Erin Leghart - Predictability of the Arctic Oscillation and its correlation to Heating Degree Days

Introduction

The leading mode of climate variability in the northern hemisphere winter circulation is the Arctic Oscillation (AO) (Stockdale et al. 2015). Geopotential height anomalies in the Arctic and their resulting winds constitute the basic structure of the AO. The strength of the AO is assessed via the AO index, which indicates the magnitude (strength) and sign (positive or negative) of the AO. A positive AO index is associated with negative geopotential height anomalies over the Arctic, and a negative AO index is associated with positive geopotential height anomalies over the Arctic. The Climate Prediction Center (CPC) constructs the daily AO index "by projecting the daily 00Z 1000mb height anomalies poleward of 20°N onto the loading pattern of the AO" (CPC) (The CPC's AO methodology can be found here: Climate Prediction Center: Teleconnection Pattern Calculation Procedures). This study makes use of the CPC's daily AO index data produced by the GEFS model.

The basic relationship between the AO and northern hemisphere winter temperatures is generally understood. Driven by the geopotential height anomalies, the AO is characterized by westerly winds circulating the Arctic which act to confine arctic air masses to the north. A positive AO leads to enhanced westerly winds, thus confining the cold, arctic air masses to the north. A negative AO leads to weakened westerly winds, which allows for "drainage" of the cold, arctic air masses to the south. Thus a positive and negative AO are associated with warmer and colder winter temperatures in the United States, respectively. These enhanced or weakened westerly winds also affect winter storm tracks and precipitation, leading to enhanced or weakened precipitation in North America, Europe and the Mediterranean. There are many hemispheric impacts of the AO on temperature and precipitation, yet only temperature effects in the United States are assessed in this study.

Due to the AO's subseasonal timescale and its intrinsic influence on northern hemisphere temperatures, it is an important factor to consider when creating a heating degree day (HDD) forecast. HDD is a measure of the difference between the mean temperature and 65°F, which is a measure of how cold a day is. 65°F is used based on the assumption that neither heating nor cooling is used when the average outside temperature is 65°F. HDDs forecasts are frequently used by those in the energy market as a proxy for energy demand, by those in the agricultural market, and any other business that is sensitive to changes in weather beyond a few days.

The goal of this study is to provide forecasters with easy-to-understand and concrete statistical relationships between the AO and HDDs to aid in the formation of HDD forecasts. First, the overall forecast skill of the GEFS AO forecast is assessed, as well as the forecast skill of the GEFS at a 15-day lead time in order to benefit subseasonal forecasters and end users. Secondly, correlations between observed AO indices and observed HDDs, and GEFS AO forecasts and observed HDDs are computed. This is done to provide the forecaster with information on the atmospheric temperature response to an AO event, and to what degree the GEFS AO forecast at different lead times can accurately predict an HDD response. Assessment of GEFS AO forecast skill and the relationship between the AO and HDDs are explored in this study, and few hypotheses to the findings are mentioned but not explored. Suggestions of future work are mentioned throughout.

Methodology

GEFS AO forecasts (courtesy of Kyle MacRitchie at CPC) from January 2006 - January 2020 and daily AO index data from 1950 - present (both from the CPC) are used to assess the forecast skill of the GEFS AO forecast at a 15-day lead time. Mean absolute error (MAE) and "% Hit Rate" (the percent of time a forecast is made correctly under a defined set of conditions) are used to assess the predictability of the AO 15-day forecast for the dataset as a whole. The MAE of an AO forecast of zero, and MAE of a normally distributed random AO forecast are compared to the MAE of AO 15-day forecast to assess the skill of the GEFS forecast. This comparison is done as a function of forecasted AO phase to show the accuracy of the AO 15-day forecast as the forecast trends away from neutral. The deviation of the AO 15-day forecast from the observed AO is illustrated via a box plot, which is shown as a function of observed AO phase in order to see biases in the GEFS forecast.

HDD data used in this study is Maxar's proprietary HDD data, which is a nationally derived HDD. Maxar's HDD data is gas-weighted, thus regions that have higher populations (and thus use more gas for heating) have a higher weight in the HDD calculation than do regions with lower populations. The data spans from August 2013 - February 2019. The relationship between the AO and HDDs was explored only during the winter months (Dec - Feb) of 2013-2019, as winter is the time of year where AO fluctuations are the most important (Thompson and Wallace 1998). The AO and HDD data was standardized in order to quantitatively compare the AO forecasts and AO observations to the HDD data.

This standardized dataset was used to assess the correlation between observed AO and HDDs, and between the AO forecast and HDDs. The correlation between the AO forecast and HDD as a function of AO forecast lead time was computed, which provides the forecaster with an understanding of how early an AO forecast can be used to generate an HDD forecast, without significant loss of correlation between the two. The correlation between observed AO and HDD were computed using HDD data from day 0-15 following an AO event, which provides insight into the atmospheric temperature response to the AO.

Results: Predictability of the AO

Table 1: Basic Skill Assessment of GEFS AO 15-Day Forecast.

GEFS AO 15-Day Forecast Phase (2006 - Jan 2020)	Mean Absolute Error (MAE)	% Hit Rate (% of forecasts that verify in the same phase as the forecast)
Positive	1.04	66%
Negative	1.20	57%

A basic skill analysis shows that the GEFS AO 15-day forecast more accurately predicts a positive AO than a negative AO. This is illustrated by the positively forecasted AO having both lower MAE and higher % Hit Rates than negatively forecasted AO. In this table, % Hit Rate is the percentage of AO 15-Day forecasts that verify in the same phase that they were forecasted to be. Table 1 also shows that overall, the GEFS is not great at predicting the AO at 15-day lead time. Regardless of the phase of the AO forecast, the GEFS over or under forecasts the AO by an index of 1.04 or more.

This 10% difference in hit rate between positive and negative forecasts may signal that the GEFS struggles more when forecasting positive geopotential height anomalies (associated with a negative AO) over the arctic than negative anomalies (associated with a positive AO), thus causing poor forecast skill for negative AO events. Previous studies (Cohen et al. 2018; Cohen et al. 2013; Wang et al. 2017), through qualitative analysis, have linked positive geopotential height anomalies in the arctic to an increase in the occurrence of severe weather in the midlatitudes. Perhaps this link between Arctic temperatures and severe weather in the United States can be a springboard into discovering the cause for this presumed low predictability of positive geopotential height anomalies in the Arctic by the GEFS.

Key Takeaways:

- The GEFS has higher predictive skill for positive AO events at 15-day lead time than for negative AO events

Table 2: MAE comparison between GEFS AO forecast, Forecast of Zero, and Random Forecast by Forecasted Phase.

GEFS AO 15-Day Forecast Phase (2006 - Jan 2020)	MAE: GEFS AO 15-Day Forecast	MAE: AO Forecast of Zero	MAE: Random AO Forecast	Sample Size of GEFS Forecasts
AO > -4 & AO ≤ -3	1.60	2.72	2.75	60
AO > -3 & AO ≤ -2	1.96	1.69	1.85	218
AO > -2 & AO ≤ -1	1.35	1.30	1.52	615
AO ≥ -1 & AO ≤ 1 (Neutral)	1.03	1.02	1.38	3262
AO > 1 & AO ≤ 2	1.19	1.67	2.12	272
AO > 2 & AO ≤ 3	1.50	2.00	2.34	64
AO > 3 & AO ≤ 4	0.89	2.75	2.61	9

Comment: The green colored boxes indicate where the MAE of a AO Forecast of Zero or a Random AO Forecast is less than the MAE of the GEFS AO 15-Day forecast.

The performance of the GEFS AO 15-Day forecast is compared to a forecast of zero, and a normally-distributed random forecast. MAE represents the absolute value of the mean error between the forecasted AO and the observed AO as a function of forecasted AO phase. The random forecast was generated by a random number generator using a gaussian distribution, which used the same mean and standard deviation of the GEFS AO 15-Day forecast in order to mimic the forecast. The green colored boxes indicate where the MAE of an AO forecast of zero or a random AO forecast is less than the MAE of the GEFS AO 15-Day forecast. Computing the MAE of a forecast of zero was done to see if the GEFS forecast over or underestimates certain phases of the AO. Computing the MAE of a random forecast was done to answer the basic question, “if I were to make a random AO forecast, how likely am I to outperform the GEFS?”

The MAE of the GEFS AO 15-Day forecast is lower for positively forecasted AO events than for negatively forecasted AO events, which is consistent with a lower MAE for positive AO forecasts in *Table 1*. Overall, MAE increases as the AO forecast becomes increasingly positive or negative. However, the MAE in the case of extreme AO events (defined here as a forecasted AO event of a magnitude of 3.0+) breaks this trend, as we observe a decrease in the MAE. This indicates that the GEFS is able to more accurately predict extreme AO events than neutral and strong AO events. The threshold for defining an extreme AO forecast of +/- 3.0 is arbitrarily chosen; however, within the AO 15-Day forecast, +/-3.0 is within the upper 0.02% and lower 1.5% of 15-Day forecast, respectively. It is possible that this break in MAE trend is influenced by the sample size of extreme AO events, as there are significantly less days that observed an extreme AO event. It is important to note that changes in the forecasted AO ranges may show a different trend than found here, yet due to time constraints this was not further explored. This may be an important area of future work, as it can illustrate variations in the predictability of the AO by the GEFS 15-Day forecast.

The MAE for a forecast of zero is less than the MAE for the GEFS 15-Day forecast when the GEFS forecast is neutral and -3 < AO Index ≤ -1 (which consists of two phase ranges in *Table 2*). This difference in MAE for neutral AO forecasts is exceptionally small at 0.01, which may be caused by the large sample size of forecasts in this range. The difference in MAE is .05 for AO forecasts in the range -2 < AO Index ≤ -1, which indicates the GEFS forecast is on average 0.05 less (more negative) than the AO verification. The difference in MAE is 0.27 for AO forecasts in the range -3 < AO Index ≤ -2, which indicates the GEFS forecast is on average 0.27 less (more negative) than the AO verification.

These findings indicate that when the GEFS predicts a negative AO at a 15-Day lead time, it tends to underpredict the AO (predicts a more negative AO). For the AO range -3 < AO Index ≤ -2, the MAE of a random forecast is less than the GEFS AO 15-Day Forecast. This indicates that the skill of the GEFS in predicting a strong negative AO event is very poor.

The comparison of MAE between the GEFS AO 15-Day forecast, a forecast of zero, and a random AO forecast further illustrates that the GEFS more accurately predicts positive AO events than negative AO events. With this being said, this further illustrates that the predictability of the AO at a 15-day lead time by GEFS is poor. Even for the neutral forecasted AO phase range, which is the most accurately forecasted with the lowest MAE, the GEFS forecast is on average over or under forecasting the AO by an index of 1.03.

<u>Key Takeaways:</u>

- When the GEFS predicts and AO in the range AO > -3 & AO ≤ -2, it is highly unreliable
- The GEFS predicts extreme AO events (+/- 3.0) better than strong AO events

Figure 1: Box plot illustrating the distribution of the deviation of the GEFS AO 15-Day forecast from AO observation, as a function of forecasted AO phase.

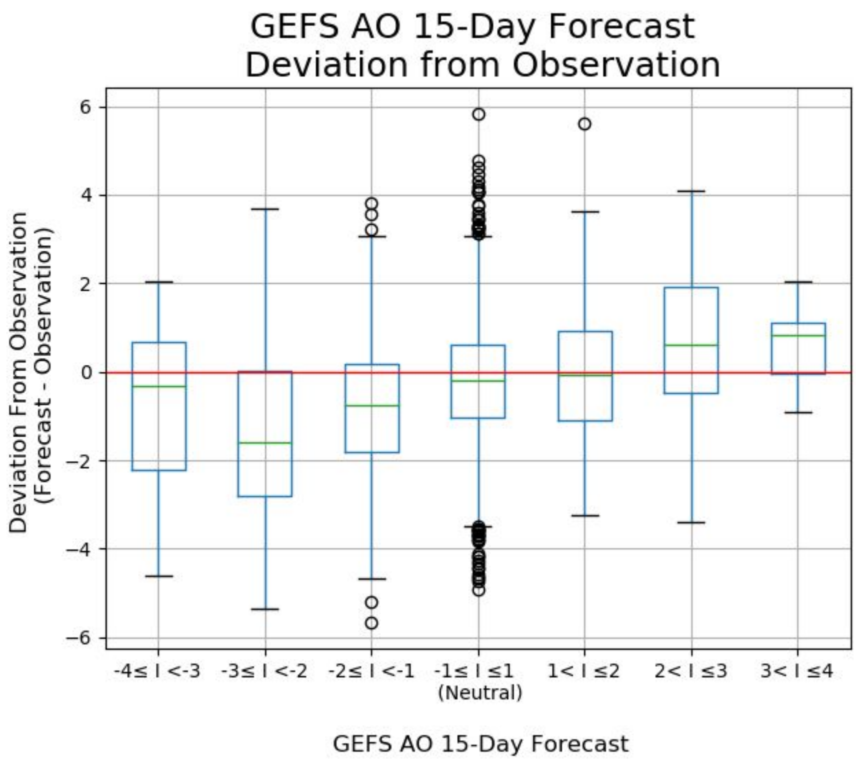


Figure 1 illustrates the trend of the GEFS AO 15-Day forecast by utilizing a box and whisker plot. The y-axis is the deviation of the GEFS forecast from the observation (referred to as "deviation" in the remainder of the study), which is easily computed by subtracting the observed AO from the GEFS AO 15-Day forecast. The line of 0 deviation is highlighted in red, which indicates where the observation equals the forecast. The box and whisker plots are computed for the same phase ranges as those in *Table 2*, where "I" represents the forecasted AO index. The boxes indicate each phase range's median, Q1, and Q3. The whiskers extend to 1.5*IQR (IQR is the interquartile range, which is Q3 - Q1) and the dots represent the outliers.

A clear positive trend is observed in the deviation, as the deviation becomes more positive as the GEFS AO 15-Day Forecast becomes more positive. The median of the negatively forecasted AO ranges is negative, indicating the GEFS underpredicts the AO (predicts a more negative AO than observed). The forecasted range -3 < AO Index ≤ -2 has the largest negative deviation. This mirrors *Table 2*, which also points out the lack of predictability of this forecasted range. In this range the majority of the IQR is contained below 0, indicating that an AO forecast made within this range has a 75% of being underpredicted. The median deviation in this range is nearing -2, indicating that a GEFS AO forecast in this range is almost 2 less than the observed AO. The median of the deviation increases towards 0 as the AO forecast becomes more positive, indicating the predictability of the AO increases as the AO forecast becomes more positive. An almost 0 deviation in the neutrally-forecasted AO & forecasted AO in the range $1 < \text{AO Index} \leq 2$ shows this increasing predictability. The predictability drops again when the AO is forecasted more positive, seen in the forecasted AO range $2 < \text{AO Index} \leq 3$ & $3 < \text{AO Index} \leq 4$. The median of these two forecasted AO ranges is positive, indicating the GEFS overpredicts the AO (predicts a more positive AO than observed) as the forecast becomes more positive. Overall, the GEFS tends to overestimate the magnitude of the AO at a 15-day lead time, regardless of positive or negative phase.

This positive trend in the median is intriguing because it points to a directional discrepancy in the GEFS AO 15-Day forecast. *Table 1* and *Table 2* led to the hypothesis that the GEFS overpredicts positive geopotential height anomalies, thus causing the poor predictability and underprediction of negative AO events. The clear trend in deviation in *Figure 1* builds upon this hypothesis by adding that the predictability of geopotential height anomalies by the GEFS is directionally skewed. Perhaps the GEFS also tends to underpredict negative geopotential height anomalies, leading to overprediction of positive AO events.

Key Takeaways:

- The GEFS 15-day forecast tends to overestimate the magnitude of the AO, in both the positive and negative direction
- GEFS has the lowest forecast skill in the forecasted AO range $-3 < \text{AO Index} \leq -2$
- An AO forecasted in the range $-3 < \text{AO Index} \leq -2$ has a 75% chance of being underpredicted
- Across all phases ranges analyzed, the median deviation is negative for 5 of the 7 phase ranges

Results: Statistical Relationship between the AO and HDDs

Table 3: Winter HDD statistics (Dec-Feb, 2013-2019) to corresponding observed AO ranges

HDD Statistics	AO < -3	AO ≥ -3 and AO < -2	AO ≥ -2 and AO < -1	AO ≥ -1 and AO ≤ 1 (Neutral)	AO > 1 and AO ≤ 2	AO > 2 and AO ≤ 3	AO >3
Count	120	134	244	600	250	113	73
Mean	31.5	29.7	28.7	27.9	28	28	26.4
Minimum	16.5	16.5	14.1	10.8	12.1	14.9	13.9
25%	27.9	24.1	24.2	23.8	23.6	22.7	22.2
50% (Median)	31.2	29.4	28.0	27.8	27.7	27.8	25.0
75%	35.3	34.6	32.8	31.7	32.2	32.4	30.9
Maximum	41.3	45.2	45.0	47.6	45.2	42.8	40.8

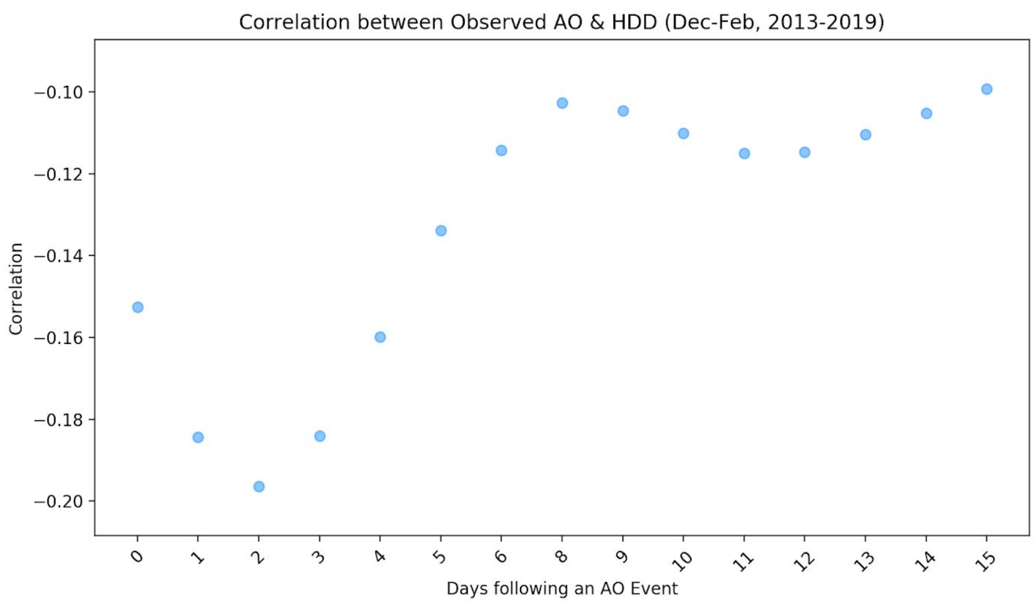
A basic statistical summary of winter HDD data, grouped by the phase of AO that was observed on the same day, was computed to get a brief understanding of how HDDs are affected by an actual AO event. The purpose of this was to see if our data follows the same general trend that positive AO events lead to warmer temperatures, and negative AO events lead to cooler temperatures. It is important to note that these AO ranges are now observed AO events, and no longer GEFS forecasted AO events.

The most telling HDD statistics in *Table 3* are the "Mean" and the "50% (Median)". A slight decreasing trend in the mean and median HDDs is observed as the AO becomes positive, with minimal change between HDDs in the AO ranges -1 ≤ AO Index ≤ -1, 1 ≤ AO Index ≤ 2, and 2 ≤ AO Index ≤ 3. This is consistent with the general understanding that a positive AO results in warmer temperatures, and thus lower HDDs. The minimum and maximum do not follow this general trend, likely because these values, if not outliers, are rather close to the 1.5*IQR threshold for an outlier. These events are likely influenced by other meteorological phenomena other than the AO. These minimum and maximum HDD values show that it is not impossible to record high HDD values during a positive AO, or to record low HDD values during a negative AO. On the contrary, other synoptic synoptic and mesoscale weather systems can have a much larger influence on temperatures than the AO. Systems such as a frigid nor'easter, a warmer system originating in low latitudes, or a midlatitude cyclone can produce temperatures inconsistent with the current AO phase.

Key Takeaways:

- AO data from winter months of Dec-Feb in 2013-2019 follow the general trend of increasing temperatures (decreasing HDDs) with a positive AO, and decreasing temperatures (increasing HDDs) with a negative AO
- Minimum and maximum HDD events are likely influenced by other synoptic conditions, and do not have the AO as the primary driver of temperature

Figure 2: Atmospheric Temperature Response to an AO Event



As previously mentioned, the influence of the AO on winter temperatures in the United States is generally understood. *Table 3* further proves this point with a simple & straightforward statistical analysis. *Figure 2* takes this analysis a step further by seeking to discover when the maximum correlation between an AO event and temperature occurs, and if there is a delayed atmospheric temperature response to an AO event. *Figure 2* is a time series that illustrates the correlation between observed AO events and HDDs from the day of the observed AO (day 0) to 15 days after the observed AO. Negative correlations are expected as HDDs decrease as temperature rises.

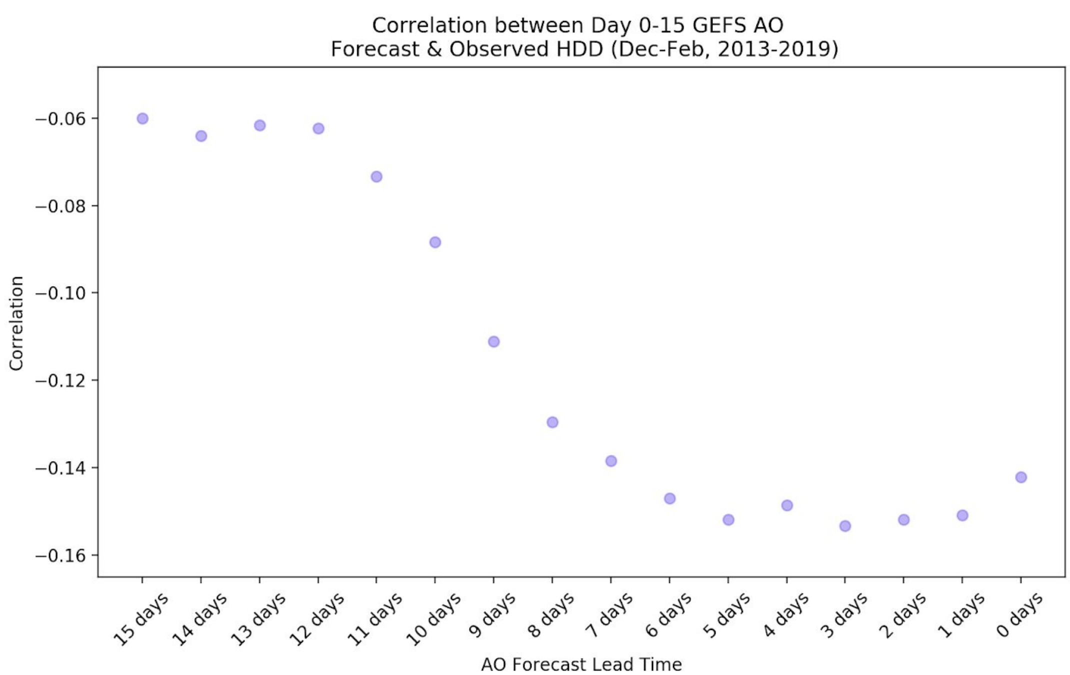
As anticipated, negative correlations are observed from day 0-15. The maximum negative correlation is observed 2 days following an AO event. This is informative for the forecaster, as it indicates that the atmospheric temperature response to an AO event in the United States is delayed by 2 days. This information is extremely useful as it can clarify when one may anticipate the most AO-influenced temperatures, and thus produce HDD forecasts with this in mind. The days preceding the maximum correlation show a steady increase to the maximum, and the days following the maximum correlation show a steady retreat from the maximum. This 2-day delay may be due to the atmospheric temperature response of the shifting of the polar jet northwards or southwards as the westerlies intensify or weaken.

The correlations observed are surprisingly small. It is important to note that although the maximum correlation of -0.2 is small, refining the dataset by removing extreme meteorological events, filtering out days affected by other subseasonal circulations (such as El Nino, which have also been linked to temperatures across the United States), and the use of a larger dataset may lead to a stronger correlation. This further refining of a dataset is an avenue for future studies, which I think will yield interesting results.

Key Takeaways:

- The maximum atmospheric temperature response to an AO event occurs two days after the AO event
- The correlation between the AO and HDDs is surprisingly low. Refining the dataset to filter out other meteorological phenomena may produce a more pronounced correlation

Figure 3: Correlation between GEFS forecasted AO events and observed HDDs



The correlation between forecasted AO events and observed HDDs is explored to shed light on the simple question, "How well will the GEFS AO forecast relate to observed HDDs?" *Figure 3* seeks to answer this question, and furthermore to discover if there is an optimal lead time to rely on the GEFS AO forecast. The purpose of this is to produce a concrete time frame where forecasters can (as accurately as possible) rely on the GEFS AO forecast when providing their end-users with an HDD forecast. The x-axis represents days before an observed HDD event (and consequently an AO event). The y-axis shows the correlation between the GEFS AO forecast at the lead times listed on the x-axis, to the observed HDD. This simple & concrete analysis allows forecasters to produce AO-influenced HDD forecasts that yield the most accurate forecasts at the longest lead time possible.

As anticipated, negative correlations are observed at all lead times. A strengthening of correlation occurs between a 12-day and 6-day lead time, and the correlation is then observed to level off as the lead time decreases. *Figure 3* illustrates that a GEFS AO forecast at a 6-day lead time and a day-of forecast (labeled “0 days”) have almost the same correlation to observed HDDs, with the day-of forecast having a weaker correlation. This is indicative that at a 6-day lead time the GEFS AO forecast is equal to, or more reliable than, a day-of AO forecast when attempting to diagnose an HDD response. One possible explanation for this leveling off of the correlation may be that once the 6-day lead time is reached, GEFS AO forecast does not significantly change thereafter. Thus leaving the correlation to observed HDDs between a 6-day lead time and the day-of seemingly unchanged.

Key Takeaways:

- The predictability of the AO’s influence on United States HDDs is reliable at a 6-day lead time when using the GEFS. The minor changes in correlation between day 5 and day 0 are likely due to changes in the AO forecast or background noise.

Conclusion

The goal of this study was to assess the predictive skill of the GEFS AO forecast, and to analyze the relationship between the AO and wintertime HDDs in the United States. The GEFS AO 15-Day forecast from 2006 - Jan 2020 was found to be more reliable in predicting positive AO events than negative events, which is hypothesized to be caused by the GEFS having less predictive skill for positive geopotential height anomalies than negative geopotential height anomalies. A relationship between wintertime geopotential height anomalies in the Arctic and severe weather in the United States has been identified (Cohen et al. 2018; Cohen et al. 2013; Wang et al. 2017), and could provide insight into how the GEFS can improve its geopotential height forecasts for the Arctic. This bias in predictive skill of the AO is further seen in *Table 2*, through the comparisons of MAE for the GEFS AO 15-Day forecast, a forecast of zero, and a normally-distributed random forecast. Neutral and some negatively forecasted AO events are shown to have higher MAE than the forecast of zero and random forecasts, indicating the poor predictive skill of the GEFS. GEFS AO forecasts in the range $-3 < \text{AO Index} \leq -2$ show the lowest predictive skill, which can also be seen in *Figure 1*. The spread of the forecast error in

Figure 1 shows a trend of the GEFS to underpredict negative AO events and overpredict positive AO events. The analysis of the predictability of the GEFS AO forecast yields the hypothesis that the GEFS has an inherent directional skew in the forecasting of geopotential height anomalies, which in turn produces a directional skew in the prediction of the AO.

Maxar's HDD data was used to verify the relationship between the AO and wintertime temperatures in the United States. A gradual decrease in median and mean HDDs is observed as the AO becomes more positive (*Table 3*). It is further discovered that the peak in atmospheric temperature response occurs two days after an AO event (*Figure 2*). Further work that would be of interest would be to remake the metrics from *Table 3* with a 2-day lag imposed on the HDD data, which I believe will show a stronger linear relationship between HDDs and AO events. In order to assess how well the GEFS AO forecast can diagnose an atmospheric temperature response, the correlation between GEFS forecasted AO events at different lead times to observed HDDs was computed. It was discovered the GEFS AO forecast at a 6-day lead time is equal to, or more reliable than, a day-of AO forecast when attempting to diagnose an HDD response (*Figure 4*). This is very useful to forecasters because it allows them to produce an HDD forecast with confidence, knowing his or her forecast is as reliable as it can be with respect to the influence of the AO on HDDs.

Future Work Ideas

Based on the findings from study, there are many avenues of future work that can provide further insight into the predictability of the GEFS AO forecast, and the relationship between the AO and HDDs. The use of CFS HDD forecasts could be used to further enhance these discoveries, and give forecasters more insight into the predictability of HDDs and their relationship to forecasted AO. To start, a similar analysis to this study can be done to assess the predictability of the CFS HDD 15-Day forecast. Next, a conditional table similar to *Table 2* and *Table 3* could be created to analyze the predictability of the CFS HDD 15-Day forecast based on the forecasted AO phase. This table could show the MAE of CFS forecasted HDDs (the average error between the forecasted HDD and observed HDD) as a function of forecasted AO events. This could possibly show historical trends in the predictability of HDDs, based on what phase the AO is forecasted to be in.

References

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.

Cohen, J., Pfeiffer, K. & Francis, J.A. Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nat Commun* 9, 869 (2018). https://doi.org/10.1038/s41467-018-02992-9

Cohen, J., J. Jones, J.C. Furtado, and E. Tziperman. 2013. Warm Arctic, cold continents: A common pattern related to Arctic sea ice melt, snow advance, and extreme winter weather. *Oceanography* 26(4):150–160, https://doi.org/10.5670/oceanog.2013.70.

Wang, S.-Y. S., Lin, Y.-H., Lee, M.-Y., Yoon, J.-H., Meyer, J. D. D., and Rasch, P. J. (2017), Accelerated increase in the Arctic tropospheric warming events surpassing stratospheric warming events during winter, *Geophys. Res. Lett.*, 44, 3806– 3815, doi:10.1002/2017GL073012.