Seasonal Prediction from Arctic Sea Surface Temperatures: Opportunities and Pitfalls

ERIK W. KOLSTAD
Uni Research Climate, Bjerknes Center for Climate Research, Bergen, Norway

MARIUS ÅRHUN
Geophysical Institute, University of Bergen, and Bjerknes Center for Climate Research, Bergen, Norway

(Manuscript received 11 January 2018, in final form 24 July 2018)

ABSTRACT

Arctic sea ice extent and sea surface temperature (SST) anomalies have been shown to be skillful predictors of weather anomalies in the midlatitudes on the seasonal time scale. In particular, below-normal sea ice extent in the Barents Sea in fall has sometimes preceded cold winters in parts of Eurasia. Here we explore the potential for predicting seasonal surface air temperature (SAT) anomalies in Europe from seasonal SST anomalies in the Nordic seas throughout the year. First, we show that fall SST anomalies not just in the Barents Sea but also in the Norwegian Sea have the potential to predict wintertime SAT anomalies in Europe. Norwegian Sea SST anomalies in spring are also significant predictors of European SAT anomalies in summer. Second, we demonstrate that the potential for prediction is sensitive to trends in the data. In particular, the lagged correlation between Norwegian Sea SST anomalies in spring and European SAT anomalies in summer is considerably higher for raw data than linearly detrended data, largely due to warming SST and SAT trends in recent decades. Third, we show that the potential for prediction has not been stationary in time. One key result is that, according to two twentieth-century reanalyses, the strength of the negative lagged correlation between Barents Sea SST anomalies in fall and European SAT anomalies in winter after 1979 is unprecedented since 1900.

1. Introduction

Seasonal weather predictions are helpful tools for risk mitigation, and can guide more efficient use of resources in various sectors of society (Palmer 2002; Troccoli 2010; Bruno Soares and Dessai 2015; De Felice et al. 2015; Clark et al. 2017). In recent years, progress in forecasting the state of the North Atlantic Oscillation (NAO), the dominant pattern of low-frequency atmospheric variability in the North Atlantic (Walker 1924; Barnston and Livezey 1987), has been documented (Scaife et al. 2014; Dunstone et al. 2016; Wang et al. 2017). Better NAO forecasts translate to more accurate predictions of variables such as pressure, temperature, and precipitation (Hurrell 1995; Athanasiadis et al. 2017), as well as storminess (Feser et al. 2015). The physical mechanisms behind the enhanced NAO forecast skill have often been shown to originate in the tropics. For instance, Cassou (2008) showed that the Madden–Julian oscillation (MJO) controls part of the distribution and sequence of the NAO on subseasonal time scales. On longer time scales, tropical rainfall anomalies have been shown to explain a large share of predicted year-to-year variations of the wintertime NAO (Scaife et al. 2017).

Climate variables outside the tropics have also been shown to influence the NAO (Smith et al. 2016, and references therein). The potential for using Arctic sea ice and sea surface temperature (SST) anomalies to predict subsequent surface air temperature (SAT) anomalies in the midlatitudes has been known for a long time, as summarized by Herman and Johnson (1978). For instance, correlations between SAT in northern Europe and sea ice extent in the Nordic seas—the marginal seas of the northeast Atlantic (see Fig. 1d for geographical references)—were identified as far back as the early 1900s, and Arctic sea ice and SST anomalies impart significant skill to modern-era empirical forecast models designed to predict climate variables in the midlatitudes (Koenigk et al. 2016; Hall et al. 2017; Wang et al. 2017). Moreover, it has been shown that proper
initialization of Arctic sea ice and SST anomalies could be important for the prediction skill in dynamical seasonal forecast models (Day et al. 2014; Scaife et al. 2014; MacLachlan et al. 2015).

It has proven useful to separate the influences of SST and sea ice anomalies on other climatic variables into dynamical and thermodynamical components. The dynamical response to a changing Arctic sea ice cover has been widely studied in recent years (e.g., Peings and Magnusdottir 2014; Vihma 2014; Barnes and Screen 2015). For the Barents and Kara Seas specifically, the transient large-scale atmospheric response to sea ice extent anomalies has often resembled the NAO pressure anomaly pattern (e.g., Deser et al. 2007; Kim et al. 2014; Nakamura et al. 2015; King et al. 2016; Screen 2017a). In particular, below-normal late summer or fall sea ice extent in the Barents and Kara Seas has been linked to negative NAO conditions and cold anomalies in Eurasia during winter (Honda et al. 2009; Petoukhov and Semenov 2010; Mori et al. 2014). At the same time, it is important to be aware of the thermodynamical effects of reduced sea ice; model results by Screen (2017a) indicate that negative NAO conditions in years with below-average sea ice cover in the Arctic are more strongly NAO-negative but give less cooling over northern Europe than negative NAO conditions in years with above-average sea ice cover.

We believe that although the latest dynamical prediction systems have started to show useful skill, there is still a need for simple empirical studies for extending our understanding of the pathways for this skill. As Folland et al. (2012) studied potential predictability of northern

![Figure 1](https://example.com/figure1.png)

**Fig. 1.** The colors show statistically significant (at the 5% level) $R^2_s$ values for each oceanic grid point, area averaged for the Europe land region (see outline in Fig. 2). (a),(b) SST in OND predicting SAT in JFM, based on raw and linearly detrended data, respectively. (c),(d) As in (a) and (b), respectively, but for SST in MAM predicting SAT in JJA. The two oceanic regions used later in the analysis are outlined in black, and the names of the marginal seas are given in (d).
European winter climate one season ahead to “develop ideas to act as a benchmark for improving the performance of dynamical climate models” (p. 801), we aim here to contribute to the ongoing debate about the influence of the Arctic on the midlatitudes by quantifying the potential for predicting European SAT anomalies from Arctic SST anomalies one season earlier. We do this by addressing three questions. First, we ask where and in which seasons European SAT anomalies can be empirically predicted from Arctic SST anomalies. We consider the whole year, not just the influence of summer/fall anomalies on winter.

Second, we know that the recent decades’ ocean heat transport into the Arctic and the concurrent decline in sea ice appear to be unprecedented in recent centuries (Kinnard et al. 2011; Spielhagen et al. 2011; Onarheim and Årthun 2017). European SAT has increased rapidly during the same period (Jones et al. 2013). If two variables exhibit positive trends that are caused by the same external mechanism (such as increased radiative forcing), the correlation between them will increase. If the variables are not linked by any physical mechanism, they may still be correlated just because of their concurrent trends. But if the variables are linked by a physical mechanism, the inflated correlation can make it seem as if that mechanism is more important than it really is. And conversely, if two variables are negatively correlated due to some physical mechanism, concurrent positive trends in the two variables can make it seem as if that mechanism is less important than it really is (because the trends make the correlation between the two less negative). These issues are often taken into account by making use of linearly detrended data (e.g., Wang et al. 2017). Our question is this: To what extent does the general warming trends of both SAT and SST influence the lagged correlations between SST and SAT? This is assessed by considering how sensitive the potential for prediction is to linear detrending.

Third, several studies have suggested that regional predictability varies with time. One point is that the calculated skill of dynamical models is dependent on the length of the analysis period (Shi et al. 2015). Another is that there are nonstationarities in the relationship between atmospheric circulation patterns and SAT in Europe on decadal time scales (Slonosky et al. 2001). For example, the NAO appears to be more predictable by dynamical models in recent decades than earlier in the twentieth century (Müller et al. 2005). Similar results have been obtained with hindcast experiments with prescribed SST (O’Reilly et al. 2017; Weisheimer et al. 2017). Ogawa et al. (2018) recently claimed that there is no significant link between reduced Arctic sea ice and Eurasian temperatures, suggesting that the observed trends in Eurasian temperatures and atmospheric circulation have been due to internal atmospheric variability. We ask, Has the predictive skill of European SAT from Arctic SST been stationary in time, or has it varied during the twentieth century?

The remainder of the paper is structured as follows. In section 2, the data and methods are described. Section 3 presents our results in five subsections. The geographical distribution of the potential for using Arctic SST anomalies to predict European SAT anomalies is explored in section 3a. The annual cycles of the potential for prediction are investigated for two oceanic regions in section 3b. In section 3c, the role of trends in influencing the potential for prediction is studied. The stationarity of the potential for prediction is assessed in in section 3d, and in section 3e we examine to which degree fall SST could have been used to predict SAT in central England in the period after 1900 and what role the NAO has played in mediating this relationship. We wrap up with a summary and a discussion of the results in section 4.

2. Data and methods

a. Data

The first part of the analysis is based on ERA-Interim data (Dee et al. 2011) for the period from January 1979 to May 2017, from which we used SAT (i.e., 2-m temperature), and SST (which was set to the freezing temperature under sea ice). We always computed SAT anomalies for season A [e.g., October–December or (OND)] and SST anomalies for season B, which followed directly after season A [e.g., January–March (JFM)]. Although we had 38 years of data, we only used 37-yr time series of these variables, as some seasons required data from the end of the years from the start of the subsequent years (this is why we used data up to May 2017). Note also that all time series were standardized (i.e., we subtracted the interannual mean and divided by the interannual standard deviation). This is important because trends and standard deviations were usually larger for SAT than for SST.

For the stationarity analysis we used data from the coupled ECMWF reanalysis of twentieth-century data (CERA-20C), the first fully coupled atmosphere–ocean reanalysis of the twentieth century (Buizza et al. 2018; Laloyaux et al. 2018), covering the period 1901–2010. After CERA-20C was produced, it was discovered that the model does not melt enough sea ice during summer, and that this led to a gradual accumulation of Arctic sea ice. But as the SSTs were relaxed toward the HadISST2 dataset (Titchner and Rayner 2014), the sea ice extent was kept under control. Data from 10 CERA-20C ensemble members are available, and we used the ensemble
mean. Note that we also performed the analysis using the atmospheric hindcast ERA-20C (Poli et al. 2016), and this gave qualitatively similar results.

We also used a wintertime NAO index computed as the standardized pressure anomaly difference between Lisbon, Portugal, and Stykkisholmur, Iceland, provided by the Climate Analysis Section at NCAR in Boulder, Colorado (Hurrell et al. 2003), as well as the aggregated series of Central England Temperature (CET; Parker et al. 1992; Parker and Horton 2005).

**b. Methods**

Three geographical regions were defined: the Norwegian Sea (westernmost outlines in Fig. 1), the Barents Sea (easternmost outlines in Fig. 1), and Europe (outlines in Fig. 2). The boundaries of the regions were chosen manually, but our results were not particularly sensitive to these subjective choices. These regions were used for area-weighted spatial averaging (just “area averaging” henceforth). In the simple case, we took the average of a variable inside a region and computed interannual time series for that variable, but we also performed slightly more complicated operations. For instance, we calculated the correlation coefficients $R$ between SST in season A in each oceanic grid point and SAT in season B in each grid point in Europe. We then computed the area-averaged $R$ value over all the land points. The resulting value, for which we obtained one for each oceanic grid point, is denoted as $R_E$, where the subscript $E$ indicates that we averaged over Europe. Similarly, we computed area-averaged $R$ values inside each of the two oceanic regions in turn. The area-averaged $R$ values for the Barents Sea are named $R_B$, and we obtained one such value for each land grid point; $R_N$ represents area-averaged values for the Norwegian Sea. To aggregate this information further, we also computed the area-averaged $R$ value inside the oceanic regions, and we call these $R_{BE}$ for the Barents Sea and $R_{NE}$ for the Norwegian Sea.

To avoid cancellations between positive and negative $R$ values when area-averaging, we also computed the area-averaged $R^2$ values as $R^2_E$, $R^2_B$, $R^2_N$, $R^2_{BE}$, and $R^2_{NE}$. These are our metrics of the potential for prediction.

To assess the statistical significance of the area-averaged $R^2$ values, we used a bootstrapping technique. The null hypothesis was that the $R^2$ values could have been produced using random data. For each combination of grid points, we created a set of $n = 1000$ artificial $R^2$ values by letting all the SAT time series consist of 37 data points picked randomly from the original time series (with replacement). If the actual area-averaged $R^2$ value was larger than the 95th percentile of the $n$ bootstrap samples, we rejected the null hypothesis and deemed it statistically significant at the 5% level.

We also used bootstrapping to compute confidence intervals for area-averaged $R^2$ values by computing $n = 1000$ artificial $R^2$ values for each combination of grid points. Each sample was calculated by drawing 37 random data points from the original time series (with replacement), but now the pairing in time between each SST and SAT data point was maintained. The 10th and 90th percentiles of the $n$ bootstrap samples then made up the bounds of the 80% confidence intervals for the area-averaged $R^2$ values.

To assess the role of the NAO in season B (hereafter NAO$_B$) in mediating the total effect of SST in season A.
(SST_A) on SAT in season B (SAT_B), we performed a statistical mediation analysis (Fritz and MacKinnon 2007), using the methods described by Kolstad (2017). Three regression models, following the notation in Fritz and MacKinnon (2007), were used:

\[
\hat{\text{SAT}}_B = \hat{\tau}_{\text{SST}}_A + \hat{\xi}_1, \tag{1}
\]

\[
\hat{\text{SAT}}_B = \hat{\tau}_{\text{SST}}_A + \hat{\beta}_{\text{NAO}} + \hat{\xi}_2, \tag{2}
\]

\[
\hat{\text{NAO}} = \hat{\alpha}_{\text{SST}}_A + \hat{\xi}_3. \tag{3}
\]

The carets indicate estimated parameters, and \(\hat{\xi}_1, \hat{\xi}_2, \text{and} \hat{\xi}_3\) are the intercepts. Full mediation, which means that the entire effect of SST_A on SAT_B was mediated by NAO_B, only occurred if four steps (or conditions) were satisfied. First, the total effect of SST_A on SAT_B (\(\hat{\tau}\)) must be significant at a certain significance level (here 5%). Otherwise, there would have been no effect to mediate. Second, the effect of SST_A on NAO_B (\(\hat{\alpha}\)) must be significant for the mediation to be meaningful. Third, the effect of NAO_B on SAT_B controlled for SST_A (\(\hat{\beta}\)) must be significant to ensure that SAT_B did not become conditionally independent of NAO_B given SST_A (which would have rendered the pathway through NAO_B invalid). Fourth, and crucially, \(\hat{\tau}\) must be nonsignificant. This means that SAT_B became conditionally independent of SST_A when SAT_B was included in the regression. The product \(\hat{\alpha}\hat{\beta}\) is known as the mediated (or indirect) effect of SST_A on SAT_B through NAO_B. Partial mediation, meaning that some (but not all) of the effect of SST_A on SAT_B was mediated by NAO_B, occurred if the fourth step is replaced by the following condition: \(|\hat{\tau}| < |\hat{\tau}|\).

3. Results

a. Geographical distributions

To assess the potential for using SST anomalies in season A to predict SAT anomalies in season B, we computed \(R_E^2\) (see section 2), which for each oceanic grid point consists of its average \(R^2\) value with each land point in the European region outlined in Fig. 2a (hence the subscript E). The calculations were done using both raw and (linearly) detrended data. The role of trends in dictating the differences between raw and detrended data will be discussed in a separate subsection below.

The \(R_E^2\) values shown in Fig. 1a indicate that SST anomalies in the northern Barents Sea in OND had a high potential for predicting European SAT anomalies in JFM. (Henceforth we will use the following shorthand for SST in season A predicting SAT in season B: A \(\rightarrow\) B, and it will become apparent below why the season pairs OND \(\rightarrow\) JFM and March–May (MAM) \(\rightarrow\) June–August (JJA) were chosen for scrutiny in Fig. 1.) Assuming that below-normal sea ice is linked to above-normal SST, the high \(R_E^2\) values for OND \(\rightarrow\) JFM are in qualitative agreement with the previously cited studies that found associations between below-normal sea ice cover in the Barents Sea and subsequent below-normal winter SAT in parts of Europe. Figure 1a also shows that SST anomalies in parts of the Greenland Sea explained some of the SAT variance in Europe during the ERA-Interim period. Figure 1b shows that when linearly detrended data were used, the geographical patterns and the magnitude of \(R_E^2\) did not change appreciably in the Barents Sea. In the Norwegian Sea, the patterns of \(R_E^2\) did change when detrended data were used, indicating a sensitivity to trends in SST, SAT, or both.

In Fig. 1c, \(R_E^2\) is shown for MAM \(\rightarrow\) JJA. High values are on display along the warm, northbound Norwegian Atlantic Current (Orvik and Skagseth 2005), but Fig. 1d reveals that these high values vanished when detrended data were used. In other words, as for Norwegian Sea SST in OND predicting European SAT in JFM, there were discrepancies between raw and detrended data. Figures 1c and 1d also show that \(R_E^2\) in the Barents Sea was quite low for MAM \(\rightarrow\) JJA, both for raw and detrended data.

We now study the geographical patterns of predictability in Europe by exploring \(R_E^2\) and \(R_N^2\), where the subscripts B and N indicate that each European land grid point’s \(R^2\) values with each oceanic grid point were area-averaged in the Barents Sea and Norwegian Sea, respectively. For OND \(\rightarrow\) JFM, \(R_B^2\) was highest in the North Sea region and in parts of eastern Europe for both raw and detrended data (Figs. 2a,b). As the North Sea is where the North Atlantic storm track typically connects with the Eurasian continent in winter, this geographical pattern suggests that the NAO may have been a mediator of the strong relationship between SST in the Barents Sea and SAT in this region.

Figure 2c shows \(R_N^2\) for MAM \(\rightarrow\) JJA, based on raw ERA-Interim data. The highest values occurred in the southeastern part of the European reference region. For detrended data (Fig. 2d), \(R_N^2\) was considerably lower, as expected from the large differences between detrended and nondetrended data already seen in Figs. 1c and 1d. These results indicate that one should take great care when using SST from the Norwegian Sea in spring to predict European SAT in summer. The potential for prediction during the ERA-Interim period was substantially higher for raw than detrended data.

b. Annual cycles

We now show \(R_{BE}, R_{BE}^2, R_{NE}, \text{and} R_{NE}^2\) for all combinations of one season predicting its subsequent season,
starting with SON → DJF, continuing with OND → JFM, and so forth.

Figure 3a shows $R_{BE}^2$, calculated using both raw (blue bars) and detrended (orange bars) data. (Note that for OND → JFM, the $R_{BE}^2$ values are the area-weighted averages of the values shown in Figs. 1a and 1b.) The $R_{BE}$ values (multiplied by 100) are listed above the bars. The highest $R_{BE}^2$ values occurred for SST anomalies in OND predicting SAT anomalies in JFM for both raw and detrended data (which is why that season pair was shown in the previous figures). Note that $R_{BE}$ was higher for detrended than raw data (see next subsection). All the season pairs between August–October (ASO) / November–January (NDJ) and NDJ / February–April (FMA), as well as January–March (JFM) / April–June (AMJ), had significant $R_{BE}^2$ values for both raw and detrended data. During the rest of the year, there was no significant potential for prediction of European SAT from Barents Sea SST.

In Fig. 3b, $R_{NE}$ and $R_{NE}^2$ are shown. The highest $R_{NE}^2$ value based on raw data occurred for MAM → JJA, with an $R_{NE}$ value of 0.25. As already indicated by Figs. 1c,d and 2c,d, $R_{NE}^2$ was substantially lower and not statistically significant for detrended data, and $R_{NE}$ was only 0.06 for detrended data. Most season pairs throughout the year exhibited higher $R_{NE}^2$ values for raw than detrended data (see next subsection). The values were significant for both detrended and raw data for the following season pairs: OND → JFM, NDJ → FMA, DJF → MAM, and FMA → May–July (MJJ).

c. The role of trends

The previous sections showed that $R^2$ was generally sensitive to detrending, and especially so for Norwegian Sea SST from winter to early summer predicting European SAT one season later. To investigate the reasons for this, we now show the trends of standardized and area-averaged SST in the two oceanic regions and SAT in Europe during the ERA-Interim period in Fig. 4. The trends are given in standard deviation units.

We recall that $R_{BE}^2$ was larger for detrended than raw data in OND → JFM (see Fig. 3a), and that $R_{BE}$ was negative. Figure 4 shows that the trends of both Barents Sea SST in OND (orange bars) and European SAT in JFM (magenta bars) were positive during the ERA-Interim period. The negative correlation coefficient $R_{BE}$ is the key to understanding why $R_{BE}^2$ was smaller for raw data than for detrended data. As mentioned, when two variables are positively correlated and have positive trends, the $R^2$ value the between the two increases. But when the two variables are negatively correlated, $R^2$ decreases if both have positive trends. We observe that $R_{BE}^2$ was also larger for detrended than raw data for the next season pair (NDJ → FMA; see Fig. 3a), at the same time that $R_{BE} < 0$ and the trends of both SST and
SAT were positive (Fig. 4). For the preceding season pair (NDJ→FMA), $R_{BE}^2$ was slightly larger for raw than detrended data (Fig. 3a), probably because the SST trend was weakly negative (Fig. 4). For the MAM→JJA season pair (see Fig. 3b), $R_{NE}^2$ was much higher for raw than detrended data. Figure 4 shows that the trend for area-averaged standardized SST in the Norwegian Sea during MAM was 0.67 per decade (orange bars) and 0.55 per decade for SAT in Europe during JJA (magenta bars). We recall that $R_{NE}^2$ for this season pair was 0.25 for raw data and 0.06 for detrended data. It appears that the strong sensitivity to detrending of $R_{NE}^2$ in the Norwegian Sea was a result of the combination of positive SST and SAT trends and positive $R_{NE}$ values.

d. Stationarity

The third question raised in section 1 was whether the potential for prediction was stationary during the twentieth century. We now use the coupled CERA-20C to quantify how much $R_{BE}^2$ and $R_{NE}^2$ fluctuated for the season pairs already investigated in Figs. 1 and 2. The period before 1979 is based on sparse data in the Arctic, but the coupling between ocean, ice, and atmosphere in

![Fig. 4. ERA-Interim trends of standardized area-averaged SST in the Barents Sea (blue bars), SST in the Norwegian Sea (orange bars), and SAT in Europe (magenta bars), all in standard deviation units per decade. The labels along the x axis indicate that the SST trends are for one season before the SAT trends. For instance, the first set of three bars show SST trends during SON and SAT trends during DJF.](image)

![Fig. 5. For each possible 37-yr period in the CERA-20C period, (a) $R_{BE}^2$ and (b) $R_{BE}$ are shown as circles for Barents Sea SST in OND predicting European SAT in JFM. The corresponding ERA-Interim values are shown as stars. In (a), the upper and lower bounds of the 80% confidence interval of ERA-Interim $R_{BE}^2$ are shown as horizontal lines, and the threshold for statistical significance at the 5% level is indicated with a dashed line. The colors denote whether the correlations were calculated using raw (blue) or detrended (orange) data, and the numbers along the x axes indicate the first year of each time series.](image)
CERA-20C inspires confidence in the physical consistency of the data.

Figure 5a shows $R^2_{BE}$ for OND→JFM during each possible period with the same length as the ERA-Interim period, the first starting in 1901 and the last in 1972, using both raw (blue) and detrended (orange) data. The CERA-20C $R^2_{BE}$ values based on detrended data were significant early in the century and during periods starting after around 1950. The raw data–based $R^2_{BE}$ values were never clearly higher than the threshold for significance. The ERA-Interim based $R^2_{BE}$ values based on both raw and detrended ERA-Interim data were higher than all the corresponding CERA-20C-based values. Only for a few periods starting around 1960 did $R^2_{BE}$ computed from CERA-20C fall inside the 80% confidence interval of ERA-Interim $R^2_{NE}$ are shown as stars. In (a), the upper and lower bounds of the 80% confidence interval of ERA-Interim $R^2_{NE}$ are shown as horizontal lines, and the threshold for statistical significance at the 5% level is indicated with a dashed line. The colors denote whether the correlations were calculated using raw (blue) or detrended (orange) data, and the numbers along the x axes indicate the first year of each time series.

CERA-20C shows $R^2_{BE}$ for OND→JFM during each possible period with the same length as the ERA-Interim period, the first starting in 1901 and the last in 1972, using both raw (blue) and detrended (orange) data. The CERA-20C $R^2_{BE}$ values based on detrended data were significant early in the century and during periods starting after around 1950. The raw data–based $R^2_{BE}$ values were never clearly higher than the threshold for significance. The ERA-Interim based $R^2_{BE}$ values based on both raw and detrended ERA-Interim data were higher than all the corresponding CERA-20C-based values. Only for a few periods starting around 1960 did $R^2_{BE}$ computed from CERA-20C fall inside the 80% confidence interval of ERA-Interim $R^2_{NE}$ are shown as stars. In (a), the upper and lower bounds of the 80% confidence interval of ERA-Interim $R^2_{NE}$ are shown as horizontal lines, and the threshold for statistical significance at the 5% level is indicated with a dashed line. The colors denote whether the correlations were calculated using raw (blue) or detrended (orange) data, and the numbers along the x axes indicate the first year of each time series.

In Fig. 6a, $R^2_{NE}$ for MAM→JJA is shown. The values based on raw CERA-20C data were significant except for periods starting from the early 1920s to 1940, but only near the end of the period did CERA-20C yield $R^2_{NE}$ values inside the ERA-Interim based 80% confidence interval. The value of $R^2_{NE}$ from detrended CERA-20C data was generally lower than that from raw data and was consistently significant only during periods starting after the end of the 1940s. It is important to note that the actual ERA-Interim $R^2_{NE}$ value fell beneath the lower boundary of its own 80% confidence interval. In other words, the actual value was lower than what would normally be expected, perhaps as a result of sensitivity to sample size (too few years available). This, along with the strong positive SST and SAT trends during the ERA-Interim period (see section 3c), helps explain the extreme sensitivity of $R^2_{NE}$ to detrending for the MAM→JJA season pair identified earlier (see Figs. 1c,d, 2c,d, and 3b). Figure 6b shows $R_{NE}$ for MAM→JJA. The values computed using raw CERA-20C data were highest in the early and latest periods, and $R_{NE}$ based on detrended data was increasing steadily throughout the twentieth century.

From the results presented in Figs. 5 and 6, it is clear that the potential for predictability of European SAT based on Arctic SST has not been stationary with respect to time. Not only did the values of $R^2_{BE}$ and $R^2_{NE}$ fluctuate during the
In the twentieth century, but their sensitivity to detrending also varied. To study this sensitivity in more detail, the standardized trends of area-averaged SST and SAT during all the periods shown in Figs. 5 and 6 are presented in Fig. 7.

Figure 7a shows that the periods when $R^2_{BE}$ and $R_{BE}$ were most sensitive to detrending (see Fig. 5) coincided with periods of high trends of Barents Sea SST during the early and late CERA-20C periods. Figure 7b tells a similar story about $R^2_{NE}$ and $R_{NE}$; the sensitivity to detrending (see Fig. 6) was largest when the trends of Norwegian Sea SST were at their highest levels. This also occurred early and late in the twentieth century.

e. An application to Central England Temperature

Figures 2a and 2b indicated that SAT in JFM in a region encompassing the British Isles could be skillfully predicted from SST in the Barents Sea in OND during the ERA-Interim period. To what degree does this relationship hold for the observational time series of CET? First, we computed the correlation between ERA-Interim SST during OND in each oceanic grid point in the Nordic seas and CET during JFM. These are shown for raw data in Fig. 8a and for detrended data in Fig. 8b. Significant negative correlations occurred in the northern Barents Sea, as well as in some parts of the Norwegian Sea.

A time series of standardized area-averaged SST anomalies inside the Barents Sea region in OND was calculated from the raw data. This is shown in blue in Fig. 8c. The correlation between this time series and the time series of standardized CET anomalies during JFM (shown multiplied by $-1$ in orange) was $-0.44$. The detrended time series of area-averaged SST and CET had a correlation of $-0.52$, as shown in Fig. 8d. If we had omitted the last year from the time series (SST in OND 2016 predicting SAT in JFM 2017), the correlation coefficients would have been $-0.58$ for raw data and $-0.61$ for detrended data. This illustrates again how volatile the correlations during short time periods can be (see also the fluctuations of $R^2_{BE}$ and $R_{BE}$ in Fig. 5).

Lagged correlations between area-averaged Barents Sea SST anomalies in OND and CET anomalies in JFM are shown for each possible 37-yr period since 1901 in Fig. 9, using SST from CERA-20C and ERA-Interim data. The results are shown for both raw data in blue and for detrended data in orange. The ERA-Interim values ($-0.44$ and $-0.52$, respectively) are shown as stars. The absolute values of the ERA-Interim correlations are unprecedented, as they never exceeded 0.4 during the CERA-20C periods. In fact, the correlations were only statistically significant for a few of those periods. We also note that, as was found for the larger European reference region (see Fig. 5b), the signs of the correlations changed over time, from positive values for the

![Fig. 7.](http://journals.ametsoc.org/images/jcli/article-pdf/31/20/8197/4697845/jcli-d-18-0016_1.pdf)
periods starting early in the twentieth century to the present-day negative values.

As mentioned, several studies have shown links between reduced sea ice in the Barents Sea in fall and negative NAO conditions in the subsequent winter. By following the steps described in section 2b and using ERA-Interim data for the period 1979–2017, it can be shown that the NAO in JFM fully mediated the total effect of Barents Sea SST anomalies in OND on CET anomalies in JFM. Using the parameters defined in Eqs. (1)–(3), the first step was satisfied because \( \hat{\tau} = -0.44 \) (\( p = 0.006 \)), the second step because \( \hat{\alpha} = -0.35 \) (\( p = 0.03 \)), the third step
anomalies in JFM. The mediated effect was \( \hat{\alpha} \hat{\beta} = -0.19 \). However, we note that the low \( p \) value for \( \hat{\tau} \) demonstrated that this result was sensitive to the choice of significance level. If a significance level of 10% had been used, the NAO would only have satisfied the criteria for partial mediation (since \( \hat{\tau} \) would not have been non-significant but \( |\hat{\tau}| < |\tau| \)). Using detrended ERA-Interim data with a significance level of 5%, the first three conditions for mediation were satisfied, but the NAO was a partial mediator because the \( p \) value for \( \hat{\tau} \) was 0.01. The mediated effect for detrended data was the same as for raw data (\( \hat{\alpha} \hat{\beta} = -0.19 \)). Going back in time, Fig. 9 showed that the total effect of Barents Sea SST anomalies in OND on CET anomalies in JFM was only significant during a few of the CERA-20C periods. Even during those periods the NAO in JFM did not satisfy the criteria for mediation (neither full nor partial).

4. Summary and discussion

We sought answers to three questions to investigate the potential for using Arctic SST to predict European SAT, and to highlight some challenges associated with empirical prediction models that assume stationarity. First, on the issue of the geographical distribution and annual cycles of the potential for prediction, we found that according to ERA-Interim, the highest potential predictability was obtained when SST in the Barents Sea region in October–December (OND) predicted European SAT in January–March (JFM). Qualitatively, this is in agreement with earlier studies (e.g., Honda et al. 2009; Petoukhov and Semenov 2010; Mori et al. 2014). A statistical mediation analysis revealed that the total effect of Barents Sea SST anomalies in OND on Central England Temperature (CET) anomalies in JFM was fully mediated by NAO anomalies in JFM during the ERA-Interim period. This suggests the following causal pathway for the lagged correlation between SST anomalies in fall and CET anomalies in winter: Barents Sea SST anomalies in OND caused NAO anomalies in JFM, which then again caused CET anomalies during the same season. We did not seek a physical mechanism for the first part of this pathway, but we note that several recent model-based studies have indicated that the polar stratosphere may have acted as a “bridge” between Barents (and Kara) Sea sea ice anomalies in fall and NAO anomalies during subsequent winters (e.g., Sun et al. 2015; Screen 2017b; Zhang et al. 2018). However, we emphasize that the mediation by the NAO was only valid for ERA-Interim data. None of the CERA-20C periods gave indications of a mediating role of the NAO. An in-depth analysis of the physical mechanisms that potentially linked Arctic SST and sea ice anomalies to subsequent midlatitude SAT anomalies through the NAO (and other atmospheric circulation anomalies) is suggested for future work.

We also identified a considerable potential for prediction of European SAT in spring and summer based on SST anomalies in the Norwegian Sea one season before, but this potential was only apparent when raw data were used. When we used detrended data, there was no clearly significant lagged correlation between SST in the Norwegian Sea and European SAT during these seasons.

Our second question touched on the influence of trends on the predictive skill. In some cases, we found that our metric for predictive skill changed considerably after the data had been detrended. This indicates that what appeared to be a significant potential for prediction when nondetrended data were used may have been artificial if there were strong trends in both the predictor and the predictand. Positive trends of both SAT in Europe (Jones et al. 2013) in summer and SST along the Norwegian Atlantic Current (Spielhagen et al. 2011) in spring have transpired over the last few decades, and those trends have contributed to the sensitivity to detrending. Another sensitivity to detrending was also revealed. For SST in the Barents Sea in late autumn predicting European SAT in late winter, our metric for potential predictive skill was higher for detrended than raw data. The reason is that the lagged correlations between the two variables were predominantly negative, while the trends of both variables were positive. These examples demonstrate that it is important to understand how trends influence the correlations between variables before those correlations are used to predict the future.

In answer to our third question, we found that the strength of the links between Arctic SST and subsequent European SAT was nonstationary with respect to time. There are at least four possible explanations for this. First, the length of the periods considered here (37 years) may be too short to obtain robust relationships. When just one year was taken out of our analysis of the lagged relationship between Barents Sea SST anomalies in fall and CET anomalies in winter, the negative correlation between the two increased from \(-0.44\) to \(-0.58\) (using raw data). Second, the fluctuations in potential skill could be linked to the aforementioned strong trends in oceanic heat transport and SST in the Nordic seas and SAT in Europe. Third, our results indicated that the physical mechanisms that have led to lagged correlations between SST and SAT have changed over time. In the
case of CET, a statistical mediation analysis revealed that the lagged relationship between Barents Sea SST in fall and CET in winter was mediated by NAO anomalies during the ERA-Interim period (1979–2017). The mediating role of the NAO is in agreement with process-based studies (Deser et al. 2007; Jaiser et al. 2012; Dunstone et al. 2016; King et al. 2016). However, according to CERA-20C, the lagged correlation between Barents Sea SST anomalies in fall and European SAT anomalies changed from being positive early in the twentieth century to negative values for periods starting around the middle of the twentieth century. The same was found for CET anomalies, and the NAO was not a mediator between Barents Sea SST and CET during any of the CERA-20C periods studied here. The implication is that while the proposed physical mechanisms may have been valid during recent decades, it is unlikely that these were operative earlier in the twentieth century. Fourth, the differences between $R_{L5}$ in CERA-20C and ERA-Interim could be due to issues with the sea ice coupling in CERA-20C, for which an excessive accumulation of Arctic sea ice has been identified (Laloyaux et al. 2018). However, relaxation toward the HadISST2 (Titchner and Rayner 2014) dataset should have neutralized this problem with regard to our analysis. But it should be noted that HadISST2 has more sea ice than some other re- constructions, and it also has a comparatively weak de-

REFERENCES

Atha, P. J., and Coauthors, 2017: A multisystem view of wint-


Acknowledgments. The authors thank three anonymous reviewers, the editor, and Erica Madonna and Camille Li for constructive comments. Funding for E.W.K. was provided by the Research Council of Norway (RCN) through the Seasonal Forecasting Engine project (Grant 270733) and Blue-Action, funded by the European Union’s Horizon 2020 research and innovation program (Grant 727852). M.A. was funded by the RCN project PATHWAY (Grant 263223) and Blue-Action. The European Centre for Medium-Range Weather Forecasts (ECMWF) provided the ERA-Interim, ERA-20C, and CERA-20C datasets.


