AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2014GL060011

Key Points:
- Variability of the Arctic Oscillation is represented in multimodel reforecasts
- Prediction skill of the Arctic Oscillation is assessed in boreal winter
- Prediction skill of the AO is higher under conditions of strong ENSO-AO coupling

Supporting Information:
- Readme
- Table S1
- Figure S1

Correspondence to:
M.-I. Lee,
milee@unist.ac.kr

Citation:
Kang, D., M.-I. Lee, J. Im, D. Kim, H.-M. Kim, H.-S. Kang, S. D. Schubert, A. Arribas, and C. MacLachlan (2014), Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems, *Geophys. Res. Lett.*, *41*, doi:10.1002/2014GL060011.

Received 24 MAR 2014
Accepted 30 APR 2014
Accepted article online 5 MAY 2014

Prediction of the Arctic Oscillation in boreal winter by dynamical seasonal forecasting systems

Daehyun Kang[1], Myong-In Lee[1], Jungho Im[1], Daehyun Kim[2], Hye-Mi Kim[3], Hyun-Suk Kang[4], Siegfried D. Schubert[5], Alberto Arribas[6], and Craig MacLachlan[6]

[1]School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea, [2]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA, [3]School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York, USA, [4]Climate Research Division, Korea Meteorological Administration, Seoul, South Korea, [5]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, [6]Met Office Hadley Centre, Exeter, UK

Abstract This study assesses the skill of boreal winter Arctic Oscillation (AO) predictions with state-of-the-art dynamical ensemble prediction systems (EPSs): GloSea4, CFSv2, GEOS-5, CanCM3, CanCM4, and CM2.1. Long-term reforecasts with the EPSs are used to evaluate how well they represent the AO and to assess the skill of both deterministic and probabilistic forecasts of the AO. The reforecasts reproduce the observed changes in the large-scale patterns of the Northern Hemispheric surface temperature, upper level wind, and precipitation associated with the different phases of the AO. The results demonstrate that most EPSs improve upon persistence skill scores for lead times up to 2 months in boreal winter, suggesting some potential for skillful prediction of the AO and its associated climate anomalies at seasonal time scales. It is also found that the skill of AO forecasts during the recent period (1997–2010) is higher than that of the earlier period (1983–1996).

1. Introduction

The Arctic Oscillation (AO) [*Thompson and Wallace*, 1998], characterized by a periodic exchange of the atmospheric mass field between the Arctic and the rest of high latitudes, is an important mode of climate variability in the Northern Hemisphere. When the Arctic region has anomalously higher atmospheric mass, the negative phase of the AO, the circumpolar jet stream weakens and shifts southward, causing abnormally severe winters in midlatitudes [*Thompson and Wallace*, 2000; *Higgins et al.*, 2002; *Wettstein and Mearns*, 2002]. For example, the negative AO in the winter of 2009/10 was unprecedentedly strong, suggesting a possible role of the rapid retreat of sea ice and the warming in the Arctic. This extreme negative AO was accompanied by severe winter storms over midlatitudes and has been analyzed in a number of diagnostic studies [*Cattiaux et al.*, 2010; *Cohen et al.*, 2010; *Fereday et al.*, 2012]. The profound impact of the AO on the surface climate over the Northern Hemisphere midlatitudes and high latitudes suggests that the accuracy of seasonal predictions in these regions is strongly tied to our ability to predict the AO. This calls for a systematic assessment of the prediction skill of the AO using forecasts made with operational forecast systems.

While the nature of the AO and the physical mechanisms underlying the phenomenon have been extensively studied [*Limpasuvan and Hartmann*, 2000; *Lorenz and Hartmann*, 2003; *Polvani and Waugh*, 2004; *Cohen et al.*, 2010; *Kim and Ahn*, 2012, among many others], studies focusing on the seasonal predictability or the prediction skill of the AO are surprisingly rare in the literature. *Johansson* [2007], *Arribas et al.* [2011], and *Kim et al.* [2012] assessed the forecast skill of the North Atlantic Oscillation (NAO)—a related mode of internal climate variability. They found the skill of NAO forecasts at intraseasonal time scales to be negligible. This contrasts with a more recent study by *Scaife et al.* [2014], which showed that the NAO in boreal winter is highly predictable at several months lead time, using the UK Met Office Global Seasonal forecasting system version 5 (GloSea5). To our knowledge, only one study examined the seasonal prediction skill of the AO exclusively [*Riddle et al.*, 2013]. That study found that the National Centers for Environmental Prediction (NCEP) climate forecast system version 2 (CFSv2) [*Saha et al.*, 2013] is capable of producing skillful forecasts of the wintertime AO for lead times exceeding 2 months. They examined the possibility that the model skill was related to a stratospheric pathway initiated by October Eurasian snow cover anomalies, but

found that these processes were poorly represented in the model. Other studies reported that the forecast of stratosphere-troposphere dynamical coupling, including stratospheric sudden warming and the AO, is improved by including an orographic drag parameterization [*Kim and Flatau*, 2010; *Kim et al.*, 2011].

Motivated by the previous studies, this study evaluates the AO prediction skill for six state-of-the-art seasonal forecast models: GloSea version 4 (GloSea4) [*Arribas et al.*, 2011], NCEP CFSv2, National Aeronautics and Space Administration Goddard Earth Observing System version 5 (GEOS-5) [*Rienecker et al.*, 2011], the Canadian Centre for Climate Modeling and Analysis (CCCma) Coupled Climate Model versions 3 and 4 (CanCM3 and CanCM4, respectively; the data set was obtained from the North American national multimodel ensemble project phase-1), and the Geophysical Fluid Dynamics Laboratory coupled model version 2.1 (CM2.1). These models have been developed independently with quite different formulations and initialization processes. The reforecasts (retrospective forecast or hindcast) from the models are independently initialized with several ensemble members in each month. The different model formulations and initialization processes are likely to affect prediction skill. By carefully examining multidecadal reforecasts produced by each forecasting system, we aim to (1) quantify the current level of AO prediction skill in modern seasonal forecast systems and (2) identify differences in skill between the systems, presumably due to differences in model formulation and initialization processes.

Section 2 describes the data and methodology. The reproducibility of AO variability in the reforecasts is presented in section 3. The prediction skill of the AO in the reforecast data sets is presented in section 4. The summary and conclusions are given in section 5.

2. Data and Methodology

2.1. Reforecast Data Set and Reanalysis

The primary data used in this research are the reforecasts from GloSea4 (1996–2009), CFSv2 (1982–2009), GEOS-5 (1981–2012), CanCM3 (1981–2010), CanCM4 (1981–2010), and CM2.1 (1982–2009). Detailed descriptions and ensemble generation methods of the systems are given in Table S1 (supporting information).

For this study, only ensemble members that were initialized on specific dates between 1 November and 1 December (2 December for CFSv2 and GEOS-5) were used to evaluate the prediction skill of the boreal winter AO. For convenience, the year of reforecast is indicated as the following year from initialization after this paragraph, for example, the 1997 forecast for the run initialized in November or December in 1996. Note that the number of ensemble members is different in the different systems (Table S1 in the supporting information). The numbers of ensemble members used are 15 for GloSea4, 28 for CFSv2, 19 for GEOS-5, and 20 for the others.

For validation, we used the Modern Era Retrospective-Analysis for Research and Applications (MERRA) [*Rienecker et al.*, 2011] atmospheric reanalysis. MERRA has a spatial resolution of 1/2°(latitude) × 2/3° (longitude), with 72 vertical levels. While the GEOS-5 model is initialized with MERRA (which itself was produced with an earlier version of the GEOS-5 atmospheric general circulation model) and so perhaps giving that model an unfair advantage, we found that our results are not sensitive to the choice of reanalysis data sets. Almost identical results for the AO loading vector and the index derived from an empirical orthogonal function (EOF) analysis using sea level pressure (SLP) are obtained using the European Centre for Medium-Range Weather Forecasts global reanalysis, ERA-Interim. The correlation coefficient of the December-January-February (DJF) AO index between ERA-Interim and MERRA is larger than 0.99. Additionally, data from Global Precipitation Climatology Project [*Adler et al.*, 2003] are used to validate precipitation from the reforecasts.

2.2. Methodology

The EOF analysis is performed with seasonal mean (DJF), area-weighted Northern Hemispheric (north of 20°N) SLP anomalies from MERRA for three periods to match the reforecasts periods (1983–2010, 1983–1996, and 1997–2010). The first EOF represents the AO mode. The associated principle component (PC) time series exhibit a large interannual variation of the AO mode. The AO index is defined in this study as the normalized PC time series by its standard deviation. The AO pattern is then obtained by regressing the SLP anomalies onto the AO index. The reforecast data sets are used to assess (1) the ability of the models to reproduce the observed AO pattern and (2) the AO forecast skill of the models.

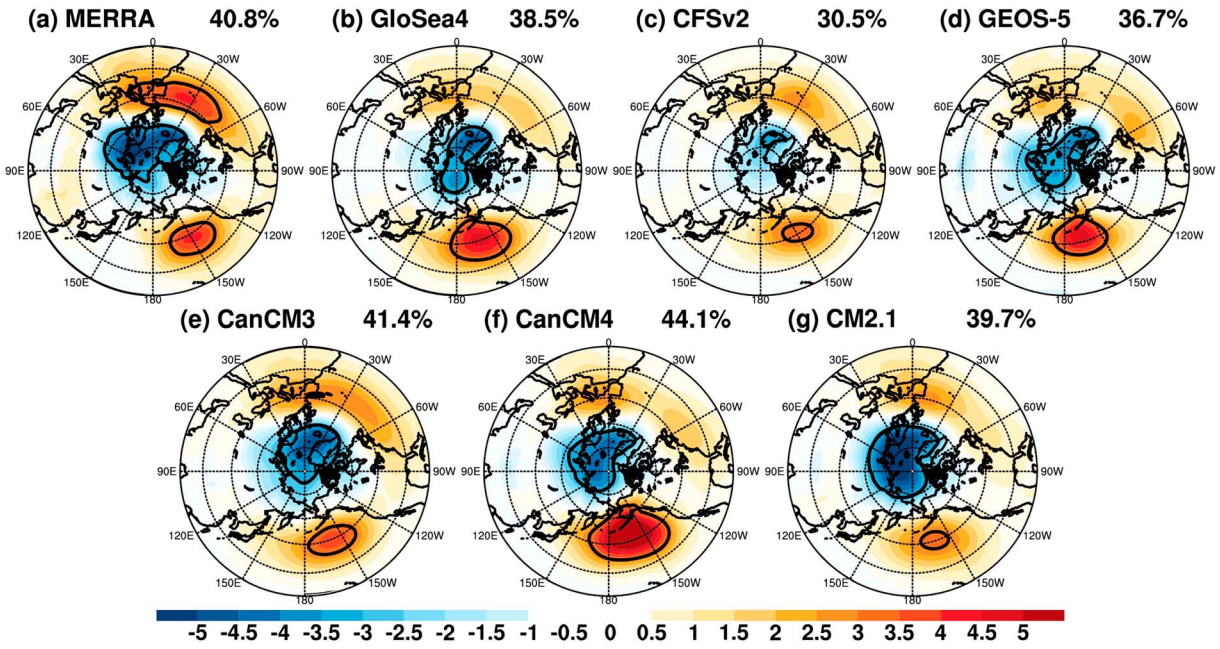

Figure 1. DJF mean sea level pressure anomaly regressed onto the AO index for 1997–2010 for (a) MERRA, (b) GloSea4, (c) CFSv2, (d) GEOS-5, (e) CanCM3, (f) CanCM4, and (g) CM2.1 (unit is hPa). Contour lines indicate 3 hPa and −3 hPa. Percentages indicate explained variance (averaged explained variance from each ensemble member) from the pattern.

In order to evaluate the AO patterns reproduced by the prediction systems, the same EOF analysis was applied to each ensemble member. In most cases, an AO-like pattern emerged as the first EOF. In some cases the second mode was used. This was done if the pattern correlation between the second EOF and the AO pattern from MERRA is higher than that of the first EOF (this never occurred for GloSea4, it occurred once for GEOS-5, CanCM3, and CM2.1, six times for CFSv2, and three times for CanCM4). We compared the AO patterns only for 1997–2010 because all model reforecasts are available for that period. After obtaining the AO mode (i.e., first or second EOF) from each ensemble member, we took an ensemble average of the AO patterns. Anomalous patterns of other variables associated with the AO were obtained by regressing the variables onto the AO index for each ensemble member, and then averaged.

To assess the prediction skill of the AO using the reforecast data sets, either seasonal or monthly averaged forecasted SLP anomalies are projected onto the observed AO loading vector in each period. The resulting ensemble-averaged and individual ensemble time series (i.e., AO indices) are normalized by the ensemble mean standard deviation for each model. We use the ensemble mean AO indices for the forecast skill assessment. Temporal correlation coefficients between the observed and forecast AO indices represent the prediction skill in this study. The persistence forecast using the November AO index provides a baseline forecast for each period, and we consider a prediction skill useful only when it exceeds that of the persistence forecast.

The Relative Operating Characteristic (ROC) score [*Mason*, 1982] is used as a probabilistic forecast skill metric. The ROC score is computed as the area under the ROC curve, which involves the probability of detection (hit rate) and false alarm rate for a particular event. The observed AO indices for each 14 year period are grouped into three terciles. On the basis of the observed tercile threshold, both the observed and forecast ensemble AO indices are categorized as an event or nonevent for each tercile. The ROC scores for the upper tercile event (i.e., positive AO) and lower tercile event (i.e., negative AO) were evaluated with probability thresholds ranging from 0% to 100% with a 20% interval, using forecasted AO indices from ensemble members of each reforecast. In other words, high ROC scores indicate the ability to discriminate tercile events and nonevents successfully in the ensemble members. In general, a ROC score above 0.5 indicates skill better than climatological probabilities.

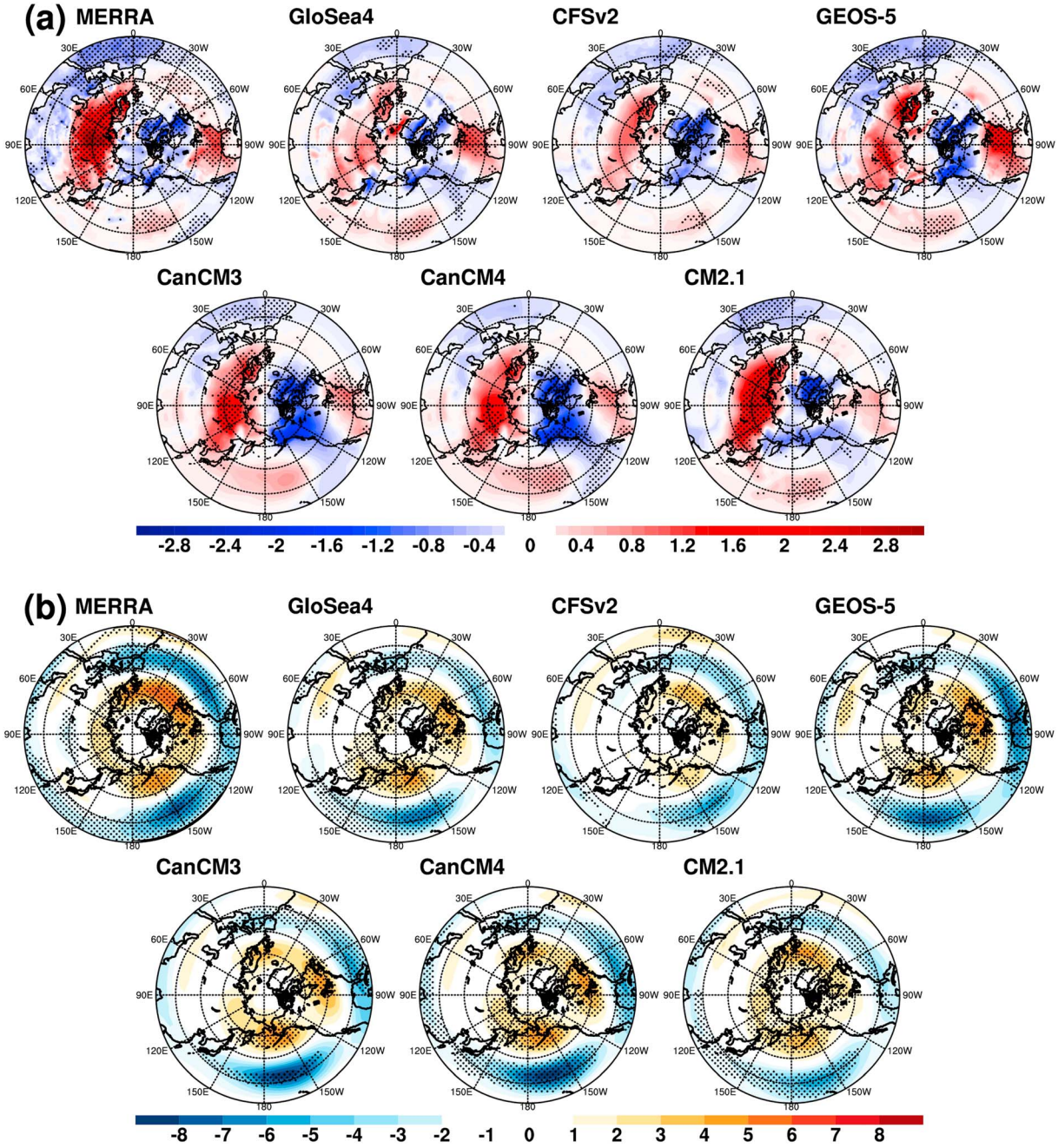

Figure 2. DJF mean (a) surface temperature (2 m temperature for CanCM3 and CanCM4) anomaly (unit is K) and (b) zonal wind anomaly at 200 hPa (unit is m/s) regressed onto the AO index of each forecast for 1997–2010. The dotted grids indicate statistical significance at the 10% level.

3. Model Representations of the AO and Its Impacts

Figure 1 compares the AO SLP patterns represented in the prediction systems and MERRA. MERRA shows a zonally symmetric pattern with clear opposite signed anomalies between the Arctic and the midlatitude oceans (North Pacific Ocean and North Atlantic Ocean). While all prediction systems are able to reproduce this pattern fairly well, they tend to overestimate the variability of the North Pacific compared to that of the North Atlantic. The pattern correlations between MERRA and each forecast have comparable values higher than 0.83. Compared to the other prediction systems, GEOS-5 and CanCM models exhibit more realistic SLP anomaly patterns over the Kara Sea and the northern Siberia.

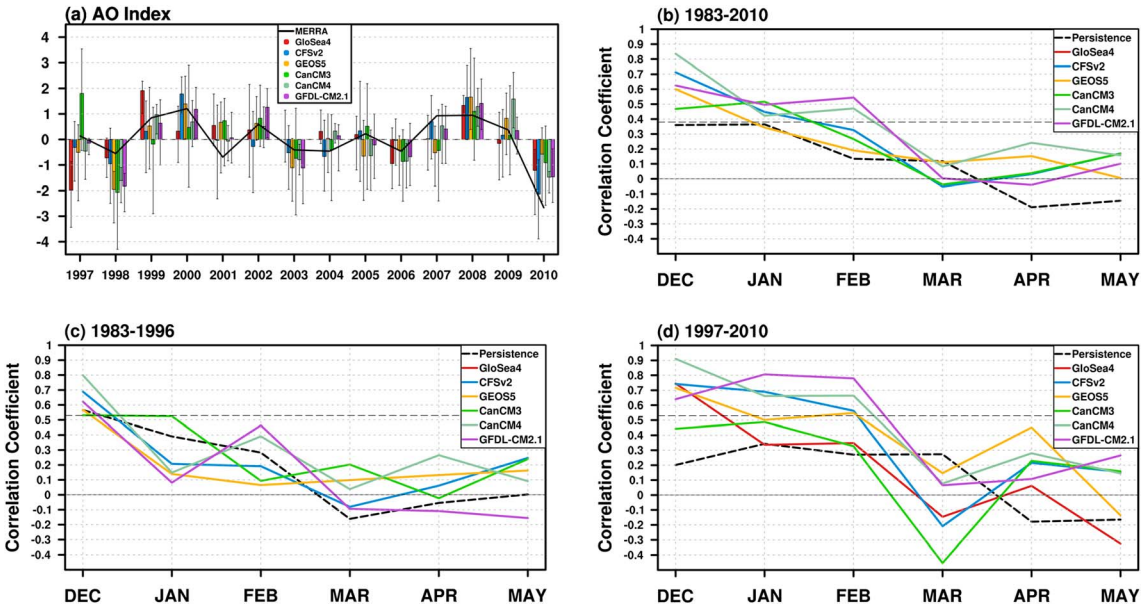

Figure 3. (a) DJF mean normalized AO index of MERRA (black solid line) and reforecasts (color bars). The error bars refer to the ensemble spread of the AO index between 25th percentile and 75th percentile (extreme ensemble spreads are not shown). Correlation coefficient of AO index as a function of forecast lead month for (b) 1983–2010, (c) 1983–1996, and (d) 1997–2010. Black dashed line refers to persistence forecast by the MERRA November AO index for each period, and colored lines indicate prediction skill for each model. Thin horizontal dashed line refers to the 95% confidence level.

The AO mode explains about 37–44% of total interannual variability in the models, except for CFSv2 (31%) which is much lower than the observed value (41%). This might be due to the greater frequency of mixing the AO signal with the second EOF mode.

Spatial patterns of surface temperature, 200 hPa zonal wind, and precipitation anomalies associated with the AO mode from each reforecast are shown in Figures 2 and S1. The north-south oriented patterns of anomalous surface temperature are represented over Eurasia and North America in MERRA (Figure 2a). This surface temperature anomaly pattern is reasonably reproduced in the reforecasts over land, although its amplitude is underestimated mostly in the reforecasts. The amplitude of the temperature variability over Siberia is more realistic in CM2.1 than in the other systems, and this might be linked to the more intense SLP variability over the Arctic and the Kara Sea (Figure 1g). The regressed upper level zonal wind onto the AO index from the forecast systems is consistent with that of MERRA with high statistical significance, describing a realistic modulation by the jet stream corresponding to the phase of the AO (Figure 2b). Nevertheless, there are system-dependent biases such as shifts in the centers of variability that correspond to biases in the SLP variability. Consistent with the jet stream shift, the precipitation is enhanced in high-latitude positive phase of the AO, but the amplitudes of the forecasts are lower than observation (Figure S1 in the supporting information). The forecast systems commonly fail to capture the precipitation anomaly in the East Asia. The overall amplitude of CFSv2 is lower than those of the other systems, possibly due to the frequent substitution of the second EOF mode.

4. AO Prediction

Above results demonstrate that the models are to large extent able to reproduce the observed AO pattern. We next focus on prediction skill. Note that, as described in section 2, we use a single AO loading vector obtained from MERRA for each period rather than those computed from the individual models. The time series of the AO index for the recent period (1997–2010) from MERRA and the reforecasts are shown in Figure 3a. The reforecasts show a reasonable prediction of the seasonal mean AO index. This includes the anomalously negative value in 2010, although most of the reforecasts underestimate the amplitude of the negative anomaly. The ensembles of the prediction systems commonly show a large spread, though they show relatively small spread in some years. Table 1 shows the correlation coefficients between the AO index

Table 1. Correlation Coefficients Between the DJF Mean AO Index From MERRA and Each Forecast[a]

	1983–1996	1997–2010	1983–2010
GloSea4	n/a	0.44	n/a
CFSv2	0.44	0.87**	0.65**
GEOS-5	0.25	0.54*	0.39*
CanCM3	0.62*	0.52	0.53**
CanCM4	0.56*	0.79**	0.67**
CM2.1	0.49	0.80**	0.66**
Persistence	−0.23	0.23	−0.25

[a]Single and double asterisks indicate that the correlation coefficient is statistically significant at the 95% and 99% confidence levels, respectively.

of MERRA and each reforecast. The AO prediction skill from the multimodel ensemble (MME, $r = 0.77$ for 1997–2010) is lower than the skill from the best-performing model (i.e., CFSv2, $r = 0.87$), which implies that the MME is not providing much benefit in this case.

Figures 3b–3d show month-to-month temporal correlation coefficients for December–May along with corresponding results for the persistence forecast. Except for GloSea4, forecasts initialized around November show higher temporal correlation coefficients in winter than persistence for 1997–2010, while the skill of the dynamical predictions do not consistently exceed that of persistence after February. The prediction skill for 1983–1996 becomes comparable to persistence after December, consistent with lower seasonal mean prediction skill during the early period (1983–1996) as shown in Table 1. The reason for the relatively low prediction skill of GloSea4 in January and February is not clear. It might be related to a model bias or it may be due to the relatively small number of ensemble members. GloSea4 shows slightly higher prediction skill when its own EOF pattern is used to derive the AO index ($r = 0.54$ for DJF mean compared to 0.44 in Table 1), which implies that deficiencies in the representation of the AO pattern obscured its prediction skill.

We note that while we find useful prediction skill of the AO in GloSea4, *Arribas et al.* [2011], who analyzed hindcasts made with an earlier version of GloSea4, reported much lower prediction skill of the NAO (which is related to the AO). The hindcast data set used in this study is made with a version of the GloSea4 seasonal prediction system upgraded in several aspects from the original version of *Arribas et al.* [2011]. The changes include improved atmospheric physics, interactive sea ice and its initialization, and enhanced vertical resolution of the atmosphere. A comprehensive description of the difference between the GloSea4 system used in *Arribas et al.* [2011] and that used in the current study is given in *Lee et al.* [2014]. A similar skill improvement in NAO prediction was found in the study of *Scaife et al.* [2014], where they incorporated those upgrades as well as a higher horizontal resolution (0.83 longitude by 0.55 latitude) for the GloSea system (i.e., GloSea5). The prediction skill increase across the GloSea versions suggests that improvements in atmospheric model physics, interactive sea ice formulation and the initialization process, and enhanced vertical resolution to better represent troposphere-stratosphere interaction, may all potentially contribute to improved AO predictions. However, the reasons for any improvements in skill can be highly dependent on the models and initial conditions. For example, increased vertical resolution does not always guarantee improvement in some models. Although better representation of the stratosphere by increasing a model's vertical resolution is likely to improve the seasonal prediction skill of the AO or NAO [e.g., *Marshall and Scaife*, 2010; *Kim and Flatau*, 2010; *Kim et al.*, 2011; *Riddle et al.*, 2013], that is not always the case. For example, CM2.1 produces skillful predictions of the AO ($r = 0.80$ for 1997–2010) with the lowest model top (10 hPa).

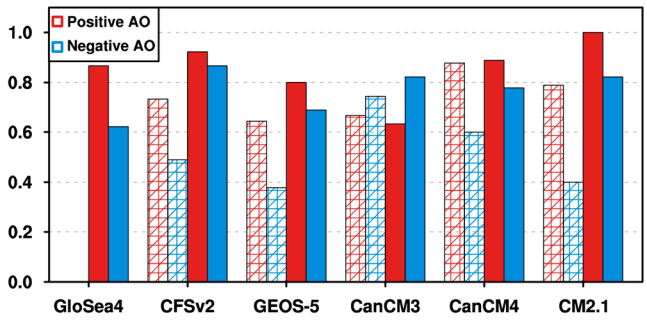

Figure 4. Relative Operating Characteristic (ROC) scores for ensemble AO index prediction for positive AO (upper tercile, red) and negative AO (lower tercile, blue). The checkered bars indicate ROC scores for 1983–1996, and the filled bars indicate ROC scores for 1997–2010.

Almost all reforecasts exhibit significantly higher skill for the recent period of 1997–2010 compared to the earlier period (i.e., 1983–1996). The skill of probabilistic forecasts, consistent with the skill of the deterministic forecasts of the AO index, also show substantial changes between the two periods (Figure 4). Each reforecast shows marginal prediction skill for both positive and negative phases of the AO for 1997–2010 (all of ROC scores exceed 0.6), while the ROC scores for 1983–1996 are mostly lower than those for the recent 14 years. Furthermore,

the ROC scores change little with calibration using cross validation (not shown). Note that the skill enhancement is greater for the negative phase of the AO than for the positive phase (on average, improvements in the scores of the five systems are 0.27 for negative AO and 0.11 for positive AO). This change is interesting because exactly the same model is used for both periods in all prediction systems.

Riddle et al. [2013], who also found this change in the CFSv2 reforecasts, speculated that the difference was caused by systematic errors associated with the initialization prior to 1998. However, given that a similar change in prediction skill occurs in other prediction systems, we need to consider other possibilities, such as changes in the predictability of the AO. Previous work supports this argument. For example, *Li et al.* [2013] suggested a strengthening in the relationship between the AO and the El Niño–Southern Oscillation (ENSO) after the mid-1990s, with possible links to interannual variability of sea ice. The correlation coefficient between the DJF mean AO index in this study and the Oceanic Niño Index of NOAA (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml) is 0.02 for 1983–1996 and −0.59 for 1997–2010. Because ENSO is a lower frequency phenomenon and is predictable up to 12 months in many seasonal prediction systems, a stronger ENSO-AO coupling means a higher predictability of the AO. *Jia et al.* [2009] showed that a strong negative AO-like response was generated by thermal forcing over the central Pacific, while the response to an eastern Pacific forcing is much weaker. And we have also observed more frequent central Pacific El-Niño events in recent decades [*Kug et al.*, 2009; *Lee and McPhaden*, 2010]. Therefore, the changes in the characteristics of ENSO and associated changes in ENSO-AO coupling during recent decades could have contributed to the change in the prediction skill of the AO index. In addition, the Madden-Julian oscillation (MJO) might also influence the tropical-extratropical coupling and intraseasonal prediction skill. For example, the AO phase [*Flatau and Kim*, 2013] and the prediction skill of the NAO [*Lin et al.*, 2010] appear to be influenced by the phase and amplitude of the MJO in boreal winter.

Among the prediction systems, only CanCM3 shows slightly lower prediction skill in the recent period compared to the earlier period. CanCM3 is very similar to CanCM4, though CanCM3 has no shallow convection scheme and has simple radiative forcing [*Merryfield et al.*, 2013]. The above improvement in skill might be associated with improved diabatic forcing in the tropics, propagating to extratropics, which acts as a source of prediction skill for the AO. In general, further studies are needed to identify the reasons for the higher prediction skill of the AO from the dynamical seasonal prediction systems in the recent period.

5. Summary and Conclusion

This study examined the skill of AO predictions using reforecast data sets from six state-of-the-art coupled ensemble prediction systems. The study focuses in particular on wintertime AO predictions using a set of reforecasts initialized around November over multiple years. The prediction systems all include interactive land, ocean, and sea ice components coupled with the atmosphere, although the details of the formulations and the initialization processes are substantially different among the systems. Our results show that most of the seasonal forecast systems used in this study exhibit useful skill in predicting the AO up to 2 months lead time for the recent 14 years (1997–2010). We also found that both the deterministic and probabilistic prediction skills of the AO are higher in the recent years (1997–2010), compared to those in the earlier 14 years (1983–1996). The apparent strengthening of ENSO-AO coupling in recent years provides one possible reason for the higher skill of AO predictions in the recent period.

Our results highlight two aspects of the AO prediction problem. First, seasonal prediction systems are able to reproduce the basic AO phenomenon itself, with high pattern correlations in SLP ranging from 0.83 to 0.95. The forecast systems also produce realistic patterns of anomalous surface temperature, upper level wind, and precipitation associated with the AO, implying that those systems are able to resolve the key physical and dynamical processes associated with the AO. Second, the seasonal prediction systems are able to forecast year-to-year variations of the AO, including the recent extreme occurrences of the AO. The prediction skill does differ among the six systems, and this likely reflects the different parameterizations and initialization processes in the different systems. The considerable spread among the model ensemble members suggests the possibility of future improvements in AO predictions. The reforecast data sets, while allowing us to demonstrate forecast skill, are however not well suited for isolating what aspects of model physics need to be improved in order to further improve AO prediction

skill. Progress on that front will require active experimentation with the individual models (e.g., sensitivity experiments), as well as assessing the quality of the initial conditions and other details of the forecast configurations (e.g., ensemble size).

Finally, we note that the higher prediction skill in recent decades has been found in previous studies for the NAO [*Rodwell and Folland*, 2002; *Bierkens and van Beek*, 2009] and the AO [*Riddle et al.*, 2013]. The short time period over which the prediction skill was evaluated, however, makes it difficult to assess any modulation of the AO from long-term variability such as the Pacific Decadal Oscillation (PDO) and the North Pacific Gyre Oscillation. Therefore, it is currently not possible to predict whether the level of skill found in this study will be same in the future.

Acknowledgments

This study was supported by the Korea Meteorological Administration Research and Development Program under grant APCC 2013–3141. The authors are grateful for the computing resources provided by the Supercomputing Center at Korea Institute of Science and Technology Information (KSC-2013-C2-011).

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

Adler, R., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979– present), *J. Hydrometeorol.*, *4*(6), 1147–1167.

Arribas, A., et al. (2011), The GloSea4 ensemble prediction system for seasonal forecasting, *Mon. Weather Rev.*, *139*, 1891–1910.

Bierkens, M. F. P., and L. P. van Beek (2009), Seasonal predictability of European discharge: NAO and hydrological response time, *J. Hydrometeorol.*, *10*, 953–968, doi:10.1175/2009JHM1034.1.

Cattiaux, J., R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron (2010), Winter 2010 in Europe: A cold extreme in a warming climate, *Geophys. Res. Lett.*, *37*, L20704, doi:10.1029/2010GL044613.

Cohen, J., J. Foster, M. Barlow, K. Saito, and J. Jones (2010), Winter 2009–2010: A case study of an extreme Arctic Oscillation event, *Geophys. Res. Lett.*, *37*, L17707, doi:10.1029/2010GL044256.

Fereday, D. R., A. Maidens, A. Arribas, A. A. Scaife, and J. R. Knight (2012), Seasonal forecasts of Northern Hemisphere winter 2009/10, *Environ. Res. Lett.*, *7*(3), 034031.

Flatau, M., and Y.-J. Kim (2013), Interaction between the MJO and polar circulations, *J. Clim.*, *26*, 3562–3574, doi:10.1175/JCLI-D-11-00508.1.

Higgins, R. W., A. Leetmaa, and V. E. Kousky (2002), Relationships between climate variability and winter temperature extremes in the United States, *J. Clim.*, *15*(13), 1555–1572.

Jia, X., H. Lin, and J. Derome (2009), The influence of tropical Pacific forcing on the Arctic Oscillation, *Clim. Dyn.*, *32*, 495–509, doi:10.1007/s00382-008-0401-y.

Johansson, Å. (2007), Prediction skill of the NAO and PNA from daily to seasonal time scales, *J. Clim.*, *20*, 1957–1975.

Kim, H.-J., and J.-B. Ahn (2012), Possible impact of the autumnal North Pacific SST and November AO on the East Asian winter temperature, *J. Geophys. Res.*, *117*, D12104, doi:10.1029/2012JD017527.

Kim, Y.-J., and M. Flatau (2010), Hindcasting the January 2009 Arctic sudden stratospheric warming and its influence on the Arctic Oscillation with unified parameterization of orographic drag in NOGAPS. Part I: Extended-range stand-alone forecast, *Weather Forecasting*, *25*, 1628–1644, doi:10.1175/2010WAF2222421.1.

Kim, Y.-J., W. Campbell, and B. Ruston (2011), Hindcasting the January 2009 Arctic sudden stratospheric warming with unified parameterization of orographic drag in NOGAPS. Part II: Short-range data-assimilated forecast and the impacts of calibrated radiance bias correction, *Weather Forecasting*, *26*, 993–1007, doi:10.1175/WAF-D-10-05045.1.

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*, 2957–2973, doi:10.1007/s00382-012-1364-6.

Kug, J.-S., F.-F. Jin, and S.-I. An (2009), Two types of El Niño events: Cold tongue El Niño and warm pool El Niño, *J. Clim.*, *22*, 1499–1515, doi:10.1175/2008JCLI2624.1.

Lee, T., and M. J. McPhaden (2010), Increasing intensity of El Niño in the central-equatorial Pacific, *Geophys. Res. Lett.*, *37*, L14603, doi:10.1029/2010GL044007.

Lee, M.-I., H.-S. Kang, D. Kim, D. Kim, H. Kim, and D. Kang (2014), Validation of the experimental hindcasts produced by the GloSea4 seasonal prediction system, *Asia Pacific J. Atmos. Sci.*, doi:10.1007/s13143-000-0000-0, in press.

Li, F., H. Wang, and J. Liu (2013), The strengthening relationship between Arctic Oscillation and ENSO after the mid-1990s, *Int. J. Climatol.*, doi:10.1002/joc.3828.

Limpasuvan, V., and D. L. Hartmann (2000), Wave-maintained annular modes of climate variability, *J. Clim.*, *13*(24), 4414–4429.

Lin, H., G. Brunet, and J. S. Fontecilla (2010), Impact of the Madden-Julian Oscillation on the intraseasonal forecast skill of the North Atlantic Oscillation, *Geophys. Res. Lett.*, *37*, L19803, doi:10.1029/2010GL044315.

Lorenz, D., and D. Hartmann (2003), Eddy-zonal flow feedback in the Northern Hemisphere winter, *J. Clim.*, *16*, 1212–1227.

Marshall, A. G., and A. A. Scaife (2010), Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution, *J. Geophys. Res.*, *115*, D16114, doi:10.1029/2009JD012643.

Mason, I. (1982), A model for assessment of weather forecasts, *Aust. Meteorol. Mag.*, *30*, 291–303.

Merryfield, W. J., W.-S. Lee, G. J. Boer, V. V. Kharin, J. F. Scinocca, G. M. Flato, R. S. Ajayamohan, J. C. Fyfe, Y. Tang, and S. Polavarapu (2013), The Canadian seasonal to interannual prediction system. Part I. Models and initialization, *Mon. Weather Rev.*, doi:10.1175/MWR-D-12-00216.1.

Polvani, L. M., and D. W. Waugh (2004), Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regime, *J. Clim.*, *17*, 3548–3554.

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.

Rienecker, M. M., et al. (2011), MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim.*, *24*, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Rodwell, M. J., and C. K. Folland (2002), Atlantic air-sea interaction and seasonal predictability, *Q. J. R. Meteorol. Soc.*, *128*, 1413–1443.

Saha, S., et al. (2013), The NCEP climate forecast system version 2, *J. Clim.*, *27*, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Scaife, A. A., et al. (2014), Skillful long-range prediction of European and North American winters, *Geophys. Res. Lett.*, *41*, 2514–2519, doi:10.1002/2014GL059637.

Thompson, D. W. J., and J. M. Wallace (1998), The Arctic Oscillation signature in the wintertime geopotential height and temperature fields, *Geophys. Res. Lett.*, *25*, 1297–1300, doi:10.1029/98GL00950.

Thompson, D. W. J., and J. M. Wallace (2000), Annular modes in the extratropical circulation, Part I: Month-to-month variability, *J. Clim.*, *13*(5), 1000–1016.

Wettstein, J. J., and L. O. Mearns (2002), The influence of the North Atlantic–Arctic Oscillation on mean, variance, and extremes of temperature in the northeastern United States and Canada, *J. Clim.*, *15*(24), 3586–3600.