Key Points:
- The Arctic Oscillation is skillfully predicted by the North American Multimodel Ensemble up to 7 months in advance
- The Arctic Oscillation and El Niño-Southern Oscillation are strongly related in the North American Multimodel Ensemble predictions
- A weaker relation in the observations implies that the models are overconfident

Supporting Information:
- Supporting Information S1

Correspondence to:
M. L. L'Heureux,
michelle.lheureux@noaa.gov

Strong Relations Between ENSO and the Arctic Oscillation in the North American Multimodel Ensemble

Michelle L. L'Heureux[1] , Michael K. Tippett[2,3] , Arun Kumar[1] , Amy H. Butler[4] , Laura M. Ciasto[1,5] , Qinghua Ding[6] , Kirstin J. Harnos[1,5] , and Nathaniel C. Johnson[7,8]

[1]National Oceanic and Atmospheric Administration, NWS/NCEP/Climate Prediction Center, College Park, MD, USA, [2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, [3]Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia, [4]NOAA/ESRL Chemical Sciences Division, University of Colorado Boulder/CIRES, Boulder, CO, USA, [5]Innovim, Greenbelt, MD, USA, [6]Deptartment of Geography, University of California, Santa Barbara, CA, USA, [7]National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [8]Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, NJ, USA

Abstract Arctic Oscillation (AO) variability impacts climate anomalies over the middle to high latitudes of the Northern Hemisphere. Recently, state-of-the-art climate prediction models have proved capable of skillfully predicting the AO during the winter, revealing a previously unrealized source of climate predictability. Hindcasts from the North American Multimodel Ensemble (NMME) show that the seasonal, ensemble mean 200 hPa AO index is skillfully predicted up to 7 months in advance and that this skill, especially at longer leads, is coincident with previously unknown and strong relations ($r > 0.9$) with the El Niño–Southern Oscillation (ENSO). The NMME is a seasonal prediction system that comprises eight models and up to 100 members with forecasts out to 12 months. Observed ENSO-AO correlations are within the spread of the NMME member correlations, but the majority of member correlations are stronger than observed, consistent with too high predictability in the model, or overconfidence.

1. Introduction

Over the middle- and high-latitude regions of North America, Europe, and Asia, the Arctic Oscillation (AO) or North Atlantic Oscillation (NAO) explains a significant fraction of temperature and precipitation variance, particularly during the Northern Hemisphere winter (Hurrell et al., 2013; Thompson & Wallace, 2001). In many regions of the Northern Hemisphere extratropics, the AO explains more climate variability than does the El Niño–Southern Oscillation (ENSO), whose seasonal predictability makes it a primary contributor to skill in seasonal climate outlooks. The AO, however, has long been considered an internal mode of climate variability, resulting from feedbacks between the zonal mean flow and synoptic-scale eddies (Lorenz & Hartmann, 2003). Thus, skill in forecasting the AO was largely believed to be limited to weather prediction time scales (Feldstein, 2000).

The recent documentation of skillful predictions of the wintertime AO/NAO in many state-of-the-art climate models is a revelation (Kang et al., 2014; Riddle et al., 2013; Scaife et al., 2014; Stockdale et al., 2015) and promises to improve climate outlooks in regions with AO/NAO teleconnections. A wide range of phenomena have been proposed as potential sources of predictability for the winter AO, with time scales of seasons to decades (e.g., Smith et al., 2016). Among other factors, AO skill could arise from tropical climate anomalies, including ENSO (Dai & Tan, 2017; Dunstone et al., 2016; Tang et al., 2007; Yu & Lin, 2016), as well as Arctic sea ice (Wang et al., 2017), Eurasian snow cover (Cohen et al., 2010), and stratospheric-tropospheric coupling (Butler et al., 2016; Sigmond et al., 2013). Whatever mechanisms are responsible for AO skill, it would seem that only now have state-of-the-art forecasting systems been able to represent them with sufficient accuracy, either through improved physics, better data assimilation, or improved observing networks.

An alternative perspective is that periods with higher skill can also occur naturally and randomly, without precursors, and do not necessarily reflect improvements in the models or observations. Shi et al. (2015) find periods of higher skill, rivaling recent findings, and also periods of lower NAO skill on subsets of a 42 year set of forecasts. Estimates of the ability to predict climate phenomena depend on sample size and choice

of historical period. For instance, Barnston et al. (2012) reveal that ENSO skill has significant decadal and subdecadal fluctuations in a set of operational climate forecasts. The relation between indices can also vary significantly as shown by O'Reilly et al. (2017) who note changing correlations among tropical Pacific SST anomalies, the Pacific/North American (PNA), and NAO indices.

Here we find that the seasonal mean AO index is skillfully predicted by the multimodel ensemble mean of the eight climate models that compose the North American Multimodel Ensemble (NMME) for lead times out to at least 5 months for winter and early spring target seasons. We define the AO using the 200 hPa geopotential height level because it is the only model level available; thus, we are examining the lower stratosphere poleward of 60°N. Ensemble mean forecasts of the AO and ENSO (Niño-3.4) indices are very strongly related, with correlation coefficients in excess of 0.9 during the winter. That this relation does not emerge nearly as strongly in the observational data suggests that the AO-ENSO covariance could be physically meaningful and a relevant factor in AO prediction skill, but small relative to the internal atmospheric variability of the AO for most lead times and seasons. Alternatively, the AO in the NMME models may be overly responsive to variability in ENSO, especially for those target seasons when the observed AO-ENSO relation is weakest.

2. Data and Methods

The NMME includes eight coupled models: GFDL-CM2p1-aer04, NASA-GMAO-062012, COLA-RSMAS-CCSM4, GFDL-CM2p5-FLOR-A06, GFDL-CM2p5-FLOR-B01, CMC1-CanCM3, CMC2-CanCM4, and National Centers for Environmental Prediction (NCEP) CFSv2 (Kirtman et al., 2014). Hindcast data are provided for the period 1982–2010, and real-time forecasts started in mid-2011. The results here are for the period 1982–2016. Monthly data are provided on a 1° × 1° grid and forecasts extend up to 12 months into the future. Seasonally averaged forecasts extend nine forecast leads beyond the starting month. For example, lead-0 seasonal forecasts initialized in early December predict the December-January-February (DJF) average. The only exception is NCEP CFSv2, which, for the hindcast period, is run every fifth day (four members each day), with the six times per month initializations occurring up to the seventh day of the start month.

Monthly data from NMME are bias corrected by removing the forecast (lead time dependent) climatology of each model separately using the entire 1982–2016 period. Observed anomalies are also calculated relative to the 1982–2016 monthly mean averages.

The ENSO index is based on Niño-3.4 sea surface temperature anomalies in the east-central equatorial Pacific Ocean (5°S–5°N, 170°W–120°W; Barnston et al., 1997). The observational monthly and seasonal averages in the Niño-3.4 index are formed from daily 0.25° × 0.25° Optimum Interpolation sea surface temperature data (Reynolds et al., 2007), which is chosen because its higher resolution more closely matches the initial conditions of several NMME models.

The AO index that we use is computed from the leading Empirical Orthogonal Function (EOF) of 200 hPa geopotential height data for 20°–90°N and over all calendar months (Thompson & Wallace, 1998). This level was selected due to NMME data availability in the International Research Institute for Climate and Society (IRI) archive (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/). The 200 hPa observational data are based on monthly NCEP/NCAR Reanalysis data (Kalnay et al., 1996). For correlations between the AO and Niño-3.4 indices, one of the indices is inverted for easier visualization, so a positive coefficient means that the AO and ENSO are inversely related (e.g., positive AO and cool phase of ENSO). Statistical significance of the correlation coefficients is assessed at the 5% significance level using a one-sided Student's t test.

Due to the size of the NMME data set (100 Members × 35 Years × 12 Monthly Lead Times × ~65,000 Grid points) and because the AO index is a large-scale mode, we employ a spatial dimension reduction technique and project the model and observational data onto the 100 leading Laplacian eigenvectors of the 20°–90°N domain (DelSole & Tippett, 2015; Saito, 2008). Laplacian eigenvectors are orthogonal and similar to spherical harmonics but are defined for general domains (access here: http://iridl.ldeo.columbia.edu/home/.tippett/.Laplacian/.NMME). The observed AO index is computed from the leading EOF of the NCEP/NCAR Reanalysis 200 hPa geopotential height. The model AO index is computed by projecting the model data onto the observed EOF. Therefore, the AO indices are formed in each of the NMME models with respect to the same observation-based pattern.
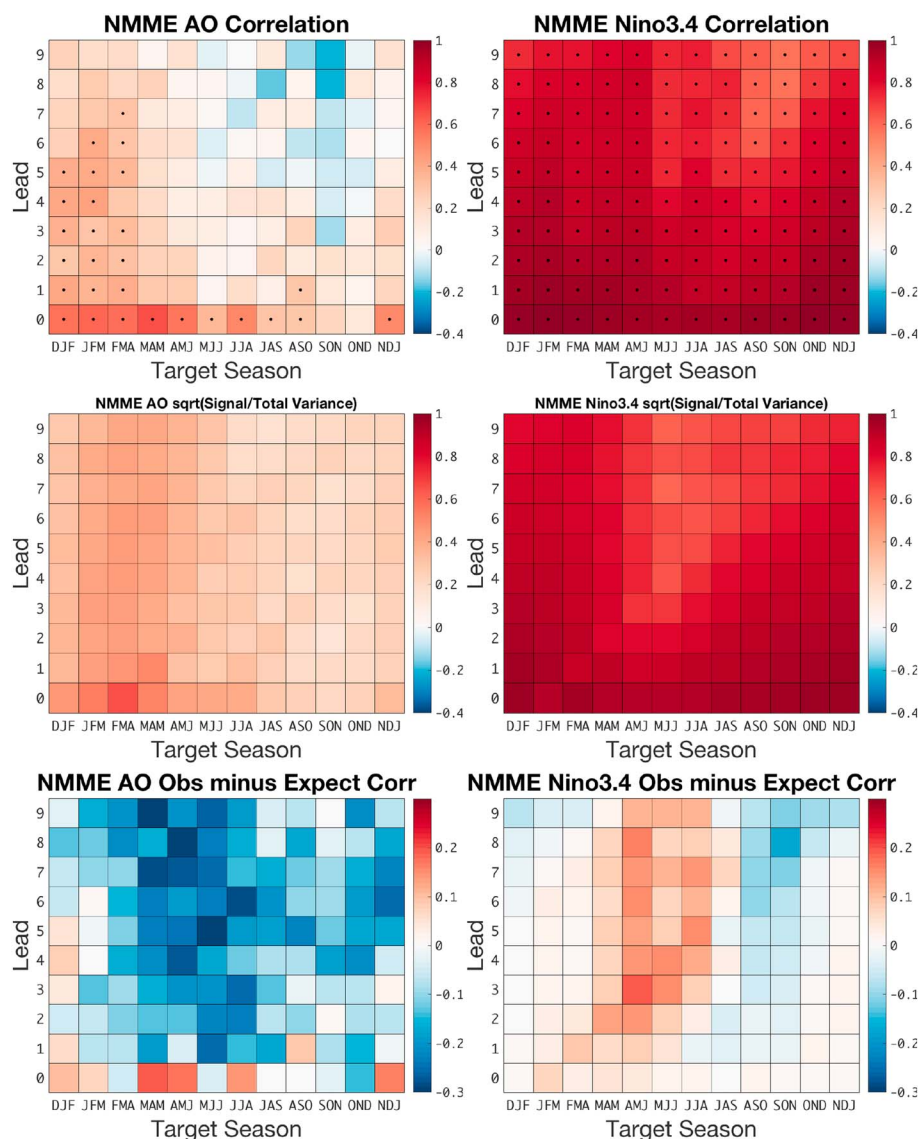
Figure 1. Anomaly correlation of the (left column) AO index or the (right column) Niño-3.4 index between seasonal mean observations and the NMME ensemble mean forecast (top row), the theoretical expected correlation based the square root of the Signal Variance (ensemble mean) divided by the Total Variance (middle row), and the absolute value of the top row minus the middle row (bottom row). In the bottom row, red shading indicates where the absolute value of the actual correlation (top row) is greater than the expected correlation (middle row) and vice versa for blue shading. Dots in the top row indicate correlations at the 5% significance level. The lead time is by month, for sliding seasonal averages.

3. Results

For the first-lead seasonal forecasts (lead-0), the correlation between the ensemble mean NMME forecasts and the observed 200 hPa Arctic Oscillation (AO) index is statistically significant in all seasons except September–November (SON) and October–December (OND; Figure 1, left column, top). The largest lead-0 correlations occur during the Northern Hemisphere winter and spring, with a correlation of 0.57 during the winter season of December–February (DJF) and maximizing at 0.66 during March–May (MAM). Interestingly, the only NMME models that do not show significant skill at lead-0 are the two GFDL FLOR models, which are initialized using only ocean observations (supporting information Figure S1 shows correlations for all individual models and monthly targets). Thus, there appears to be an advantage to initializing the atmosphere, which can provide skillful information on the state of the AO for the upcoming month or season.
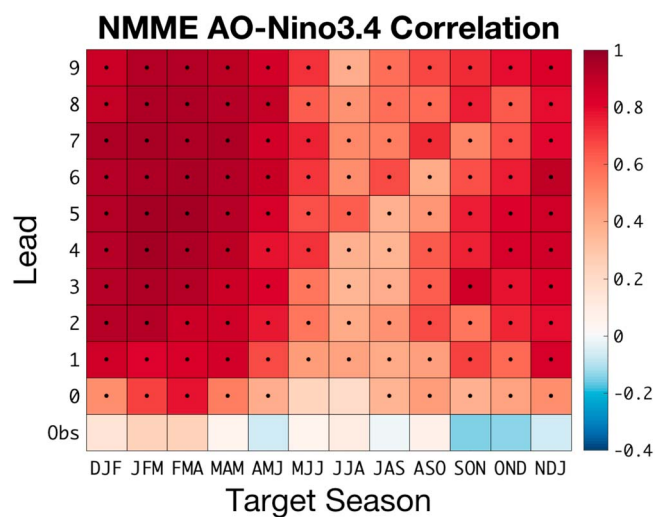
NMME AO-Nino3.4 Correlation

Figure 2. Anomaly correlation between of the NMME ensemble mean Niño-3.4 and AO index values. Dots indicate the 5% significance level. The bottom row (Obs) indicates the contemporaneous correlation between the Niño-3.4 index and the AO index. The AO index has been inverted. The lead time is by month, for sliding seasonal averages.

Beyond lead-0, NMME skill is mostly insignificant except for forecast targets during the three overlapping seasons of DJF to FMA (Figure 1, left column, top). The lead-0 and lead-1 correlations appear to be smaller than the ∼0.6 correlations for DJF documented for individual modeling systems (Kang et al., 2014; Riddle et al., 2013; Scaife et al., 2014; Stockdale et al., 2015). As the lead time increases to 5–7 months, the correlations drop to ∼0.4, indicating that roughly 16% of the observed winter to spring AO variability is captured by the NMME mean for forecasts made as early as July.

The skill of the forecast depends on the specific time interval (in this case, 1982–2016) and may not reflect the intrinsic predictability of the climate system. Scaife et al. (2014) and Eade et al. (2014) note that the signal-to-noise ratio, a measure of predictability, is much lower in the UKMet model during December–February (∼0.2, corresponding to a correlation of ∼0.4) than one would expect given correlations between their model and the observations. They argue that their model may be underconfident and argue that the real world is more predictable than what the model signal-to-noise would suggest. Based on theoretical considerations, Kumar (2009) and Tippett et al. (2010) provide a derivation for an "expected correlation" for a given signal-to-noise ratio. To compare the actual correlation between the forecast and the observations (Figure 1, left column, top) to the expected AO correlation (Figure 1, left column, middle), the absolute value is taken and the difference is shown in Figure 1 (left column, bottom). For most seasons and leads, the expected correlations are larger than the actual correlations (more negative values), indicating higher predictability in the model, or overconfidence. This overconfidence is especially prominent for forecasts of spring and summer targets, which tend to have lower variance relative to the winter in both the models and the observations (Figure S2). Because the observed and model total variances are similar during these seasons, it follows that the model signal may be too large (or equivalently that the noise is too small), leading to higher expected correlations or predictability.

In contrast with the spring and summer, the differences for wintertime targets are smaller and imply that the predictability implied by the NMME ensemble is reasonably capturing that of the real AO (Figure 1, column left, bottom). For DJF only, there is a weak indication of forecast under confidence out to lead-5. At longer leads (lead-7 and beyond), there is a suggestion of overconfidence or higher predictability in the model, as indicated by the larger expected correlations. Supporting information Figure S2 corroborates that the total variance of the model is, in fact, lower than the observations, implying slight underdispersion, which is a result that often emerges for other variables and domains (e.g., Shi et al., 2015).

In contrast to the AO, the correlations associated with ENSO are statistically significant for all seasons and leads (Figure 1, right column, top). Skills are in excess of 0.9 up to the lead-4 forecasts (for targets outside of the summer/early fall) and minimize at ∼0.6 for the longest lead time (lead-9) for the SON target, reflecting the weakness in skill for forecasts traversing the well documented "spring barrier" (Barnston et al., 2017; Tippett et al., 2017). In addition to Niño-3.4 predictions having higher skill than the AO predictions, the expected Niño-3.4 correlations (Figure 1, right column, middle) strongly resemble the actual correlations between the observations and ensemble mean (Figure 1, right column, top). The differences between the expected and actual correlations are generally very small for short lead times (leads 0 and 1) and for all lead times for targets in the winter and early spring (Figure 1, right column, bottom). Thus, for most seasons, the observed correlations match the predictability estimated by the NMME. Interestingly, the largest gap is evident for targets during the spring and summer, when the actual correlations are higher than the expected ones. This occurs during seasons and leads with a relative minimum in skill, and yet from this perspective, the implication is that the model's estimate of predictability may still be slightly too low. Lower expected correlations are consistent with the models tending to have higher variance than observations (supporting information Figure S2).

Despite the relatively high level of skill in predicting ENSO and low level of skill in the AO, Figure 2 shows that the correlation between the ensemble means of the Niño-3.4 index and the AO index is statistically significant
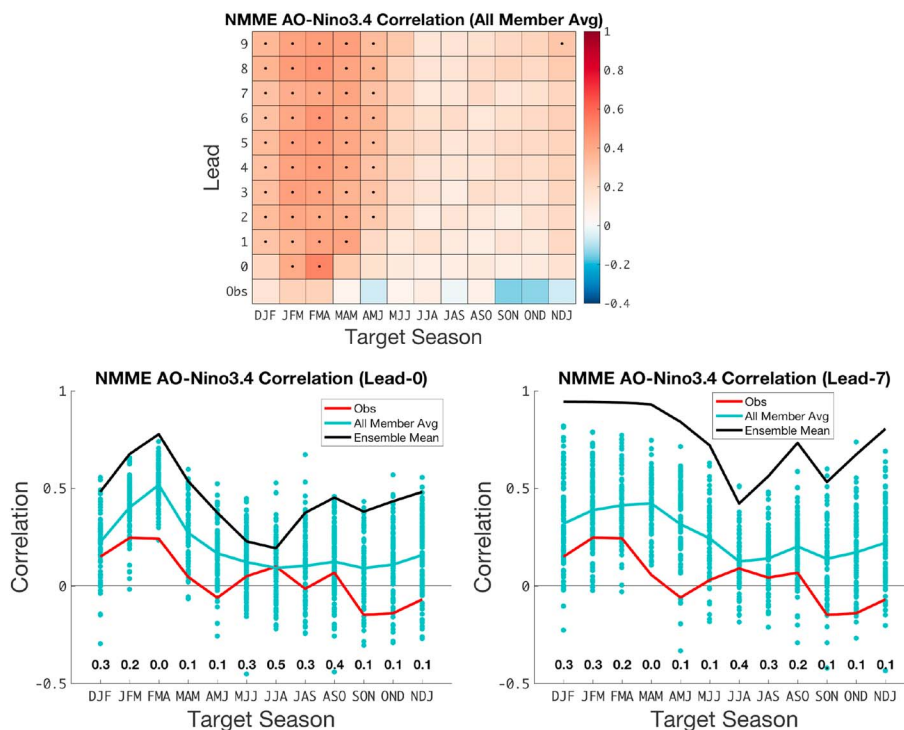
Figure 3. All member average of 100 correlations between seasonal averages of the (top) NMME ensemble mean Niño-3.4 and AO index values. Bottom row shows the observed correlations and dots indicate the 5% significance level. For forecast Lead-0 (bottom left) and forecast Lead-7 (bottom right), each individual Niño-3.4 and AO correlation is shown for the 100 members (blue dots). The red line shows the observed correlations, the blue line shows the all member average (same as top panel), and the black line is the ensemble mean correlations indicated in Figure 2. The AO index has been inverted. The lead time is by month, for sliding seasonal averages. Numbers above the abscissa indicate the fraction of ensemble members whose Niño-3.4 and AO correlation is less than the observed correlation.

for nearly all lead times and seasons in the NMME (supporting information Figure S3 shows correlations for all individual models and monthly targets). The correlations between the two indices exceed 0.9 for targets during DJF through MAM for lead-3 to lead-9, with a maximum correlation of 0.96 for lead-5 in JFM and FMA. A minimum in correlations between the two indices occurs during the summer and early fall (JJA through ASO), which are the seasons when both the AO and Niño-3.4 skills are relatively lower (Figure 1). The high correlations found in the NMME stand in striking contrast with the substantially weaker contemporaneous correlations found in the observed data, which range from 0.25 (JFM and FMA) to −0.15 during SON. The observed correlations are reproduced when we use other SST and reanalysis data sets (supporting information Figure S4).

However, the observations contain both signal and noise components, whereas the ensemble mean relationship uncovered here isolates only the correlations among the signals in the predicted ENSO and AO indices. Thus, the lack of a relation in the observations may simply reflect, at least in part, the noise of internal atmospheric variability, whereby a truly physical connection between the two phenomena is metaphorically drowned out.

Though the correlations are highest between the ensemble mean indices, significant relations in the NMME also extend to the individual members, primarily during the winter and spring seasons. Figure 3 shows the correlation between Niño-3.4 and the AO index in all 100 members of the NMME. The average correlation of the individual members is shown in the top panel, with the bottom panels showing the all member average (blue line) and associated spread of the member correlations (blue dots) for lead-0 (left panel) and lead-7 (right panel). Lead-7 is displayed because of the generally high level of AO-Niño-3.4 association. The observational data represent just one possible outcome, so the forecast ensemble should ideally encompass that one outcome. In fact, this is the case, where the observed relationship between Niño-3.4 and the AO indices (red line) appears to mostly lie within the spread of the model ensemble (blue dots), although consistently below the average.
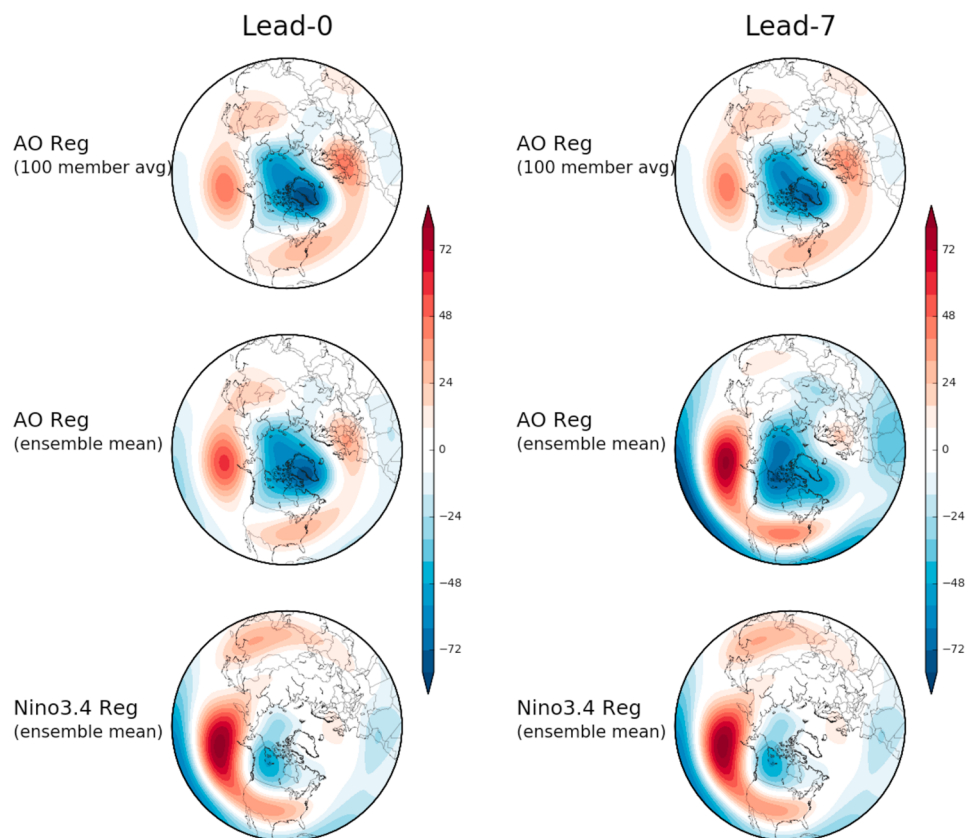
Figure 4. NMME z200 anomalies regressed onto standardized indices computed for NMME lead-0 (left column) and lead-7 (right column) based on monthly data from January 1982 to December 2016. (top row) The z200 regressed onto the AO index but first computed for all 100 members individually and then averaged together. (middle row) The z200 ensemble average regressed onto the AO ensemble average. (bottom row) The z200 ensemble average regressed onto the Niño-3.4 ensemble average. The bottom row is multiplied by a factor of 3 to aid comparison with the middle row and the Niño-3.4 index has been inverted.

The relatively lackluster correlations seen in the observed indices may be reflecting internal atmospheric variability that is only present to a lesser degree in the models. The idea that sampling variability is also playing a role in the observed AO-ENSO correlations is also hinted at by the spread of correlations in moving 35 year sliding windows in data sets going back to 1950 (supporting information Figure S4, bottom). Especially during the boreal winter season, the correlations over the most recent 1982–2016 period appear to be on the lower end relative to other historical periods.

Nevertheless, Figure 3 shows that the average of the individual member correlations (blue line) and the correlations of the ensemble mean (black line; same data as in Figure 2) are stronger than in the observations. Thus, even when taking the noise in the NMME into account, the ENSO-AO relationship in the model is—on average—stronger than the observed relationship, which is on the low end of the model spread for initializations during the 1982–2016 period. That the observations lie on the lower end of the spread for all target seasons seems unlikely to be due to chance. For half of the seasons, the observed correlations are lower than 10% of the model correlations. The results in Figure 1 may indicate that the model's estimate of AO predictability is too large, or alternatively, the model estimate may be correct or even too small, but the model skill is deficient because other predictable signals are missing or distorted. Thus, one possible interpretation is that the model ENSO-AO relationship is unrealistically strong in the NMME models.

Figure 4 shows the lead-0 and lead-7 forecasts of the 200 hPa geopotential height anomalies regressed onto the predicted AO and Niño-3.4 indices over all months. Figure 4 (top row) shows the average of 100 regression maps based on the AO indices created for all 100 members. For both lead times, the height pattern matches the conventional AO pattern featured in Thompson and Wallace (2000), among many others, with out-of-phase height anomalies over the Arctic and the middle latitudes. The similarity between both leads

suggests that an individual NMME member of the AO is an appropriate analog to the observational AO index, containing both signal and noise. However, after averaging together the 100 AO indices to suppress the noise, and then regressing the single ensemble mean AO index onto the height field, a different picture emerges (Figure 4, middle row). In this case, the lead-7 AO index projects onto a height pattern more similar to the anomalies associated with ENSO (Figure 4, bottom row). Supporting information Figure S5 shows the lead-7 AO and the inverted Niño-3.4 ensemble mean indices for February-April targets (selected because skill in the AO and Niño-3.4 extends out to lead-7) and also reveals a very close correspondence between the temporally evolving signals of Niño-3.4 and the AO. Yet this similarity is not as obvious for the lead-0 ensemble mean indices, which have noticeably higher amplitudes, signifying that the drivers of the AO, in addition to ENSO, are increasingly incorporated into the forecasted signal.

4. Discussion and Future Work

The results here are presented for the NMME, which encompasses a diverse range of models and a large set of ensemble members. There are inherent advantages in using such a forecast system to study the skill of ENSO and AO because the models are constructed at different institutions, so are at least somewhat structurally independent (though there can be some shared components). Furthermore, a larger set of members enables better estimates of predictability. Notwithstanding the advantages of the NMME, it is still subject to some of the same limitations as other individual models and multimodel combinations. The limited hindcast record length (36 years) means that the AO-ENSO behavior may be a function of the sample period. There may be biases in the model, and so the AO-ENSO relation could be unrealistically strong, though the member spread is inclusive of the observations. And finally, while NMME ensemble is — to our knowledge — the largest to date used to examine AO skill, the signal and noise are still unlikely to be perfectly known. But because real-world observations only contain one realization, we rely on the models to provide the best estimate of predictability (Deser et al., 2012; Jha et al., 2016).

Despite these limitations, the strong relations between ENSO and the AO documented herein are provocative, indicating that a part of the predictable signal associated with the AO depends on ENSO, especially at longer lead times. Whether these relations are present in other models and multi-modeling systems remains to be seen. The long-lead relation is also suggested by Dunstone et al. (2016) who show, using multivariate regression and correlation, that ENSO and the stratospheric polar vortex account for NAO skill out to ~13 months. Because our analysis is restricted to the 200 hPa level, other levels may provide additional insight. Previous research indicates that current state-of-the-art modeling systems underestimate stratospheric-tropospheric coupling (Riddle et al., 2013; Stockdale et al., 2015). Correctly capturing these linkages may be even more important because the high-latitude response to ENSO is likely also modulated by the stratosphere (Butler et al., 2014; Ineson & Scaife, 2009; Polvani et al., 2017). The high correlation between ENSO and the AO during the winter and spring may suggest stratospheric-tropospheric coupling, so it would be useful to diagnose whether AO-ENSO covariability arises from this mechanism or from tropospheric teleconnections alone.

Overall, understanding the skill and predictability of the AO is critical to improving middle- to high-latitude climate outlooks. The AO predictability is likely overestimated with overconfident forecasts during most seasons and lead times, and this reflects a model response that is too sensitive to ENSO. This sensitivity may arise from the models underrepresenting or missing other sources of predictability. Alternatively, the strong AO skill could be the consequence of observed sampling variability, in which skill fluctuates naturally and randomly over time. Even if the physics of the AO-ENSO connection could be firmly established, the low correlations in the observations relative to those in the models still prompts the question of whether the change in the forecasted ensemble means is large enough to be of consequence. This question, among many others that arise in light of this ENSO-AO association in NMME, justifies further testing methods to constrain our estimates of predictability and to better understand the sources of prediction skill in the middle to high latitudes.

References

Barnston, A. G., Chelliah, M., & Goldenberg, S. B. (1997). Documentation of a highly ENSO-related SST region in the equatorial Pacific: Research Note. *Atmosphere-Ocean*, *35*(3), 367–383. https://doi.org/10.1080/07055900.1997. 9649597

Barnston, A. G., Tippett, M. K., L'Heureux, M. L., Li, S., & DeWitt, D. G. (2012). Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, *93*, 631–651. https://doi.org/10.1175/BAMS-D-11-00111.1

Barnston, A. G., Tippett, M. K., Ranganathan, M., & L'Heureux, M. L. (2017). Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, 1–20. https://doi.org/10.1007/s00382-017-3603-3

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., … Yasuda, T. (2016). The Climate-system Historical Forecast Project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quarterly Journal of the Royal Meteorological Society*, *142*(696), 1413–1427. https://doi.org/10.1002/qj.2743

Butler, A. H., Polvani, L. M., & Deser, C. (2014). Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections. *Environmental Research Letters*, *9*(2), 024014.

Cohen, J., Foster, J., Barlow, M., Saito, K., & Jones, J. (2010). Winter 2009–2010: A case study of an extreme Arctic Oscillation event. *Geophysical Research Letters*, *17*, L17707. https://doi.org/10.1029/2010GL044256

Dai, P., & Tan, B. (2017). The nature of the Arctic Oscillation and diversity of the extreme surface weather anomalies it generates. *Journal of Climate*, *30*, 5563–5584. https://doi.org/10.1175/JCLI-D-16-0467.1

DelSole, T., & Tippett, M. K. (2015). Laplacian eigenfunctions for climate analysis. *Journal of Climate*, *28*(18), 7420–7436. https://doi.org/10.1175/JCLI-D-15-0049.1

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*, 527–546. https://doi.org/10.1007/s00382-010-0977-x

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., … Knight, J. (2016). Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nature Geoscience*, *9*(11), 809–814.

Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*, 5620–5628. https://doi.org/10.1002/2014GL061146

Feldstein, S. B. (2000). The timescale, power spectra, and climate noise properties of teleconnection patterns. *Journal of Climate*, *13*(24), 4430–4440. https://doi.org/10.1175/1520-0442(2000)013<4430:TTPSAC>2.0.CO;2

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (2013). An overview of the North Atlantic oscillation. In J. W. Hurrell et al. (Eds.), *The North Atlantic Oscillation: Climatic significance and environmental impact* (pp. 1–35). Washington, DC: American Geophysical Union. https://doi.org/10.1029/134GM01

Ineson, S., & Scaife, A. A. (2009). The role of the stratosphere in the European climate response to El Niño. *Nature Geoscience*, *2*, 32–36.

Jha, B., Kumar, A., & Hu, Z.-Z. (2016). An update on the estimate of predictability of seasonal mean atmospheric variability using North American Multi-Model Ensemble. *Climate Dynamics*, 1–13. https://doi.org/10.1007/s00382-016-3217-1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kang, D., Lee, M.-I., Im, J., Kim, D., Kim, H.-M., Kang, H.-S., … MacLachlan, C. (2014). Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems. *Geophysical Research Letters*, *41*, 3577–3585. https://doi.org/10.1002/2014GL060011

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., … Wood, E. F. (2014). The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Kumar, A. (2009). Finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, *137*, 2622–2631. https://doi.org/10.1175/2009MWR2814.1

Lorenz, D. J., & Hartmann, D. L. (2003). Eddy zonal flow feedback in the Northern Hemisphere winter. *Journal of Climate*, *16*(8), 1212–1227. https://doi.org/10.1175/1520-0442(2003)16<1212:EFFITN>2.0.CO;2

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

Polvani, L. M., Sun, L., Butler, A. H., Richter, J. H., & Deser, C. (2017). Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *Journal of Climate*, *30*(6), 1959–1969. https://doi.org/10.1175/JCLI-D-16-0277.1

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Riddle, E. E., Butler, A. H., Furtado, J. C., Cohen, J. L., & Kumar, A. (2013). CFSv2 ensemble prediction of the wintertime Arctic oscillation. *Climate Dynamics*, *41*(3), 1099–1116. https://doi.org/10.1007/s00382-013-1850-5

Saito, N. (2008). Data analysis and representation on a general domain using eigenfunctions of Laplacian. *Applied and Computational Harmonic Analysis*, *25*(1), 68–97. https://doi.org/10.1016/j.acha.2007.09.005

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., … Williams, A. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*, 2514–2519. https://doi.org/10.1002/2014GL059637

Shi, W., Schaller, N., MacLeod, D., Palmer, T. N., & Weisheimer, A. (2015). Impact of hindcast length on estimates of seasonal climate predictability. *Geophysical Research Letters*, *42*, 1554–1559. https://doi.org/10.1002/2014GL062829

Sigmond, M., Scinocca, J. F., Kharin, V. V., & Shepherd, T. G. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98–102.

Smith, D. M., Scaife, A. A., Eade, R., & Knight, J. R. (2016). Seasonal to decadal prediction of the winter North Atlantic Oscillation: Emerging capability and future prospects. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 611–617. https://doi.org/10.1002/qj.2479

Stockdale, T. N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophysical Research Letters*, *42*, 1173–1179. https://doi.org/10.1002/2014GL062681

Tang, Y., Lin, H., Derome, J., & Tippett, M. K. (2007). A predictability measure applied to seasonal predictions of the Arctic Oscillation. *Journal of Climate*, *20*(18), 4733–4750. https://doi.org/10.1175/JCLI4276.1

Thompson, D. W. J., & Wallace, J. M. (1998). The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters*, *25*(9), 1297–1300. https://doi.org/10.1029/98GL00950

Thompson, D. W. J., & Wallace, J. M. (2000). Annular modes in the extratropical circulation. Part I: Month-to-month variability. *Journal of Climate*, *13*(5), 1000–1016. https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2

Thompson, D. W. J., & Wallace, J. M. (2001). Regional climate impacts of the Northern Hemisphere annular mode. *Science*, *293*(5527), 85–89. https://doi.org/10.1126/science.1058958

Tippett, M. K., Barnston, A. G., & DelSole, T. (2010). Comments on finite samples and uncertainty estimates for skill measures for seasonal prediction. *Monthly Weather Review*, *138*(4), 1487–1493. https://doi.org/10.1175/2009MWR3214.1

Tippett, M. K., Ranganathan, M., L'Heureux, M., Barnston, A. G., & DelSole, T. (2017). Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dynamics*, 1–22. https://doi.org/10.1007/s00382-017-3721-y

Wang, L., Ting, M., & Kushner, P. J. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, *7*(1), 279. https://doi.org/10.1038/s41598-017-00353-y

Yu, B., & Lin, H. (2016). Tropical atmospheric forcing of the wintertime North Atlantic oscillation. *Journal of Climate*, *29*(5), 1755–1772. https://doi.org/10.1175/JCLI-D-15-0583.1