</tr>
<tr>
<td>JJASO SST (PC2)</td>
<td>Had</td>
<td>0.24 (0.05)</td>
<td>5 (0.03)</td>
<td>0.25 (0.05)</td>
<td>6 (0.03)</td>
</tr>
<tr>
<td>JJASO SST (PC2)</td>
<td>NCEP</td>
<td>0.24 (0.05)</td>
<td>5 (0.03)</td>
<td>0.24 (0.05)</td>
<td>5 (0.03)</td>
</tr>
<tr>
<td>JJ T$_{SP}$</td>
<td>CRUT</td>
<td>0.33 (0.02)</td>
<td>10 (0.01)</td>
<td>0.27 (0.05)</td>
<td>6 (0.03)</td>
</tr>
<tr>
<td>1950–2001 JJ T$_{SP}$</td>
<td>NCEP</td>
<td>0.46 (-0.01)</td>
<td>21 (-0.01)</td>
<td>0.36 (0.02)</td>
<td>12 (0.01)</td>
</tr>
<tr>
<td>JJ T$_{SP}$</td>
<td>NCEP–CRUT</td>
<td>0.44 (-0.01)</td>
<td>19 (-0.01)</td>
<td>0.36 (0.02)</td>
<td>12 (0.01)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>CRUT</td>
<td>0.21 (0.10)</td>
<td>4 (0.07)</td>
<td>0.27 (0.06)</td>
<td>7 (0.04)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>NCEP</td>
<td>0.33 (0.02)</td>
<td>10 (0.01)</td>
<td>0.32 (0.04)</td>
<td>9 (0.02)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>NCEP–CRUT</td>
<td>0.29 (0.04)</td>
<td>8 (0.03)</td>
<td>0.33 (0.03)</td>
<td>10 (0.02)</td>
</tr>
<tr>
<td>JJ T$_{SP}$</td>
<td>CRUT</td>
<td>0.54 (-0.01)</td>
<td>29 (-0.01)</td>
<td>0.57 (-0.01)</td>
<td>32 (-0.01)</td>
</tr>
<tr>
<td>JJ T$_{SP}$</td>
<td>NCEP</td>
<td>0.58 (-0.01)</td>
<td>34 (-0.01)</td>
<td>0.56 (-0.01)</td>
<td>31 (-0.01)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>NCEP–CRUT</td>
<td>0.59 (-0.01)</td>
<td>35 (-0.01)</td>
<td>0.57 (-0.01)</td>
<td>33 (-0.01)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>CRUT</td>
<td>0.45 (0.03)</td>
<td>20 (0.02)</td>
<td>0.49 (0.02)</td>
<td>23 (0.01)</td>
</tr>
<tr>
<td>1972–2001 MJJAS T$_{SP}$</td>
<td>NCEP</td>
<td>0.51 (0.01)</td>
<td>26 (-0.01)</td>
<td>0.47 (0.01)</td>
<td>22 (0.01)</td>
</tr>
<tr>
<td>MJJAS T$_{SP}$</td>
<td>NCEP–CRUT</td>
<td>0.48 (0.02)</td>
<td>22 (0.01)</td>
<td>0.47 (0.02)</td>
<td>22 (0.01)</td>
</tr>
<tr>
<td>A–0 EU snow cover</td>
<td>Rut</td>
<td>0.47 (0.01)</td>
<td>21 (0.01)</td>
<td>0.36 (0.02)</td>
<td>11 (0.03)</td>
</tr>
<tr>
<td>JJ NH snow cover</td>
<td>Rut</td>
<td>0.53 (0.01)</td>
<td>28 (&lt;0.01)</td>
<td>0.51 (0.01)</td>
<td>24 (&lt;0.01)</td>
</tr>
</tbody>
</table>

* Data available 1900–98.
Therefore, the modest skill seen over the full 100-yr period could be drawn entirely from the 1972–2001 skillful period. To test this we first perform a rolling cross-correlation analysis for both the JJ and MJJAS TSP indices against the CRU NAO{\textsubscript{D\&F}} index for all possible 30-yr periods 1900–2001 (Fig. 4). A linear trend is removed from both time series prior to each correlation value being computed. The MJJAS TSP index peaks at the beginning and end of the twentieth century. By contrast, JJ TSP is most skillful since 1950. Further investigation reveals that it is the August–September period that contributes to the early 1900s peak in MJJAS (not shown). The strengthening correlations since 1950 may be linked to an increase in decadal NAO{\textsubscript{D\&F}} variability over the same period (Hurrell and van Loon 1997).

Despite the coincidence between periods of high correlation and periods of strongly positive NAO index in the early and late twentieth century (Fig. 1), the TSP indices perform equally as well predicting either above- or below-median NAO{\textsubscript{D\&F}} events. Hindcasts using MJJAS TSP are correctly above and below median in 54% of observed above- and below-median CRU NAO{\textsubscript{D\&F}} events 1900–2001. For the period 1950–2001 the JJ TSP index is correct in 64% of cases, while for 1972–2001 the figure is 71%. We therefore conclude that the hindcast skill from the summer TSP indices is nonstationary and that highest positive skill is observed during the early and late twentieth century.

c. Dataset dependence

Table 3 shows that, with the exception of May SST (SVD), the NAO{\textsubscript{D\&F}} hindcast skill exhibits little sensitivity to the choice of SST or 2-m air temperature dataset. Specifically, the JJASO SST (PC2) predictor produces similar skill against the three NAO{\textsubscript{D\&F}} indices whether it is calculated using HadISST or NCEP–NCAR data 1950–2001. The largest skill sensitivity to dataset is observed for the TSP indices calculated using CRUTEM2 and NCEP–NCAR data 1950–2001. Using CRUTEM2 data, the JJ TSP predictor explains 11% of the variance in the CRU NAO{\textsubscript{D\&F}} index compared with 21% using NCEP–NCAR data. A possible explanation for this difference is that, unlike the reanalysis data, the CRUTEM2 data are missing over the open ocean. The SG TSP region is more than 50% over ocean and has double the weighting of the EU and NA regions. Therefore, SG is likely to contribute the largest errors in TSP. However, only ~50% of JJ TSP skill comes from SG, which explains why CRUTEM2 produces some positive skill.

4. Summer TSP influence on upcoming NAO{\textsubscript{D\&F}}

a. Relationship between NAO{\textsubscript{D\&F}} lagged predictors

Quantifying the links between the different NAO{\textsubscript{D\&F}} lagged predictors will help to explore the physical basis for the observed lagged predictability. Cross correlations are performed for all predictors over the three assessment periods (not shown). For all periods, stable and significant (p < 0.05) relationships are found between the summer TSP indices and the summer/autumn (JJASO) SST (PC2) predictor. Since 1972, similar correlations are observed between TSP, JJASO SST (PC2) and the two warm season snow cover predictors. These results imply a potential physical relationship between the predictors both in time (summer–autumn) and in space (North America, Eurasia, to the North Atlantic sector).

The NAO{\textsubscript{D\&F}} exhibits persistence from one winter to the next (Johansson et al. 1998; Wanner et al. 2001). We therefore ask how much of the observed NAO{\textsubscript{D\&F}} hindcast skill results from the influence of the previous NAO{\textsubscript{D\&F}} on the predictors? Table 4 shows that only the Hurrell NAO{\textsubscript{D\&F}} 1900–2001 and Z{\textsubscript{500,D\&F}} 1950–2001 have a significant (detrended) lag-1 autocorrelation. The table also shows how much hindcast predictability the lag-1 NAO{\textsubscript{D\&F}} offers for predicting the May SST (SVD), JJASO SST (PC2), and JJ TSP time series. For the period 1900–2001, the skill values from the lag-1 NAO{\textsubscript{D\&F}} are comparable to those achieved using the predictors themselves to hindcast NAO{\textsubscript{D\&F}} (Table 3). For 1950–2001 and 1972–2001, with the exception of May SST (SVD), the predictors provide significantly higher NAO{\textsubscript{D\&F}} skill than does lag-1 NAO{\textsubscript{D\&F}}, based on all years of data. However, when composited on JJ TSP.
terciles, the lag-1 NAO$_{DJF}$ skill for JJ $T_{SP}$ 1950–2001 increases to $r_s \sim 0.5$. This suggests that when JJ $T_{SP}$ is large, it responds more to the NAO$_{DJF}$ signal from the previous winter. However, JJ $T_{SP}$ is a better NAO$_{DJF}$ predictor than lag-1 NAO$_{DJF}$ alone (Tables 3 and 4). This implies that the signal from the previous winter intensifies during the summer.

The results in Table 4 are similar using data with trends included (not shown). However, some sensitivity to the choice of dataset was found for the May SST (SVD) predictor 1950–2001. Using NCEP–NCAR re-analysis SST data, the lag-1 NAO$_{DJF}$ skill drops to $r_s = 0.3$ ($p > 0.10$) for both 1950–2001 and 1972–2001. The JJASO SST (PC2) predictor does not display such sensitivity, which suggests that greater differences exist for the May SST data than for the JJASO SST fields when comparing the NCEP–NCAR and HadISST datasets.

b. Physical basis for summer $T_{SP}$ influence on upcoming winter NAO$_{DJF}$

Seasonal atmospheric predictability in the extratropics is normally assumed to be lower than for the Tropics because of the presence of baroclinic instability. This appears to contradict the links we observe between summer $T_{SP}$, snow cover, and the upcoming winter climate. There is little persistence intrinsic to the extratropical atmosphere, which has a decorrelation time of $\sim 10$ days. Therefore, to enable a summer–winter link the memory of the summer atmospheric state must persist in a slowly varying boundary variable that can feed back onto the atmosphere at a later time.

Following Saunders et al. (2003), we propose the following mechanism for the influence of summer climate on that of the upcoming winter: summer (JJ) NH snow cover anomalies are negatively correlated with JJ subpolar near-surface air temperature ($T_{2m}$) over northern Siberia and northwest Canada and positively correlated with JJ $T_{2m}$ over Southern Greenland. The resulting subpolar temperature pattern induces a contemporaneous atmospheric response and lagged SST response centered on the North Atlantic. The atmospheric response is characterized by anomalies in MSLP and midlatitude zonal wind. The SST response is characterized by basin-scale anomalies and meridional gradients south and east of Newfoundland. Persistence of these SST gradients into autumn or early winter could force either a direct thermal NAO$_{DJF}$ response or initiate a response via a third variable.

There is good observational evidence to support the link between snow cover, $T_{SP}$ and North Atlantic SSTs. First, there is a significant correlation ($p < 0.01$) between changes in JJ NH snow cover and JJ $T_{SP}$ for the
period 1972–2001 (Saunders et al. 2003). Second, the atmosphere adjacent to the TSP centers is significantly perturbed during June and July. Figure 5 shows zonally averaged zonal wind anomalies as a function of height above the TSP regions and the North Atlantic in months before and during high minus low JJ TSP tercile years 1950–2001. Prior to June and July the only consistent signal for all three regions is located in the stratosphere at high latitudes during April. The signal in June and July over Eurasia and North America is characteristic of a strengthened polar vortex and a weakened midlatitude jet. A corresponding teleconnected signal is seen over the North Atlantic in June and July, which extends from the surface to the lower stratosphere. A band of significant anomalous westerlies is centered between 50° and 60°N, with anomalous easterlies to the north and south. This highlights that during extreme years of TSP, contemporaneous subpolar atmospheric teleconnections are formed between northern Eurasia, northern Canada, and the North Atlantic during June and July.

Our proposed mechanism is further supported by Fig. 6, which shows the hemispheric-scale anomaly patterns in height-averaged (925–200 hPa) zonal wind,
Fig. 6. Composite of (a) vertically averaged (925–200 hPa) North Atlantic sector zonal wind anomalies, (b) North Atlantic sector MSLP anomalies and (c) North Atlantic SST anomalies based on high minus low terciles of JJ T_sp index. Contour intervals are (a) 1 m s⁻¹, (b) 1 hPa, and (c) 0.2°C. Dashed contours denote negative anomalies and the zero contour is labeled. Shaded areas denote significance (p < 0.05) as determined by a Monte Carlo resampling test.
MSLP, and SST in months before, during, and following high-minus-low JJ $T_{SP}$ tercile years 1950–2001. Over the North Atlantic, the atmospheric signal strengthens in June and the SST strengthens after July. The zonal wind, MSLP, and SST signals appear causally connected, with the zonal wind anomalies collocated with gradients in MSLP anomalies and with SST anomalies. The atmospheric signal is prominent in June and July but dissipates thereafter, although it is apparent in the JJASO mean. In contrast, the SST anomaly pattern strengthens in July and persists until October. The timing, lag, and spatial pattern of the SST signal is consistent with its forcing by the anomalous atmospheric circulation associated with JJ $T_{SP}$. This conclusion is further supported by the significant ($p < 0.05$ for all time periods) link between JJ $T_{SP}$ and JJASO SST (PC2).

Figure 7 shows the mean anomalies in vertically averaged zonal wind, MSLP and SST for the winter seasons before and following high-minus-low JJ $T_{SP}$ terciles. The situation preceding JJ $T_{SP}$ shows a moderate positive NAO pattern in the zonal wind anomalies, which is further evidence that a proportion of the NAO$_{DJF}$ predictability from JJ $T_{SP}$ arises through interannual NAO persistence (see also section 4a). However, SSTs are close to climatology except near the equator, showing that the involvement of SST does not begin until after summer. The SST pattern following extreme JJ $T_{SP}$ events exhibits a strong meridional gradient off Newfoundland, which has persisted from summer. This pattern resembles the NAO-driven North Atlantic SST tripole anomaly pattern (e.g., Czaja and Frankignoul 2002).

Next, we attempt to show whether the evolving anomalies seen in Fig. 6 indicate a real response to JJ $T_{SP}$ or merely represent increased climatological variance in those months. This is quantified by plotting the percentage of grid cells in each month that are locally significant ($p < 0.05$). Figure 8 shows the percentages for zonal wind over the region ($30^\circ$–$80^\circ$N, $120^\circ$W–$40^\circ$E) and for SST over the region ($0^\circ$–$65^\circ$N, $100^\circ$W–$0^\circ$E). The horizontal lines denote approximate levels of field significance ($p < 0.05$) following the method of Livezey and Chen (1983). For each month, we assume 30 spatial degrees of freedom (DOF) in the zonal wind field and 15 DOF in the SST field. These values are appropriate because Livezey and Chen (1983) state that hemispheric atmospheric fields contain ~30–60 DOF. The actual DOF in our fields may oscillate around these estimates depending on the month. However, even with a conservative estimate for DOF of 15, the zonal wind and SST maxima are both field significant.
The zonal wind signal peaks in June and July, consistent with Fig. 6. The SST response develops 1 month later, peaking in July and August, and persisting until October. The timing of the lagged SST response is consistent with other observational studies of atmosphere–ocean interaction (e.g., Deser and Timlin 1997). By October, the majority of the SST signal comes from the subtropics. The SST response decreases to December, which is associated with a more variable atmospheric pattern August–October. We therefore conclude that the anomalous atmospheric circulation associated with JJ $T_{SP}$ is leading North Atlantic SST variability with a time lag of $\sim$1 month.

5. Discussion

NAO$_{DJF}$ hindcast skill is nonstationary during the twentieth century, implying that the predictive relationships vary in time. Data quality may also contribute to the variations in NAO$_{DJF}$ skill. During the early part of the record, observations are more sparse than during the 1972–2001 period of highest skill. This reduction in data quality may degrade NAO$_{DJF}$ hindcast skill and contribute to nonstationarity.

Nonstationarity in predictive relationships may arise from low-frequency multiannual or decadal oscillations in the predictor and/or predictand indices. An analysis of spectral coherency (Bloomfield 2000) was performed on the predictors and NAO indices to determine the dominant time scales of interaction. The results (not displayed) show that the preferred oscillatory period for each predictor with the NAO$_{DJF}$ differs depending on the assessment period. For example, the MJJAS $T_{SP}$ index and CRU NAO$_{DJF}$ index are significantly coherent at a period of 7 yr (1900–2001), 4 yr (1950–2001), and 8 yr (1972–2001). A further cause of the nonstationarity may be competing influences from SST forcings outside the North Atlantic, in particular, variations in the strength of ENSO (Sutton and Hodson 2003).

The May SST (SVD) predictor shows significant skill with HadISST data 1900–2001 and with NCEP–NCAR SST data 1950–2001. However, no significant skill is seen with HadISST over the latter period. Furthermore, our reported skill levels are lower than those found by Rodwell and Folland (2002), who employed an earlier version of HadISST 1948–98. Their original skill ($r_s = 0.45$) is achieved using data with trends included, which suggests that 30%–60% of that skill comes from linear trend 1950–2001 (cf. Table 1). Furthermore, this original skill is affected by serial correlation in the Z500$_{DJF}$ index. Using data with trends included, the lag-1 autocorrelation of the Z500$_{DJF}$ index rises from 0.29 to 0.42. We also note that despite optimizing covariance against the 500-hPa DJF North Atlantic geopotential height field, May SST (SVD) produces highest skill against the Hurrell NAO$_{DJF}$ index. This results because the Z500$_{DJF}$, CRU, and Hurrell NAO$_{DJF}$ indices cross correlate at $r > 0.85$ (1950–2001).

Significant SST persistence associated with JJ $T_{SP}$ is shown to occur from summer into October. The persistent SST signal located southeast of Newfoundland in Fig. 6 lies adjacent to the main region of North Atlantic cyclones, which marks the beginning of the North Atlantic storm track. Anomalous meridional SST gradients and associated turbulent heat, moisture, and momentum fluxes in this region could therefore initiate a direct thermal response in NAO$_{DJF}$ (Rodwell et al. 1999; Peng and Whittaker 1999; Peng et al. 2003; Cassou et al. 2004). However, the weakened SST signal from October to December (Fig. 8) suggests that the $T_{SP}$-induced SST anomalies alone may not be sufficient to force subsequent changes in NAO$_{DJF}$. One possibility is that autumn SST anomalies could affect the upward propagation of anomalous wave activity into the subpolar stratosphere. This would cause a downward-propagating response during early winter, which could influence the surface (Polvani and Waugh 2004). However, typically an AO, rather than an NAO$_{DJF}$, signal is observed in the winter stratosphere. Another theory is that autumn SST could influence either local sea ice outflow (Hilmer and Jung 2000; Kvämtso et al. 2004) and/or remote snow cover extent (Kitaev et al. 2002; Robock et al. 2003) through contemporaneous atmospheric circulation anomalies. These variables have sufficiently long time scales to feed back onto the atmosphere during the subsequent winter. An alternative possibility is that anomalous atmospheric flow could be the lone cause of summer NH snow cover anomalies, $T_{SP}$, and subsequent North Atlantic SST anomalies.

6. Summary and conclusions

In this study, a detailed assessment of the current levels of empirical seasonal hindcast skill for the North Atlantic Oscillation (NAO) is presented. A standardized hindcast procedure is employed to validate four previously published lagged predictors of the upcoming winter NAO. An additional predictor based on summer Northern Hemisphere subpolar 2-m air temperature ($T_{SP}$) is examined. Over three extended assessment periods out to 100-yr summer $T_{SP}$ is most skillful in predicting the upcoming winter NAO. For the period 1900–2001, May–September mean $T_{SP}$ offers the highest skill (a $\sim$6%–9% improvement over climatology). Since 1950, June–July mean $T_{SP}$ produces the highest skill (up to 22% improvement). Significant skill is also
observed 1900–2001 and 1950–2001 using patterns of May and late summer/autumn North Atlantic SST. Two warm season snow cover predictors also produce significant skill 1972–2001. Twentieth-century NAO$_{DJF}$ hindcast skill is nonstationary and highest since 1972 for all predictors except those derived using SSTs. This coincides with a period of large decadal NAO$_{DJF}$ variability. The highest skill from $T_{SP}$ also coincides with positive trends in the NAO$_{DJF}$ index during the early and late twentieth century. However, $T_{SP}$ performs equally well predicting above- or below-median NAO$_{DJF}$ events.

Evidence is presented supporting a physical link between summer $T_{SP}$ and the upcoming winter NAO. First, there is a significant contemporaneous association during summer between $T_{SP}$ and Northern Hemisphere snow cover extent. During subsequent months, the atmospheric response to $T_{SP}$ is centered on the mid-latitude North Atlantic. Circulation anomalies over the ocean are associated with a persistent pattern of North Atlantic SST into October. SST persistence is strong off Newfoundland, which coincides with the main region of North Atlantic cyclogenesis. This suggests that $T_{SP}$ either produces a direct NAO$_{DJF}$ response to local SST gradients or initiates a feedback onto NAO$_{DJF}$ from a third variable.

Further investigation using coupled dynamical models is required to fully understand the link between summer and winter North Atlantic climate. The large repository of global historical temperature observations allow for more extended analyses of the $T_{SP}$/NAO$_{DJF}$ mechanism than is possible with seasonal NAO$_{DJF}$ predictive links involving shorter time series (e.g., snow cover). Future studies must also examine whether the nonstationarity in this link is explained by observed variations in twentieth-century climate. Experiments using coupled GCMs with realistic snow–atmosphere interaction and subpolar teleconnections are required to further investigate the dynamics of how snow cover, $T_{SP}$, and SST link to NAO$_{DJF}$.

Our findings suggest that the NH subpolar regions may provide extended-range NAO$_{DJF}$ seasonal predictability in addition to the midlatitudes or the Tropics. This contrasts with recent thinking based on atmospheric GCM experiments, which indicate that variations in tropical SSTs are of primary importance for explaining the NAO$_{DJF}$ trend 1950–2000 (Hurrell et al. 2004). We believe that these results offer exciting new possibilities for the future of seasonal forecasting in the extratropics.

Acknowledgments. CGF was supported by a Research Studentship from the U.K. Natural Environment Research Council. We thank B. Lloyd-Hughes for helpful discussions. We acknowledge NOAA–CERES Climate Diagnostics Center, Boulder, Colorado, for NCEP–NCAR global reanalysis project data; the Snow Data Resource Center at Rutgers University for snow extent records; the Met Office Hadley Centre for HadISST and HadSLP data; the Climatic Research Unit, University of East Anglia, for CRUTEM2 and NAO data; and the Climate Analysis Section, NCAR, Boulder, Colorado, for Hurrell (1995) NAO data.

REFERENCES


Johansson, A., A. Barnston, S. Saha, and H. Van den Dool, 1998:


