AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL073736

Key Points:
- PNA and NAO hindcast skill show marked variability over the twentieth century, following the varying link between tropical Pacific SST and the PNA
- Variability in PNA and NAO skill is linked on interannual time scales, notably during the mid-century when the hindcast skill drops
- PNA- is less predictable, arising more from atmospheric variability; PNA+ is more clearly forced by warm tropical Pacific SST anomalies

Supporting Information:
- Supporting Information S1

Correspondence to:
C. H. O'Reilly,
christopher.oreilly@physics.ox.ac.uk

Citation:
O'Reilly, C. H., J. Heatley, D. MacLeod, A. Weisheimer, T. N. Palmer, N. Schaller, and T. Woollings (2017), Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century, *Geophys. Res. Lett.*, *44*, 5729–5738, doi:10.1002/2017GL073736.

Received 23 DEC 2016
Accepted 21 MAY 2017
Accepted article online 24 MAY 2017
Published online 12 JUN 2017

Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century

Christopher H. O'Reilly[1], James Heatley[1], Dave MacLeod[1], Antje Weisheimer[1,2], Tim N. Palmer[1], Nathalie Schaller[3], and Tim Woollings[1]

[1] Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK, [2]European Centre for Medium-Range Weather Forecasts, Reading, UK, [3]Center for International Climate and Environmental Research, Oslo, Norway

Abstract Seasonal hindcast experiments, using prescribed sea surface temperatures (SSTs), are analyzed for Northern Hemisphere winters from 1900 to 2010. Ensemble mean Pacific/North American index (PNA) skill varies dramatically, dropping toward zero during the mid-twentieth century, with similar variability in North Atlantic Oscillation (NAO) hindcast skill. The PNA skill closely follows the correlation between the observed PNA index and tropical Pacific SST anomalies. During the mid-century period the PNA and NAO hindcast errors are closely related. The drop in PNA predictability is due to mid-century negative PNA events, which were not forced in a predictable manner by tropical Pacific SST anomalies. Overall, negative PNA events are less predictable and seem likely to arise more from internal atmospheric variability than positive PNA events. Our results suggest that seasonal forecasting systems assessed over the recent 30 year period may be less skillful in periods, such as the mid-twentieth century, with relatively weak forcing from tropical Pacific SST anomalies.

1. Introduction

Forecasts of the wintertime extratropical circulation over the Northern Hemisphere have historically exhibited low levels of skill. The North Atlantic Oscillation (NAO), which is the dominant mode of variability over the Euro-Atlantic sector [*Marshall et al.*, 2001], has proven particularly challenging for seasonal forecast models to predict. Analysis of NAO hindcasts (or reforecasts) in a previous generation of models, used in the DEMETER project [*Palmer et al.*, 2004], found that none of the models exhibited skill over the full 42 year hindcast period [*Müller et al.*, 2005; *Shi et al.*, 2015]. However, over the final 14 year DEMETER hindcast period several models did exhibit significantly skillful NAO hindcasts, highlighting the possibility that forecast skill may depend on the period of the hindcasts. The latest generation of seasonal forecast models has demonstrated improved and significant hindcast skill for the NAO [*Scaife et al.*, 2014] and the Arctic Oscillation [*Stockdale et al.*, 2015; *Riddle et al.*, 2013]. The Met Office's decadal prediction system has recently been shown to exhibit significantly skillful hindcasts of the NAO at lead times of over a year, in large part through accurately capturing tropical Pacific variability [*Dunstone et al.*, 2016]. However, skillful hindcasts of extratropical wintertime circulation in these models has only been demonstrated for the most recent ~30 year period (i.e., from 1980). Therefore, it is important to investigate whether the hindcast skill of the recent period holds going further back in the twentieth century.

A recent study by *Weisheimer et al.* [2017] presented seasonal hindcast experiments for all boreal winter seasons between 1900 and 2010, initialized at the beginning of November, using a state-of-the-art seasonal forecast model (albeit using prescribed sea surface temperatures (SSTs)). In these hindcasts, variability in both ensemble mean and probabilistic NAO skill measures over 30 year periods was demonstrated. The NAO hindcasts in this system have comparable skill for the early and late twentieth century periods, but there is an apparent decline in skill during a period in the mid-century. Although the skill of the NAO in the most recent period of this hindcast ($r \approx 0.4$) is lower than in some other seasonal forecasting systems (e.g., $r \approx 0.6$) [*Scaife et al.*, 2014], the results of *Weisheimer et al.* [2017] nonetheless suggest that some periods may indeed be more/less predictable than others. The reasons for the variable levels of skill, however, are not clear. In this study we use the hindcast experiments presented in *Weisheimer et al.* [2017] and analyze the variable predictability of the Pacific/North American index (PNA) and the NAO over the twentieth century. To investigate

the source of the variable predictability, we analyze the observed link between the extratropical circulation and the tropical Pacific, which is the dominant source of skill in wintertime seasonal forecasting [*Smith et al.*, 2012].

2. Methods

2.1. Twentieth Century Hindcasts

In this study we analyze the hindcast experiments presented by *Weisheimer et al.* [2017]. The simulations were performed using the European Centre for Medium-Range Weather Forecasting (ECMWF) Integrated Forecast System [*Molteni et al.*, 2011], model cycle 41r1 (a newer model cycle than that currently used for ECMWF's operational seasonal forecasting system, System 4). The simulations were performed at T255 horizontal resolution (∼80 km grid spacing), with 91 vertical levels (up to 0.01 hPa). The hindcasts were initialized with 51 ensemble members on 1 November and run through winter to the end of February, for each year from 1900 to 2009. The model was initialized using data from ERA-20C [*Poli et al.*, 2016], which is a reanalysis product produced by assimilating only surface pressure and marine wind observations over the period 1900–2010. The forecast model also includes explicit stochastic physics schemes, which represent the uncertainties of parameterized physical processes [*Weisheimer et al.*, 2014]. The stochastic parameterizations are responsible for generating the spread between ensemble members in these experiments, as initial condition perturbations are not used. The forecast model also includes time-varying greenhouse gas, solar cycle, and volcanic aerosol forcing [*Molteni et al.*, 2011], which helps to improve the simulation of trends over the hindcast period [*Weisheimer et al.*, 2017].

The simulations were performed with sea surface temperature (SST) and sea ice boundary conditions from the HadISST 2.1 data set [*Rayner et al.*, 2003] applied throughout each simulation. Since SSTs are prescribed, rather than modeled, these simulations are somewhat idealized and approximate the potential predictability under the assumption that SSTs force the atmosphere on seasonal time scales, particularly in the tropics, and are perfectly predicted. However, seasonal forecast models run with a coupled ocean model typically do exhibit high levels of skill in simulating SSTs in the tropics [e.g., *Weisheimer et al.*, 2009; *Kim et al.*, 2012; *Barnston et al.*, 2012; *Xue et al.*, 2013; *Scaife et al.*, 2014].

2.2. PNA and NAO Indices

The PNA is one of the dominant teleconnection patterns in the extratropics during the Northern Hemisphere winter season [*Wallace and Gutzler*, 1981] and is strongly associated with regional temperature and precipitation anomalies across North America [*Leathers et al.*, 1991]. Here we use the grid point definition of the PNA as defined in *Wallace and Gutzler* [1981] as PNA = $\frac{1}{4}[Z'(20°\text{N}, 160°\text{W}) - Z'(45°\text{N}, 165°\text{W}) + Z'(55°\text{N}, 115°\text{W}) - Z'(30°\text{N}, 85°\text{W})]$, where Z' is the normalized 500 hPa geopotential height anomaly, averaged over the DJF winter season. The PNA indices are calculated in this manner for ERA-20C, hereafter called the reference PNA index, and for each ensemble member in the hindcast. The reference PNA is normalized by its standard deviation, and the PNA indices for each ensemble member are also normalized using the standard deviation of the reference PNA. The reference PNA index calculated from the ERA-20C data set is very similar to PNA indices calculated from other data sets, using either a pointwise or rotated empirical orthogonal function (EOF) method (Figure S1 in the supporting information).

The NAO indices analyzed in this study are the same as those used in *Weisheimer et al.* [2017]. The reference NAO index is calculated as the principle component time series of the leading EOF of the wintertime (i.e., DJF mean) area-weighted 500 hPa geopotential height anomalies over the Atlantic sector (90°E–30°W, 30°E–90°W). As with the PNA, the reference NAO index is calculated from the ERA-20C data set and is in very good agreement with other estimates of the index [*Weisheimer et al.*, 2017]. The NAO indices for each ensemble member are calculated by projecting the 500 hPa geopotential height anomalies onto the EOF pattern calculated for the reference NAO. The reference NAO index is normalized to have unit standard deviation, and the NAO indices for each ensemble member are also normalized using the standard deviation of the reference NAO.

2.3. Confidence Intervals

The confidence intervals displayed were calculated using a bootstrapping with replacement method. For example, to estimate the confidence intervals for a correlation over a 30 year window, 30 years are selected at random from the 30 year window, with replacement, and used to produce a pair of dummy 30 year time series.

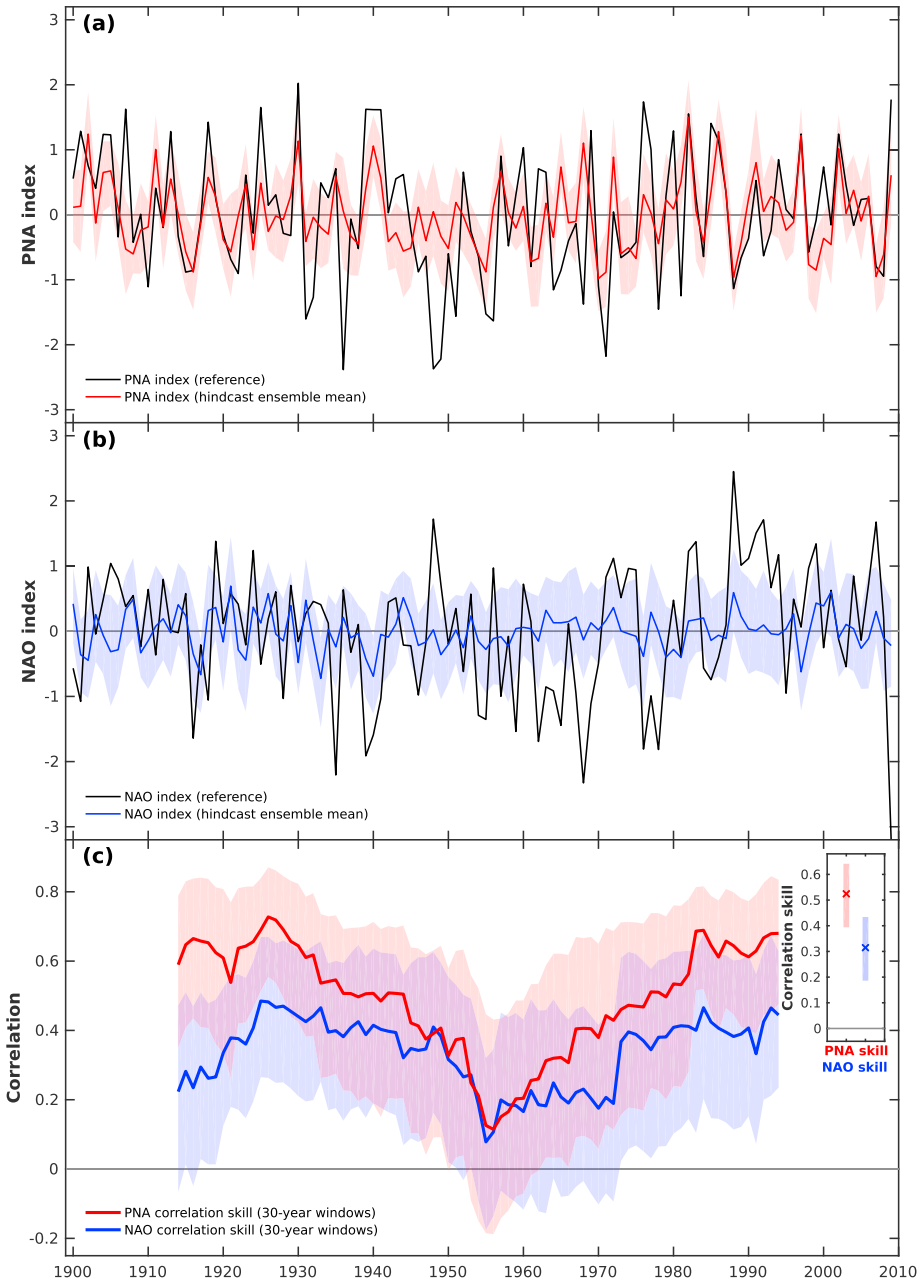

Figure 1. (a) The reference PNA index (black) and the ensemble mean PNA from the hindcast (red). (b) The reference NAO index (black) and the ensemble mean NAO from the hindcast (blue). In Figures 1a and 1b the shading indicates the interquartile spread of the ensemble members. (c) The correlation skill of the hindcast ensemble mean with the reference indices over moving 30 year windows for the PNA (red) and NAO (blue). The shading around each curve in Figure 1c represents the 5–95% uncertainty of the correlation, calculated using a bootstrapping method. Also shown in Figure 1c is the PNA and NAO correlation skill over the full hindcast period, along with the 5–95% uncertainty range.

The correlation between the dummy time series is saved and the process repeated 10,000 times. This distribution of 10,000 values is then used to estimate the sampling uncertainty of the correlation coefficient.

3. Results

3.1. Variability of PNA and NAO Hindcast Skill

We first consider the ensemble mean correlation of the PNA and NAO over the hindcast period. The time series of the reference PNA and NAO indices are plotted in black in Figures 1a and 1b, respectively. The ensemble mean PNA and NAO hindcasts for each year are also shown. The correlation between the ensemble mean

hindcasts and the reference index is shown in Figure 1c. While both the PNA and NAO are significantly skillfully forecast over the entire 110 year hindcast period, overall the PNA ($r = 0.52$) is significantly more skillfully forecast than the NAO ($r = 0.31$). Higher skill for the PNA is consistent with other seasonal forecast systems [e.g., *Scaife et al.*, 2014] as the phase of the PNA is strongly influenced by SST anomalies (and associated heating anomalies) in the tropical Pacific [*Horel and Wallace*, 1981; *Wallace et al.*, 1992; *Trenberth et al.*, 1998].

The analysis of *Weisheimer et al.* [2017] highlighted potential variability in skill when measured over shorter periods such as those used to assess the current generation of seasonal forecast models. To investigate the variability of skill, we analyze the ensemble mean correlation skill over moving 30 year windows, shown in Figure 1c. NAO ensemble mean skill exhibits a drop in the mid-century, but this variability is relatively modest compared to the confidence intervals. The variability of the ensemble mean PNA skill varies more than the NAO, with high levels of skill in the early and late parts of the hindcast period. The PNA skill also exhibits a mid-century minimum where it approaches zero, closely matching the mid-century minimum in NAO skill, suggesting that the drop in skill in the two indices during this period is related.

3.2. Variable Link Between the PNA and the Tropical Pacific

To investigate the source of the variability in PNA hindcast skill, we analyze the variability of the SSTs in the tropical Pacific. The time series of SST anomaly averaged over the Niño-3 region (defined as 150°–90°W, 5°S–5°N) is shown in Figure 2a. The Niño-3 SSTs are significantly positively correlated with the reference PNA index ($r = 0.47$), which has similarly been highlighted in previous studies [e.g., *Horel and Wallace*, 1981]. To assess the variability in the relationship between the PNA and the tropical Pacific SSTs, the correlation between the reference PNA index and the Niño-3 index over moving 30 year windows is plotted in Figure 2b (green line); this variability is also seen in PNA indices using different reanalysis data sets (Figure S2). The variability of the correlation between the Niño-3 index and the reference PNA is almost identical to the variability in the PNA hindcast skill (red line). The Niño-3/PNA relationship very clearly captures the mid-century minimum in PNA hindcast skill. This suggests that the lack of PNA hindcast skill during the mid-century period is due to a weakening of the relationship between tropical SSTs and the extratropical circulation during this period. The hindcast ensemble mean PNA response to the Niño-3 SST anomalies is consistently high ($r > 0.8$) throughout the hindcast period (shown by orange line in Figure 2b). This high correlation is a result of averaging over many ensemble members, as the correlation between all the individual ensemble members (i.e., $N = 51$ members $\times$ 110 years = 5610) and the Niño-3 index over the full hindcast period is much smaller ($r = 0.49$, shown in Figure 2c) and is comparable with the reference PNA index ($r = 0.47$, also shown in Figure 2c). Since the model PNA responds the same way to SST anomalies in the tropical Pacific throughout the hindcast period, the failure of the model to skillfully forecast the PNA during the mid-century is due to the weakening of the observed relationship between the tropical Pacific SST and the PNA, which the model does not capture.

Similar analysis of the reference NAO index and the Niño-3 index indicates that the variability in the NAO hindcast skill is not clearly related to a variable relationship with the tropical Pacific (Figure S4). The Niño-3 index is only very weakly negatively correlated with the reference NAO and model NAO over the full hindcast period (Figure 2c). In fact, the NAO seems to be slightly more strongly linked to the PNA index than the Niño-3 index in both the reference data set and the model, which further suggests that the hindcast skill of the PNA and NAO indices could be linked.

3.3. The Link Between PNA and NAO Hindcast Skill

To probe the link between PNA and NAO skill over the hindcast period, we first calculate a measure of the hindcast error. For each year, we fit Gaussian distributions to the 51 PNA and NAO hindcasts from the ensemble members using the mean and standard deviation of the hindcasts. We then calculate how different the reference PNA and NAO values are from the respective ensemble mean hindcast, in units of standard deviation, $|\sigma|$. The $|\sigma|$ for the PNA and NAO hindcasts are plotted for each year in Figure 3a. In the decades during the middle of the hindcast period there are a larger number of high $|\sigma|$ events for both the PNA and NAO, consistent with the dropoff in correlation skill (i.e., Figure 1c).

The correlation between the PNA and NAO $|\sigma|$ values over moving 30 year windows is shown in Figure 3b. The PNA and NAO errors are most strongly correlated during the mid-century period when the PNA and NAO hindcast skill is the lowest (Figure 1c). The extended mid-century period when the hindcast error of the PNA and NAO are most closely linked, based on Figure 3b, seems to be the 45 year period between 1936 and 1980 (though the results are not qualitatively different if this period is moved by a few years in either direction). Scatterplots of the PNA and NAO $|\sigma|$ values during this 1936–1980 period (Figure 3c) and all other years

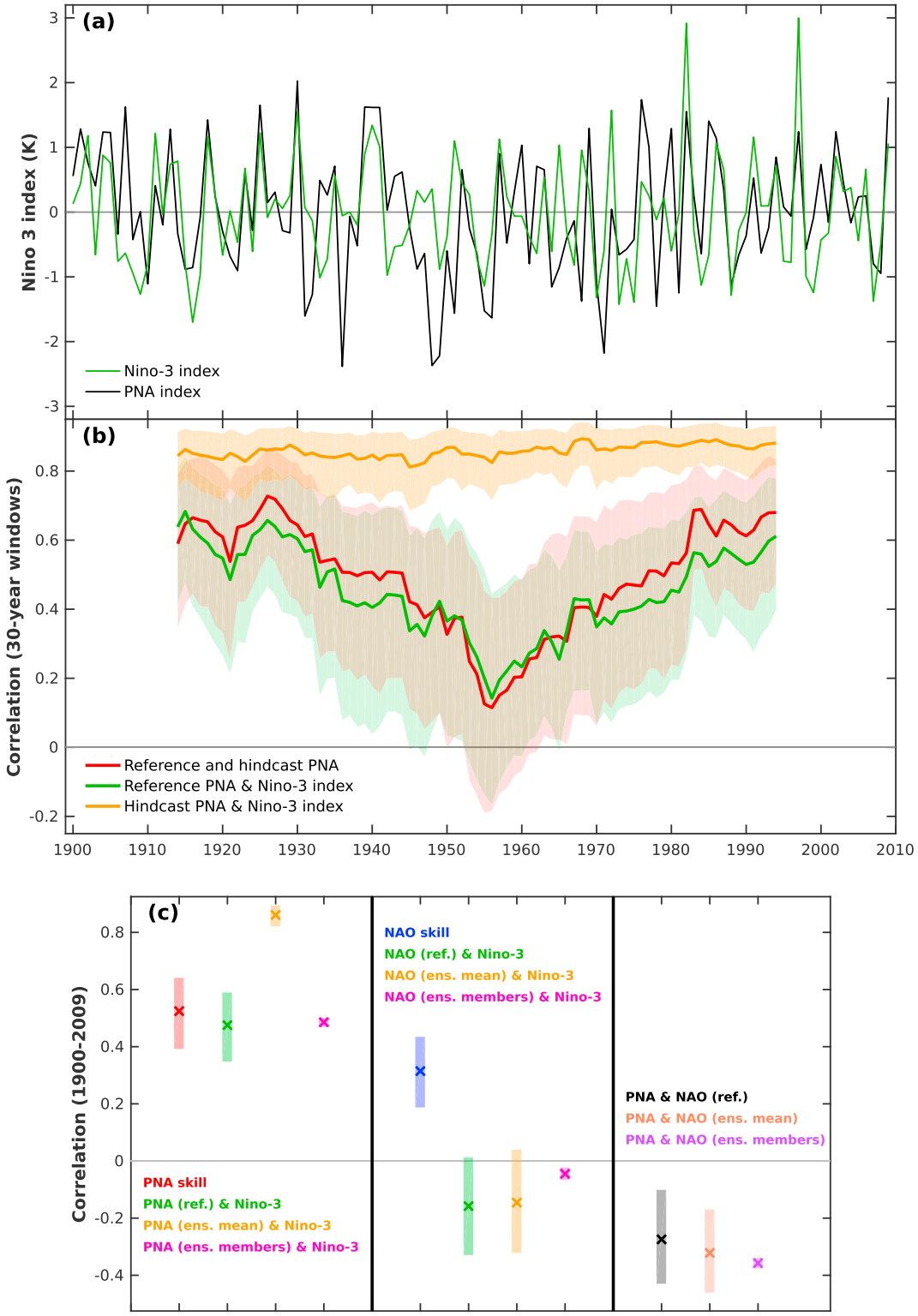

Figure 2. (a) The boreal winter Niño-3 index and reference PNA index over the hindcast period. (b) Correlation between the Niño-3 index (green) and the hindcast ensemble mean PNA (red, as in Figure 1c) and the reference PNA index over moving 30 year windows. Also plotted in Figure 2b is the correlation between the Niño-3 index and the hindcast ensemble mean PNA (orange). (c) Correlation values over the whole hindcast period: the left section shows the values for the PNA hindcast skill and the PNA correlation with the Niño-3 index in observations and the model; the middle section is the same as the left section but for the NAO; and the right section shows the correlation between the PNA and the NAO in the reference and the model. The shading in Figures 2b and 2c represents the 5–95% confidence intervals.

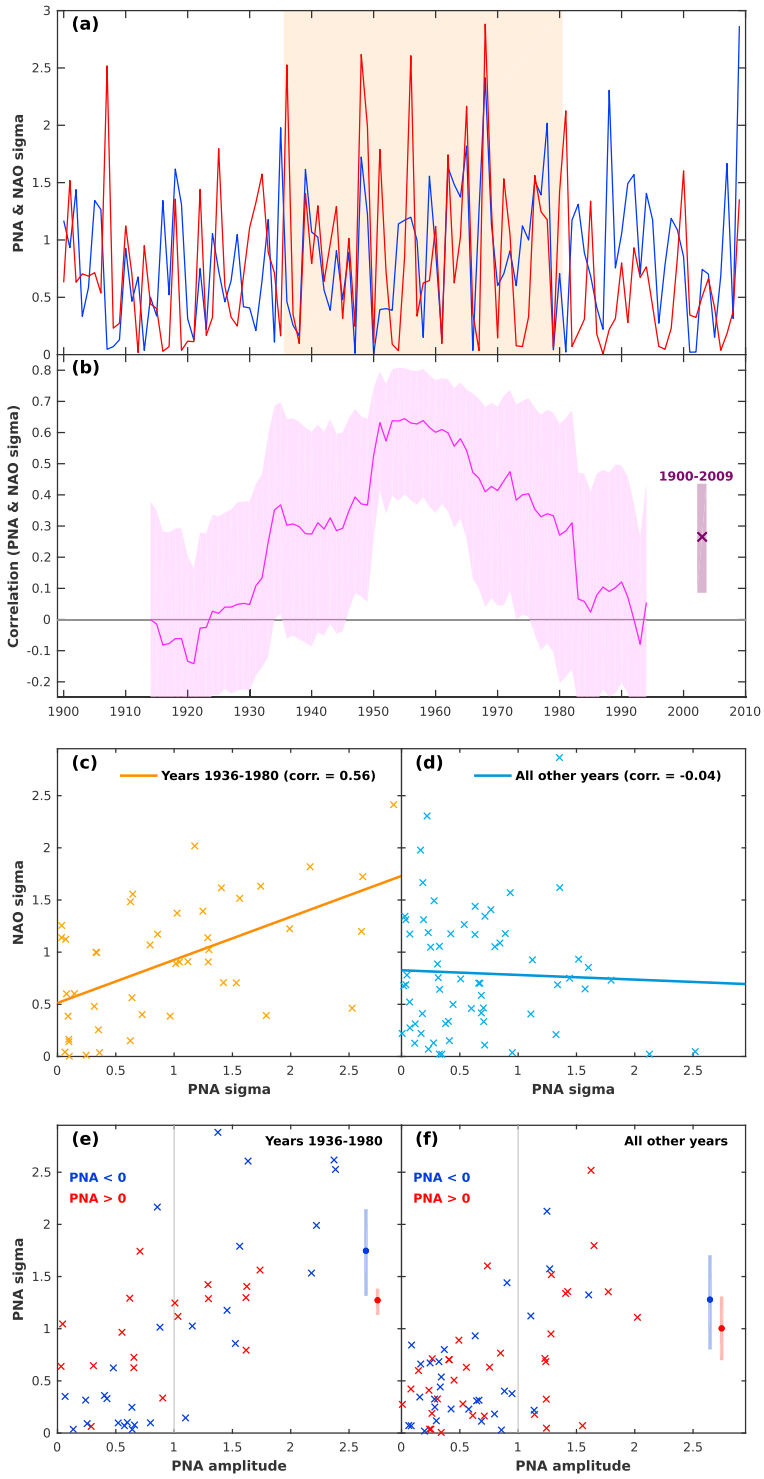

Figure 3. (a) Distance of the reference PNA (red) and NAO (blue) from the hindcast ensemble mean, in $|\sigma|$ (i.e., standard deviation), calculated from a normal distribution fitted to the ensemble members. (b) Correlation between the $|\sigma|$ values of the PNA and NAO hindcasts over moving 30 year windows, shading indicates the 5–95% confidence intervals (the correlation over the full hindcast period is also shown). The mid-century period where the PNA and NAO $|\sigma|$ values are closely related, 1936–1980, is shaded in Figure 3a. Scatterplots of the PNA and NAO $|\sigma|$ values and lines of best fit are shown (c) for years 1936–1980 and (d) for all other years. Scatterplots of the PNA $|\sigma|$ values are plotted against the reference PNA index (e) for years 1936–1980 and (d) for all other years. The filled circles on the right side of Figures 3e and 3f show the mean PNA $|\sigma|$ values for the positive (red) and negative (blue) reference PNA events greater than unit amplitude (with 5–95% confidence intervals).

(Figure 3d) demonstrate that the hindcast errors of the PNA and NAO are strongly linked during the mid-century. However, the hindcast errors are only weakly related in the earlier and later periods, when the hindcast exhibits the highest levels of skill. We also analyzed the development of PNA and NAO hindcast error on subseasonal time scales, to see if the PNA error leads or follows NAO error. However, the hindcast skill degrades when analyzed over shorter periods (particularly for the NAO) and the results were inconclusive. Nonetheless, it does seem that the NAO might be more strongly related to the PNA than the tropical Pacific SSTs, although the difference in not clearly significant (Figure 2c). This suggests that the errors in the PNA region are possibly influencing the NAO forecasts farther downstream on subseasonal time scales, through the downstream influence of the Pacific circulation on the Atlantic highlighted in previous studies [e.g., *Honda and Nakamura*, 2001; *Pinto et al.*, 2011; *Drouard et al.*, 2015].

During the mid-century period when the PNA and NAO skill drops, there are a number of years with particularly large PNA hindcast errors (Figure 3c). During the 1936–1980 period, the largest PNA $|\sigma|$ years are almost all strongly negative PNA events (Figure 3e). In the other years there are fewer large PNA $|\sigma|$ years and the association between the phase of the PNA and the hindcast error is not so pronounced (Figure 3f). Over the whole period the hindcast error of PNA events greater than unit amplitude are significantly larger for negative PNA events compared to positive PNA events. In contrast, the NAO hindcast errors exhibit no significant phase dependence (not shown).

3.4. The Asymmetry of PNA Hindcast Skill and Forcing by the Tropical Pacific

The majority of the worst PNA hindcasts identified in Figure 3 were negative PNA years in the reference (i.e., Figures 3e and 3f), indicating that negative PNA years are not as well forecast as positive PNA events. PNA phase has been found to influence predictability in medium range forecasts, with forecasts initialized in the negative phase of the PNA being less skillful [e.g., *Palmer*, 1988; *Lin and Derome*, 1996; *Sheng*, 2002]. To further examine the phase dependence of seasonal PNA hindcast skill, we calculated the relative operating characteristic (ROC) skill score for different threshold PNA events. The ROC curve provides a measure of the probabilistic forecast skill of a binary event [e.g., *Buizza and Palmer*, 1998; *Mason and Graham*, 1999; *Kharin and Zwiers*, 2003], which in this case is defined as the PNA exceeding (or not) a particular threshold. The ROC skill score (defined as double the area under the ROC curve, minus one) is positive for skillful forecasts, equal to one for perfect forecasts and negative for forecasts that are less skillful than climatology; it is plotted in Figure 4 for various PNA thresholds. Over the hindcast period it is clear that positive PNA years are generally more skillfully forecast, which is particularly true for relatively high amplitude PNA years. For example, the hindcasts of upper quintile (i.e., eightieth percentile) PNA events are significantly more skillful than lower quintile (i.e., twentieth percentile) PNA events (at the 5% significance level, using the bootstrapping with replacement method).

Composite SST anomalies for the upper and lower quintile PNA events are shown in Figure 4. The upper quintile PNA events are strongly associated with warm SST anomalies in the Tropical East Pacific, indicating a clear link to El Niño winters. The lower quintile PNA events, however, exhibit a much weaker association with cold tropical Pacific SST anomalies. This indicates an asymmetrical response of the PNA to tropical Pacific SST anomalies (and associated heating/precipitation anomalies, Figure S5), similar to that highlighted in previous studies [e.g., *Hoerling et al.*, 1997; *Straus and Shukla*, 2002; *Peng and Kumar*, 2005]. The atmosphere exhibits a relatively strong, predictable response to positive SST anomalies in the eastern tropical Pacific and as a result, the positive PNA events are more predictable. On the other hand, negative PNA events are less strongly constrained by SST anomalies in the tropical Pacific, which is consistent with the idealized modeling study by *Abid et al.* [2015]. Given that there are significant midlatitude SST anomalies in the negative PNA composites (Figure 4d), one might consider whether these could have a role in forcing the PNA. To first order, however, the midlatitude SST anomalies associated with the PNA are primarily forced by the atmosphere through anomalous turbulent heat fluxes of opposite sign, associated with the PNA circulation anomalies (Figure S6). Therefore, it seems that the negative PNA events are more associated with internal atmospheric variability, rather than forced from the tropics as in positive PNA events.

This asymmetry in PNA hindcast skill is likely related to the skewed distribution of SST anomalies in the tropical Pacific, with larger amplitude SST anomalies occurring in El Niño winters in comparison with La Niña winters [e.g., *Kumar et al.*, 2000]. Composite SST anomalies for upper and lower quintile PNA events in the model (i.e., 612 ensemble members) reveal a weaker asymmetry than in the observations (Figures 4c and 4e). However, the magnitude of the composite SST anomaly for the upper quintile PNA events is about 0.2°C

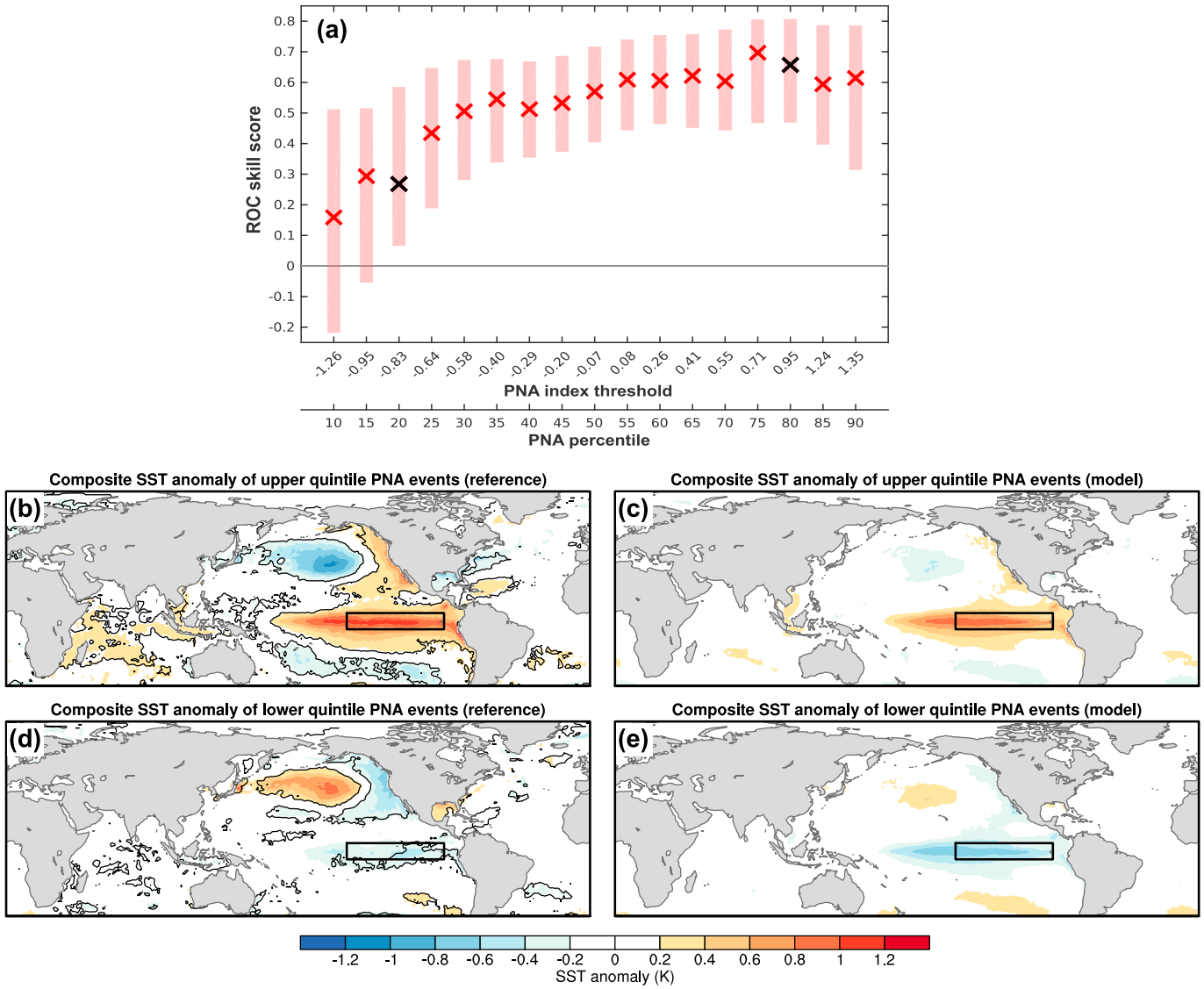

Figure 4. (a) The ROC skill score for PNA events at different thresholds, defined using percentiles of the reference PNA index. Upper/lower quintile thresholds are indicated by black crosses. The 5–95% uncertainty range is also shown. (b–e) Composite SST anomalies for the upper quintile (Figures 4b and 4c) and lower quintile (Figures 4d and 4e) PNA events in the reference data set (Figures 4b and 4d) and for all ensemble members of the model (Figures 4c and 4e). In Figures 4b and 4d values that are significant at the 5% level according to a *t* test are shown in black contours for the reference composites. In Figures 4c and 4e all shaded values exceed the 5% significance level.

greater than the composite SST anomaly for the lower quintile, which is found to be significant, suggesting that the model is able to somewhat replicate the asymmetric response.

Weisheimer et al. [2017] showed that the Pacific Decadal Oscillation (PDO) exhibited similar variability to the NAO hindcast skill. However, the PDO index is largely forced by the PNA on interannual time scales (Figure S7). The variability of the multidecadal PNA index, therefore, closely follows the multidecadal PDO index and also the multidecadal variability in PNA skill (Figure S7), because hindcasts of negative PNA events are less skillful. The relationship between the NAO and PNA skill variability, consequently, explains the apparent relationship between NAO skill variability and the PDO.

4. Summary and Discussion

We have analyzed global seasonal hindcast experiments for the Northern Hemisphere winter from the beginning of the twentieth century onward. The hindcast skill for the ensemble mean PNA exhibits dramatic variability over the hindcast period, with correlation dropping to nearly zero during the mid-twentieth century,

similar to the variability in NAO hindcast skill. The variability in the hindcast PNA skill closely follows the observed correlation between the observed PNA index and SST anomalies in the eastern tropical Pacific. In the observations there is a distinct weakening of the link between the Niño-3 index and the PNA during the mid-twentieth century. During the mid-century period, errors in the PNA hindcast are closely correlated with errors in the NAO hindcast, indicating that the drop in skill in the two indices is closely related. The drop in PNA predictability is related to a series of poorly forecast mid-century negative PNA events, which were not forced in a predictable manner by SST anomalies in the tropical Pacific. Our findings indicate that the negative phase of the PNA is somewhat less predictable and likely arises more from internal atmospheric variability than the positive phase of the PNA, which is more strongly forced by warm SST anomalies in the tropical Pacific.

A limitation of the hindcast setup used in this study is that it does not capture any source of skill from the initialization in the stratosphere, as has been cited in other seasonal forecast systems [e.g., *Scaife et al.*, 2014; *Stockdale et al.*, 2015; *Dunstone et al.*, 2016]. This is because the initialization from ERA-20C does not provide useful information in the stratosphere, where it differs most greatly from reanalysis products that assimilate upper air observations such as ERA-Interim [*Poli et al.*, 2016]. Another limitation is the temporal variability in the observational quality, particularly of the prescribed SST and sea ice boundary conditions, which are particularly uncertain around the second world war [*Kennedy*, 2014] and could be influencing the skill during the mid-century period.

The results presented in this study have implications for the analysis of other seasonal forecast systems. They suggest that seasonal forecasts that are assessed by hindcasts over only the most recent period (i.e., the last 30 years or so) may be less skillful in periods, such as the mid-twentieth century, when there was relatively weak forcing from SST anomalies in the tropical Pacific. As a result, the skill of future forecasts from seasonal forecast systems may be lower than has been reported for the most recent generation of these models [e.g., *Scaife et al.*, 2014; *Stockdale et al.*, 2015]. SST anomalies in the tropical Pacific were found to be crucial in the skillful prediction of the NAO in both the first and second winter in the Met Office's decadal prediction system (DePreSys) [*Dunstone et al.*, 2016]. We would expect DePreSys, or any other seasonal forecasting system which derives appreciable extratropical skill from the link with tropical Pacific SST anomalies, to demonstrate diminished levels of skill during a period characterized by negative PNA phase and weak tropical SST forcing. As demonstrated here, the observed link is seemingly less predictable.

Acknowledgments
This work was supported in part by the Natural Environmental Research Council (C.O.R, T.W., and A.W.; grant NE/M005887/1), the EU-funded EUCLEIA project, and the EU-funded SPECS project (D.M. and A.W.; EU-FP7, grant agreement 308378). J.H. was supported by the Met Office Academic Partnership scheme. The model output from the hindcasts in this study is available on request.

References

Abid, M. A., I.-S. Kang, M. Almazroui, and F. Kucharski (2015), Contribution of synoptic transients to the potential predictability of PNA circulation anomalies: El Niño versus La Niña, *J. Clim.*, *28*(21), 8347–8362.
Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt (2012), Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?, *Bull. Am. Meteorol. Soc.*, *93*(5), 631–651.
Buizza, R., and T. N. Palmer (1998), Impact of ensemble size on ensemble prediction, *Mon. Weather Rev.*, *126*(9), 2503–2518.
Drouard, M., G. Rivière, and P. Arbogast (2015), The link between the North Pacific climate variability and the North Atlantic Oscillation via downstream propagation of synoptic waves, *J. Clim.*, *28*(10), 3957–3976.
Dunstone, N., D. Smith, A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews, and J. Knight (2016), Skilful predictions of the winter North Atlantic Oscillation one year ahead, *Nat. Geosci.*, *9*, 809–814.
Hoerling, M. P., A. Kumar, and M. Zhong (1997), El Niño, La Niña, and the nonlinearity of their teleconnections, *J. Clim.*, *10*(8), 1769–1786.
Honda, M., and H. Nakamura (2001), Interannual seesaw between the Aleutian and Icelandic lows. Part II: Its significance in the interannual variability over the wintertime Northern Hemisphere, *J. Clim.*, *14*(24), 4512–4529.
Horel, J. D., and J. M. Wallace (1981), Planetary-scale atmospheric phenomena associated with the Southern Oscillation, *Mon. Weather Rev.*, *109*(4), 813–829.
Kennedy, J. J. (2014), A review of uncertainty in in situ measurements and data sets of sea surface temperature, *Rev. Geophys.*, *52*, 1–32, doi:10.1002/2013RG000434.
Kharin, V. V., and F. W. Zwiers (2003), On the ROC score of probability forecasts, *J. Clim.*, *16*(24), 4145–4150.
Kim, H.-M., P. J. Webster, and J. A. Curry (2012), Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, *Clim. Dyn.*, *39*(12), 2957–2973.
Kumar, A., A. G. Barnston, P. Peng, M. P. Hoerling, and L. Goddard (2000), Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO, *J. Clim.*, *13*(17), 3139–3151.
Leathers, D. J., B. Yarnal, and M. A. Palecki (1991), The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations, *J. Clim.*, *4*(5), 517–528.
Lin, H., and J. Derome (1996), Changes in predictability associated with the PNA pattern, *Tellus A*, *48*(4), 553–571.
Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravanan, and M. Visbeck (2001), North Atlantic climate variability: Phenomena, impacts and mechanisms, *Int. J. Climatol.*, *21*(15), 1863–1898.
Mason, S. J., and N. E. Graham (1999), Conditional probabilities, relative operating characteristics, and relative operating levels, *Weather Forecasting*, *14*(5), 713–725.
Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer, and F. Vitart, (2011), The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo, European Centre for Medium Range Weather Forecasts, Reading, Berkshire.

Müller, W., C. Appenzeller, and C. Schär (2005), Probabilistic seasonal prediction of the winter North Atlantic Oscillation and its impact on near surface temperature, *Clim. Dyn.*, *24*(2–3), 213–226.
Palmer, T. (1988), Medium and extended range predictability and stability of the Pacific/North American mode, *Q. J. R. Meteorol. Soc.*, *114*(481), 691–713.
Palmer, T., et al. (2004), Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bull. Am. Meteorol. Soc.*, *85*(6), 853–872.
Peng, P., and A. Kumar (2005), A large ensemble analysis of the influence of tropical SSTs on seasonal atmospheric variability, *J. Clim.*, *18*(7), 1068–1085.
Pinto, J. G., M. Reyers, and U. Ulbrich (2011), The variable link between PNA and NAO in observations and in multi-century CGCM simulations, *Clim. Dyn.*, *36*(1–2), 337–354.
Poli, P., et al. (2016), ERA-20C: An atmospheric reanalysis of the twentieth century, *J. Clim.*, *29*(11), 4083–4097.
Rayner, N., D. E. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.
Riddle, E. E., A. H. Butler, J. C. Furtado, J. L. Cohen, and A. Kumar (2013), CFSv2 ensemble prediction of the wintertime Arctic Oscillation, *Clim. Dyn.*, *41*(3–4), 1099–1116.
Scaife, A., et al. (2014), Skillful long-range prediction of European and North American winters, *Geophys. Res. Lett.*, *41*, 2514–2519, doi:10.1002/2014GL059637.
Sheng, J. (2002), GCM experiments on changes in atmospheric predictability associated with the PNA pattern and tropical SST anomalies, *Tellus A*, *54*(4), 317–329.
Shi, W., N. Schaller, D. MacLeod, T. Palmer, and A. Weisheimer (2015), Impact of hindcast length on estimates of seasonal climate predictability, *Geophys. Res. Lett.*, *42*, 1554–1559, doi:10.1002/2014GL062829.
Smith, D. M., A. A. Scaife, and B. P. Kirtman (2012), What is the current state of scientific knowledge with regard to seasonal and decadal forecasting?, *Environ. Res. Lett.*, *7*(1), 015602.
Stockdale, T. N., F. Molteni, and L. Ferranti (2015), Atmospheric initial conditions and the predictability of the Arctic Oscillation, *Geophys. Res. Lett.*, *42*, 1173–1179, doi:10.1002/2014GL062681.
Straus, D. M., and J. Shukla (2002), Does ENSO force the PNA?, *J. Clim.*, *15*(17), 2340–2358.
Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski (1998), Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures, *J. Geophys. Res.*, *103*(C7), 14,291–14,324.
Wallace, J. M., and D. S. Gutzler (1981), Teleconnections in the geopotential height field during the Northern Hemisphere winter, *Mon. Weather Rev.*, *109*(4), 784–812.
Wallace, J. M., C. Smith, and C. S. Bretherton (1992), Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, *J. Clim.*, *5*(6), 561–576.
Weisheimer, A., F. Doblas-Reyes, T. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel (2009), ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs, *Geophys. Res. Lett.*, *36*, L21711, doi:10.1029/2009GL040896.
Weisheimer, A., S. Corti, T. Palmer, and F. Vitart (2014), Addressing model error through atmospheric stochastic physical parametrizations: Impact on the coupled ECMWF seasonal forecasting system, *Philos. Trans. R. Soc. A*, *372*(2018), 20130290.
Weisheimer, A., N. Schaller, C. O'Reilly, D. A. MacLeod, and T. Palmer (2017), Atmospheric seasonal forecasts of the 20th Century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution, *Q. J. R. Meteorol. Soc.*, *143*, 917–926, doi:10.1002/qj.2976.
Xue, Y., M. Chen, A. Kumar, Z.-Z. Hu, and W. Wang (2013), Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2, *J. Clim.*, *26*(15), 5358–5378.