University of Wisconsin Milwaukee
UWM Digital Commons

Theses and Dissertations

September 2018

Using Self-Organizing Maps as a Forecasting Tool

Andrea Honor
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

 Part of the Atmospheric Sciences Commons

Recommended Citation
Honor, Andrea, "Using Self-Organizing Maps as a Forecasting Tool" (2018). *Theses and Dissertations*. 1826.
https://dc.uwm.edu/etd/1826

This Thesis is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact open-access@uwm.edu.

www.manaraa.com

USING SELF-ORGANIZING MAPS AS A FORECASTING TOOL

by

Andrea Honor

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Atmospheric Science

at

The University of Wisconsin-Milwaukee

August 2018

www.manaraa.com

ABSTRACT

USING SELF-ORGANIZING MAPS AS A FORECASTING TOOL

by

Andrea Honor

The University of Wisconsin-Milwaukee, 2018
Under the Supervision of Professor Paul J. Roebber

Some extreme weather events, such as the early season heavy snow and cold weather outbreak of early November 2014, can be traced back to the influence of tropical or extratropical cyclones on the planetary scale flow. Such planetary scale reorganization also occurs in conjunction with serial extratropical cyclogenesis. Potential temperature on the dynamic tropopause (defined by the 2 PVU surface) allows for a dynamically compact characterization of the flow. NCEP Climate Forecast Systems Reanalysis data spanning 32 years are used to provide this measure, and Self-Organizing Maps (SOM) are then constructed to identify our atmospheric regimes. Key elements of this analysis are the transitions between SOM regimes, which provide means for identifying increased regime predictability at medium and extended ranges. In this study, it was found that 30 regimes defined through the 32 year period were subjectively reasonable for characterizing the variety of hemispheric flow patterns that are observed. The probability of transitions between these regimes over certain time scales (e.g., 10 days, 20 days, and 30 days) was estimated with these same data. This analysis revealed a statistically significant tendency (at the 95% confidence interval) for recurrent patterns at the 30 day lead time, which presents some additional information that may be used in the context of extended range forecasting.

www.manaraa.com

TABLE OF CONTENTS

www.manaraa.com

LIST OF FIGURES

www.manaraa.com

www.manaraa.com

www.manaraa.com

www.manaraa.com

www.manaraa.com

ACKNOWLEDGEMENTS

I would like to thank Professor Paul Roebber for the constant help and mentorship on this research project. I would also like to thank Dr. Sergey Kravtsov for the additional help with coding and ideas with this project. Finally, I would like to thank the UW-Milwaukee atmospheric science department for providing me with teaching assistantship funding, and Dr. Roebber for providing research assistant funding through the NOAA/NGGPS NA16NWS4680018 grant.

www.manaraa.com

I. INTRODUCTION

Northern Hemispheric flow predictability at subseasonal time scales beyond one week depends greatly on interactions between tropical, midlatitude, and polar disturbances. Cool season atmospheric predictability on subseasonal time scales over the Continental United States (CONUS) are critically dependent on the structure, configuration, and evolution of the North Pacific jet (NPJ) stream (Winters et al. 2017). While the NPJ can be perturbed on its tropical side on synoptic time scales from tropical cyclones and on subseasonal time scales by longitudinally varying convection, the North Pacific jet stream can be perturbed on its poleward side due to midlatitude and polar disturbances on a synoptic scale that originate over the Asian Continent. According to Roebber et al. (2016), these midlatitude and polar disturbances in turn act as the catalyst for downstream Rossby wave propagation across the North Pacific, North America, and the North Atlantic.

These downstream Rossby waves that are triggered by midlatitude and polar disturbances are seen as a causation of downstream extreme weather events that can occur over the CONUS. Two specific events acted as the motivation for this project, the first being a major forecast bust that can be traced back to the extratropical transition of Supertyphoon Nuri upstream in the western Pacific in early November 2014. Figures 1-2 show a series of schematic representations of the large scale upper-level flow evolution preceding and following extratropical transition of Supertyphoon Nuri, across the North Pacific and North America between 4-18 November, as first shown by Lance Bosart and collaborators (Bosart, personal communication, 2017). Prior to the extreme weather over the CONUS, Supertyphoon Nuri achieved a minimum central pressure of around 910 hPa just southeast of Japan. Subsequently, this storm underwent extratropical transition after November 6th and then explosively deepened as an extratropical cyclone at a rate

www.manaraa.com

of 60 hPa in 30 hours to a central pressure of 924 hPa. This reintensification was due in part to a cold surge from the Asian continent that had established a strong baroclinic zone across the western North Pacific (Figure 1). Meanwhile downstream over the CONUS, the explosive extratropical cyclone reintensification of former Supertyphoon Nuri triggered an eastward-propagating Rossby wave train and downstream flow amplification. Next, secondary downstream cyclogenesis over the Gulf of Alaska associated with the Rossby wave train resulted in a high-amplitude omega block over western North America by November 11th, with a full omega block observed by November 15th (Figure 2), and a "bent-back" omega block by November 18th (Roebber et al. 2016). Three linked and clustered arctic cold surges subsequently occurred over the conus in conjunction with three polar anticyclones, allowing progressively colder air into the CONUS. During this time period, an early season heavy snow and cold weather outbreak took place, resulting in 150-200 cm of snow in Buffalo, NY, as well as 2,677 minimum temperature records (Figure 3) broken across the CONUS in November of 2014 (Roebber et al. 2016). This event in particular posed significant troubles for forecasters weeks prior to the event, and was not anticipated more than a few days in advance. NCEP CFSv2 2-m temperature forecasts run between 12-21 October and 21-31 October 2014 indicated above average temperatures, while the verification showed below average temperature for everywhere east of the Rockies.

The second example of clustered cyclogenesis occurred over the eastern CONUS between late January to mid-February 2015 and can be seen in Figures 4-5. This event resulted in anomalously strong ridges over the northwestern Pacific and western North America as well as anomalously strong troughs over the Arctic Ocean, northern Canada, the east-central North Pacific and along the east coast of the CONUS. This particular setup favored two storm tracks,

www.manaraa.com

the first being a moisture rich track that extended from the subtropical eastern Pacific poleward toward the Gulf of Alaska then back over the western North American ridge and back southeastward across the CONUS east of the Rockies (Figure 4). The second track extended from northern Russia across the western Arctic ocean southeastward toward the eastern CONUS (Figure 4). Unlike the first track, the second track lacked moisture but had the advantage of enhanced low-level baroclinicity across western Canada and parts of the eastern CONUS (Figure 5). These two favored tracks were important because as they united, they increased the likelihood of disturbances having the capability of reaching the eastern CONUS due to the moisture rich subtropical air from track 1 as well as the enhanced downstream low-level baroclinicity from track 2. This setup led to nearly 250 cm of snowfall in Boston, MA between 24 January and 16 February (Roebber et al. 2016).

Considering these two scenarios, the goal of this study is to study the life cycle of similar weather regimes in order to improve medium-range weather prediction. In past studies work has been done to analyze the physical mechanism of surface cyclone intensification (e. g., Gyakum 1983ab; Bosart 1981; Bosart and Lin 1984), while some more recent studies have considered the relationship between these cyclones and planetary-scale circulation features (e.g., Anderson and Gyakum 1989; Roebber 2009; Cordeira and Bosart 2010, 2011). In these studies, the presence of clustering of surface cyclones during 7-20 day periods is noted, with some focus also being on a week-long period while several meteorological ‘bombs’ occurred over the North Atlantic basin in which this period also followed a record-breaking cold-air outbreak over North America (Turner et al. 2013). In addition, more recent studies have examined significant serial clustering of extreme explosively developing cyclones during late January to early February 1972, after which a strongly negative Arctic Oscillation occurred. During this time, significant anomalously

www.manaraa.com

cold air dominated the northern part of the basin, while anomalously warm air dominated the southern region of the basin. This resulted in a significant zonal jet at 250 hPa across the basin that averaged more than 120 knots from February 1st to the 4th (Roebber et. al. 2016). This jet streak was significant in that two explosively developing cyclones developed on the right entrance region of the planetary-scale jet in a region of near-neutral effect static stability. During this event, a positive feedback cycle was supported between the explosive cyclogenesis and the jet amplification. These types of events are significant, in that they may provide insight into development of a persistent weather regime, or a blocking pattern.

Roebber (2009) found that changes in the strength of zonal flow following cyclogenesis are connected to flow patterns reminiscent of PV filament shedding and tightening of the PV gradient on one hand, and PV wave-breaking anticyclonic roll up on the other. These findings are significant in that they provide opportunities to develop dynamical-statistical forecast products that form the operational basis of this work. Physical processes associated with the relationship between planetary-, synoptic-, and mesoscale processes need to be investigated to determine the dynamical and thermodynamical processes that govern the generation, maintenance, and decay of the planetary-scale jet as the planetary-scale environment that facilitates the life cycles of mesoscale convective systems and synoptic-scale cyclones associated with extreme weather.

As a first step, this study will analyze historical reanalysis data, and categorize individual weather regimes using a machine learning tool called a Self-Organizing Map (SOM). The purpose of doing this is to form the foundation for a suite of hybrid, dynamical-statistical forecast products that will help forecasters to improve weeks two-to-four operational predictions. Considering surface cyclone clustering and other extreme weather phenomena, the goal of this

www.manaraa.com

study is to analyze and categorize phenomena that are directly linked to environmental metrics from synoptic to subseasonal scales. These examples may include surface cyclone clustering, flow regimes associated with blocking, anomalous large-scale flow regimes (e.g. major teleconnection indices such as PNA, AO, and NAO beyond +/- 1 sigma) and MJO-related intraseasonal variability.

Once the atmospheric regimes are classified, we will estimate the transition probability between regimes for a range of time periods. Keeping this in mind, the goal is to be able to better predict regime transitions after the 2- to 4-week time scale, and thereby better predict the probability of extreme weather associated with each regime. Future work will connect regime behaviors to cyclogenesis.

www.manaraa.com

II. DATA AND METHODS

NCEP Climate Forecast Systems Reanalysis (CFSR) data for the 32 year period from January 1979 to December 2010 is used to derive theta on the dynamic tropopause (DT) defined by the 2 PVU surface and theta-e at 850 hPa were used in this study. These data were further restricted to northern hemisphere latitudes from 30-90°N and coarsened from the original 0.5 degree grid spacing to 4 degrees in order to focus on synoptic-to-planetary scale rather than mesoscale features.

The next data treatment step was to normalize these data using standardized anomalies (Grumm and Hart 2001) and to remove seasonality, which was accomplished as follows. First, the standard deviation (σ) was computed:

$$\sigma = \sqrt{\frac{1}{N}\sum(\bar{\theta} - \theta)^2} \qquad (1)$$

where θ is the data variable (such as theta on the DT or theta-e at 850 hPa), and N is the total number of days in each month. The monthly means were also calculated for each variable as:

$$C = \frac{1}{N}\sum\theta \qquad (2)$$

and then the standardized anomalies (k) were computed:

$$k = \frac{\theta - C}{\sigma} \qquad (3)$$

Using standardized anomalies rather than raw variables allows for feature-based comparisons across seasons. Since this method offers comparisons across all seasons throughout the 1979 to 2010 period, however, it does not account for climate non-stationarity.

Next, these data are analyzed using Self-Organizing Maps (SOMs). SOMs allow for the identification of nearly continuous synoptic categorizations, which will be useful when

www.manaraa.com

considering the transitions between synoptic states over the time scales of interest in this study (up to one month). After training, we can inspect the resulting SOM classifications through an objective classification called a Biplot, an enhanced scatterplot that uses both points and vectors to represent structure and similarity via Principal Component Analysis (Young 1999). This plot allows a visualization of regime similarity.

The SOM algorithm uses competitive machine learning to represent the probability density function of a dataset using a two-dimensional grid of map nodes (Gervais et al. 2016). This allows for the classification of large volumes of data into a predetermined number of archetypes that are then categorized based on their similarities. When comparing a SOM over a traditional EOF, there are a few advantages. First, there is a lack of restrictions to orthogonality and stationarity of identified patterns. Processes in the atmosphere are not always represented by orthogonal patterns, and hence a SOM can identify sets of nonorthogonal patterns in data more effectively that may also be more physically relevant.

The main purpose of a SOM is to capture the internal variability of a certain variable(s) over a special scale. When utilizing a SOM, the grid size is user defined, and chosen as a way to represent the distribution required for the study. Many different studies have used a wide range of node numbers and arrangements. For example, Lennard and Hegerl (2015) used a 3 by 4 node SOM (12 nodes) for their study, while Cassano et al. (2015) argued that a larger SOM over a smaller SOM provides optimal interpretability over their region of interest and employed a 5 by 7 node SOM (35 nodes). In our study, a 5 row by 6 column SOM (resulting in 30 regimes) were visually inspected to insure that the regimes captured the breadth of broader scale flow states and the transitions between them. If we were to increase the size of the SOM, smaller scale features would become more dominant in the regime classifications. Conversely if we were to decrease

www.manaraa.com

the SOM size, then large scale atmospheric features would dominate and we would capture less of the internal variability through the period.

When analyzing theta on the DT and theta-e at 850 hPa, specific signatures such as tropospheric coupling can be analyzed. Any sign of a coupling feature is indicative of low static stability leading to more response to a given level of dynamical forcing and thereby increased cyclogenesis. If a specifically classified regime contains a coupling signature, one can better infer that increased cyclogenesis may occur within that current regime. Another advantage of using theta on the DT and theta-e at 850 hPa is one can infer associated thermodynamics and wind fields with a particular PV anomaly (Hoskins et al. 1985).

Once the regimes are defined, the next main step will be to estimate the transition probabilities. In order to do this, we will analyze the frequency of each regime transition from an initial regime to a final regime after an allotted amount of time. For each day throughout the data period, a single regime will be assigned based on its correlation to the mean regime centroids. Therefore, the boundaries for the initial and final regimes is defined as the assigned regime at the initial time step to the assigned regime at the final time step. Using this method, we will analyze the probabilities of regime transitions after 10 days, 20 days, and 30 days. Once these probabilities are calculated, we can test statistical significance using a simple bootstrap statistics test. Understanding if there are favored regime transitions and associated pathways between particular regimes is a fundamental goal of this research, as that understanding can form the basis for a probabilistic suite of forecast tools.

As stated in the previous section, one of the goals of this study will be to analyze and categorize phenomena that are directly linked to environmental metrics from synoptic to subseasonal scales. These phenomena may include surface cyclone clustering as well as

www.manaraa.com

atmospheric flow regimes associated with blocking, but in particular we will look into large-scale flow regimes such as major teleconnection indices such as PNA, AO, and NAO. To do this, we will correlate the daily observed regime assignment (out of our 30 total regimes) with the observed teleconnection indices. Ideally, this analysis will highlight the individual regimes, if any, that correlate well with large scale flow and hence feature dominant large scale features.

www.manaraa.com

III. RESULTS AND DISCUSSION

As stated previously in the Data and Methods section, 30 total regimes were classified using a 5x6 SOM matrix. The main dimensions of the SOM were specified in this way in order to capture large scale features, while also including smaller scale atmospheric features such as intermediate wave scale patterns. When creating these atmospheric regimes using the SOM, we started by keeping the DT data separate from the 850-hPa data by creating a SOM strictly with the DT data, then another separate SOM with the 850 hPa data. Once that was completed, we created a SOM using the DT data overlaid with the 850-hPa data. When comparing the overlaid data regimes with the DT data regimes, it was found that the 30 regimes for all three separate SOM maps were nearly identical in structure. The similarity between the DT SOM maps and the 850 hPa maps can be seen as an indication of scale. The previously discussed coupling is most expressed at synoptic and sub-synoptic scales, and therefore the DT and 850 hPa data maps on to each other. Given the focus of this research on planetary scale wave transitions, hereafter we define regimes using only the DT data.

After defining the 30 regimes in this way, the next step was to analyze the frequency of a regime transition, or transition probability (TP). When analyzing TPs, we looked at 3 different time scales: 10 day lead time, 20 day lead time, and 30 day lead time. Starting with the 10 day transition probabilities as shown in Figure 6, the initial regime is numbered on the x-axis while the y-axis displays the final regime, where each corresponding box represents the TP from the initial regime to the final regime. Looking closely into the 10 day TPs, we can see that the highest TP is from regime 1 to regime 30 with a TP of 10%. The next highest TP was found to be around 8.1% with a transition from regime 23 to regime 24. Following those higher percentages, there is no clear pattern or grouping found in areas with higher or lower TPs.

www.manaraa.com

The next step was to analyze the TPs on a longer time scale such as 20 days. Referring to Figure 7 we can see similar characteristics in TPs as what was seen in our 10 day TPs. That is, there is no significant grouping or significantly high TPs to note. The highest probabilities found for the 20 day time range peaked at 8.1% with a transition from regime 2 to 24, with the next highest at 7.4% for the transition from regime 11 to regime 30. Overall the main data percentages ranged from 1% to 5%.

The last time scale that was analyzed for the TPs was 30 days. Referring to Figure 8, we can see an increase in higher TPs. For example, for 6 total transitions, the TP was greater than 8%. For the 30 day time period, we can also spot individual groupings. Starting on the left side of the diagram, there are clear higher probabilities for the transitions from regimes 1 to 9, 1 to 10, 1 to 18, 2 to 13, 3 to 6, 3 to 14, 4 to 17, and 6 to 10. In addition to these individual higher TPs, we can see an interesting tendency for recurrence with higher TPs for regimes 23 to 27. This result is distinct from the previous time scales discussed (10 day and 20 day). These higher TPs were also found to be statistically significant. A bootstrap statistics test for the 30-day TPs found numerous high and low transitions that are statistically significant at the 95% confidence interval (Figure 9).

As a check on climate non-stationarity (subject to the limitations of the period-based mean and standard deviation calculations previously discussed), we also considered temporal changes in the regime frequencies of occurrence. For example, if the year 1994 had a higher frequency of regime 14 occurring, then there may be some hint of some potential climate influence during that time period. Upon analyzing this temporal frequency, we did not detect any non-stationarity in climate that shifted individual regime frequencies (Figure 10).

www.manaraa.com

The next step was to visually see how these regimes relate to one another spatially by using a Biplot. A Biplot uses points to represent the scores of the observations on the principal components, and it uses vectors to represent the coefficients of the variables on the principal components. On the x-axis, which represents the first principal component, the data is aligned orthogonal to this axis and positioned to represent the maximum/minimum variation through the data. Then along the y-axis, which represents the second principal component, the data is aligned to represent the next highest variation through the data. In our case, both axis' will represent the orientation of the data spatially. The x-axis will represent the varying differences in sign of the standard anomalies by latitude. For example, if one regime resembles a positive anomaly dominating the arctic while another resembles opposite characteristics with a negative anomaly dominating the arctic, then these two regimes will be on opposing ends of the x-axis. The y-axis then resembles the orientation and location of the PV anomalies longitudinally in the meridional flow patterns. For example, if one regime were to resemble a strong positive anomaly somewhere over North America with a strong negative anomaly over Europe and Asia, while another represents opposite characteristics with a strong negative anomaly over North America and a positive anomaly over Europe and Asia, then these two regimes would be on opposite ends of the y-axis. The relative location of points corresponds to similarities or differences, i.e. the closer the points the more similar the components are.

Referring to the BiPlot in Figure 11, the relationships between the regimes can be noted. Starting with the right side we see regimes 2, 3 and 4 as the right-most regimes. In Figures 13-15 one can note that these regimes are very linear with a positive anomaly dominating the pole and negative anomalies to the south. Conversely if we refer to the regime 28 circle, we can note in Figure 39 that this regime is a mirror image of regimes 2, 3, and 4, where a negative anomaly

www.manaraa.com

dominates the pole with positive anomalies to the south. This confirms that the x-axis represents the latitudinal orientation of the negative and positive anomalies. Meanwhile the y-axis represents the longitudinal variation of the sign of the PV anomalies. We can spot this by comparing regime 1 and regime 5. Referring to Figure 12, regime 1 features a deep positive anomaly stretching from the pole down to northern Canada and Alaska while a deep negative anomaly dominates east of Greenland and north of Europe. Regime 5 however as noted in Figure 16 displays differing features. A negative anomaly can be spotted over northern Canada while a positive anomaly can be spotted east of Greenland and north of Europe, opposite than that of Regime 1. This confirms that the y-axis orientation and location of the PV anomalies longitudinally in the meridional flow patterns.

Now understanding how the orientation of the BiPlot relates to regime spatial relationships, we can infer the similarities between the regimes with the high TPs from the 30 day lead time that were statistically significant. Starting with regime 1 transitioning into 9, 10, or 18, we can first spot that regimes 9 and 10 are close to a one another on the BiPlot while regime 18 is slightly more to the left. Interestingly, all of these regimes are distant from the initial regime 1, suggesting that there is still a substantial difference in spatial orientation of these regimes. When looking at the regimes with a recurring pattern (23-27) there are some similarities suggested by the BiPlot. In this chart, regimes 23-27 all lie on the left side of the chart. While there is less variation in the x-axis (regime 25 is more neutral while regime 23 is more negative), the greater differences are spotted along the y-axis where 23, 27, and 26 are positive while 24 and 25 are negative. The greatest similarities can be spotted with regimes 23, 26 and 27 as they are all positioned in the positive y, negative x quadrant together. Altogether, this BiPlot suggests stronger similarities for this recurrent flow group than that of the 1-9, 10, 18 group.

www.manaraa.com

Given that there are recurrent flows, it is of interest to determine whether there are preferred pathways or "tracks" that are followed during the 30 day period. In other words, if an initial regime is 23 and after 30 days, the flow returns to 23, what are the regimes that take place in between these times, and are these transitions consistent from one case to the next? In order to test this, a subset was taken from the data only including cases where the initial regime was within regimes 23 to 27, and ended at any regime in that same grouping. Referring to Figure 42, our initial regime is denoted as Initial and is in blue representing the starting point at the bottom row of the graph. As we move upwards each bar represents the next 5 days, i.e. pink represents the spread of the new regimes after 5 days, light green represents the spread of the new regimes after 10 days, etc, until we reach the 30 day mark represented in purple at the top of the chart. It is evident that there is no preferred track or pathways for these transitions on the way to recurrence. As denoted by each boxplot for each transition time step, there is considerable spread for each case. We can however see some hinting at groupings for the 5-day and 25-day time steps, where there appears to be a higher frequency of transitions in the 23 to 27 regime group.

In a study done by Plaut and Vautard (1993), a multichannel version of the singular spectrum analysis (M-SSA) is used to identify dynamically relevant space-time patterns, much like the way a principal component analysis identifies the spatial patterns dominating variability. Three major low-frequency oscillations were found where the periods ranged from 70 days, 40-45 days, and 30-35 days. It was found by Plaut and Vautard that in particular, the occurrence of the Euro-Atlantic blocking regime was strongly favored by the particular phases of the 30-35 day mode which were found to be naturally confined over the Atlantic region. When looking at the transitions of regimes, it was found that a transition from a zonal to blocking regime was highly connected to the life cycle of the 30-35 day mode. This was found to indicate that regime

www.manaraa.com

transitions do not result only from random occurrence of particular transient eddy forcing, and that there are preferred paths between weather regimes. This finding by Plaut and Vautard (1993) supports our findings of recurrent weather regimes after 30 days. Considering these results, the potential for long-range forecasting for large-scale environment-favoring weather regimes has the potential to be improved, particularly 30 days in advance. As stated, the phases of the 30-35 day mode found by Plaut and Vautard (1993) were found to be confined over the Atlantic region. Interestingly, the North Atlantic region coincides with the location of the peak anomalies in most of our recurring regimes (23-27), an important exception being regime 27 where the peak anomaly is in the North Pacific. This latter result may suggest a connection to North Pacific blocking but this is an aspect beyond the scope of the current study. Plaut and Vautard (1993) found that the conditional probability of the occurrence of blocking is enhanced with a 30 day lead time relative to climatological probability. This further supports the value of our findings for significant transition probabilities up to 30 days in advance.

Simply based on how Figures 12-41 are structured, one may wonder if any of these regimes are related to teleconnection indices. For example, regime 11, as seen in Figure 22, appears to show characteristics of a positive Pacific/North American (PNA) pattern suggesting that relationships to teleconnection patterns could be possible in some of these regimes. Correlations with three separate teleconnection indices were analyzed, including the Arctic Oscillation (AO) and the North Atlantic Oscillation (NAO) in addition to the PNA teleconnection. Throughout the data spanning 32 total years, each day had an observed teleconnection index for each of the three teleconnections analyzed, as well as a regime that was assigned based on the individual day's correlation to the mean regime centroid. We then compared the individual days that had assigned regimes ranging from 1 to 30 to the observed

www.manaraa.com

teleconnections based on those assigned regimes. The overall intent was to find any positive or negative index trend for any of the individual regimes. Starting with the PNA index compared with the original data as seen in Figure 43, there was no such correlation found. In this Figure, the spread of the observed PNA with each assigned regime grouping was found to be large, indicating that there was no discernable strong relation to the PNA index with the raw data. We performed this same analysis with the AO and NAO indices, as shown in Figures 44 and 45, and found similar results where no discernable strong relation between these regimes and the AO or NAO teleconnections patterns was found. This result was not entirely surprising, particularly when you consider what each individual day PV anomaly field looked like.

For example, considering Figure 46 we can see the observed PV anomalies on 28-30 June 2010. On these days, similar to all other days in the data, there was a lot of noise spatially, and therefore no strong and discernible relationship to the characteristic positive or negative teleconnection patterns. In order to get a better sense as to the correlation between regime groups and observed teleconnection indices, filtering was done on the individual days. With the primary goal of filtering out the days in the data with small scale noise, we looked at the correlation of each day with the 5 Leading Principal Components (PCs). If the correlation of the 5 Leading PCs to an individual day was greater than or equal to 0.6, that individual day was included, otherwise all other days with correlations of less than 0.6 were excluded. This filtering alone helped cut out the noisy day data which excluded regime groups 9-15, 17-26, and 30, indicating that the remaining regimes have characteristic dominating large scale features. Starting with PNA as seen in Figure 47, we start seeing more indications of positive or negative correlations. For example, the regime 1 group displays a positive PNA index correlation, while the regime 5 group displays a negative PNA index correlation. While some individual regime groups displayed positive or

www.manaraa.com

negative correlations, others, such as regime group 2, still had a bit of noise and didn't suggest any particular relationship to a PNA pattern. Similar findings were seen for the NAO and AO teleconnection pattern correlations. Figure 48 displays the same data as Figure 47 but with the spread of the AO and NAO teleconnection patterns in addition to PNA for comparison purposes. Similar to the PNA pattern, the AO and NAO patterns show signs of correlating with individual regime groups. For example, regime group 16 shows signs of a correlation with positive NAO and AO while a negative PNA pattern is hinted. While this particular group has distinct relationships between the teleconnections displayed, other regime groups still have a bit of noise associated with the teleconnection patterns. For example, looking at regime group 1, we see an association with a positive PNA index, however while that particular pattern has a more discernible correlation, the other two teleconnection patterns still show quite a bit of spread and noise. This indicates that while some regime groups, such as group 16, appear to be a possible combination or interaction between teleconnection patterns, others, such as regime group 1, may be more associated with one teleconnection pattern and not others. While much of the filtered data still did not show any discernible relationship, there are many other teleconnection patterns to consider that were not analyzed in this study that may be related to those other regime groups.

While it's clear that many of the regimes did not have strong connections with teleconnection patterns, previous research by Vigaud et al. (2018) proved that atmospheric weather regimes on a sub-monthly time scale may have clear connections to large-scale teleconnection patterns. In their study, day-to-day variability in atmospheric circulation over North America is examined during the extended winter season using daily 500 hPa geopotential from 1982 to 2014 with the primary focus on recurrent weather regime reproducibility and predictability using ECMWF ensemble forecasts. In their study four weather regimes were

www.manaraa.com

classified using a K-means clustering technique on daily Z500 fields from MERRA data. Of those four regimes, regime 2 was found to correspond to the negative phase of the NAO, while regime 4 closely resembled the negative phase of the PNA pattern. Of those two, regime 2 by Vigaud et al (2018) correspond well with our regimes 5 and 16 in terms of correlations with the PNA and NAO indices. That is, our regimes 5 and 16 corresponded with negative NAO indices and positive PNA indices. Interestingly enough, Vigaud et al. (2018) found that regimes 2 and 4 corresponded well with the variability of the mean position of the jet over the North Atlantic. While this resembles somewhat well with our regime 5 (Figure 16), our regime 16 (Figure 27) appears to resemble the primary modes of variability in the mean position of the jet over the North Pacific Basin. One important factor to note in these results compared to Vigaud et al (2018), is the comparative number of regimes. As stated previously, an increased SOM size results in greater amount of internal variability to be realized in each regime. Since Vigaud et al (2018) used 4 regimes to classify atmospheric variability as compared to our 30, it should be assumed that their 4 regimes contain the more of the dominant large scale features with natural correlations with teleconnection patterns. Any realization in a teleconnection pattern is important in an atmospheric regime, as Vigaud et al (2018) found evidence of accuracy in model's deterministic forecasts in the 1-7 day range, where an ensemble mean proved best while also providing a more robust representation of large-scale teleconnection patterns.

While Vigaud et al (2018) showed that forecast skill can be improved on a shorter time scale from 5-20 days due to correlations with teleconnection patterns, additional insight can be added to 30 day lead times when considering the high transition probability groups found, including the recurrent patterns that were found to be statistically significant. As observed in Vigaud et al (2018), for week 1 and 2 lead times, regime counts were forecasted with little bias

www.manaraa.com

for the extended winter season. However, by the week 3 lead time and beyond, significant differences began to appear between the observed and forecasted monthly counts. In this case, ECMWF forecasts were underestimating days spent in their regimes 1 and 3 and overestimating days spent in their regimes 2 and 4. This may reflect the lesser predictability of intermediate scale wave pattern after 2 weeks compared to larger scale teleconnections. This is an area where our smaller scale regimes appear to have an advantage. While our 30 regimes lack in regards to a strong correlation with large scale teleconnection patterns compared to Vigaud et al (2018), our regimes capture more internal variability and therefore can detect intermediate scale wave patterns. Referring back to our transition probabilities discussed previously, our transitions with higher percentages increase in number and magnitude as we near the 20 day lead time and especially the 30 day lead time. These transition probabilities provide insight on preferred transitions after 20 or 30 days.

Agel et al. (2018) performed a very similar study in which large-scale meteorological patterns associated with extreme precipitation in the northeastern CONUS were examined. This study included very similar methods, including the use of a 5x6 SOM to organize the patterns that may be related to extreme weather by using pressure anomalies on the Dynamic Tropopause. In that study, the extent to which large-scale meteorological patterns can "explain" extreme precipitation is addressed, and whether there were any distinguishing features associated with each SOM pattern that occur on extreme-precipitation days. To do so, the mean correlation of individual days (both extreme precipitation and non-extreme precipitation days) to their assigned SOM pattern was calculated. Of the 30 total SOM patterns created by Agel et al. (2018), the spatial correlation was slightly higher on extreme days than on non-extreme days for only 4 of the 30 SOM patterns and were found to be significant at the 0.05 significance level. These SOMs

www.manaraa.com

consisted primarily of trough patterns, however there were some ridge patterns that also had higher correlations on extreme days but these were not found to be statistically significant at the 0.05 level. There was also one SOM pattern that had fewer extreme precipitation days than non-extreme precipitation days, where the correlations were also found to be small on extreme precipitation days. Understanding how regime patterns are associated with extreme weather is another vital piece to this puzzle. While in our research the use of transition probabilities is addressed to better predict regime transitions, Agel et al. (2018) proposes a practical approach to further improving forecasting skill through associating each regime with a measure of extreme weather probability. In addition to being able to better predict the onset of extreme weather, we may better predict the persistence of extreme weather through the use of SOMs. According to Agel et al. (2018) nearly every SOM pattern of their 30 featured extreme precipitation days with a longer mean pattern duration than that of the non-extreme precipitation days, however only 2 of the 30 SOM patterns were found to be statistically significant based on a Monte Carlo approach. This too provides additional value to the use of a SOM as a forecasting tool. Altogether understanding which regimes have a tendency to produce significant extreme weather for a longer duration holds value, and combining that knowledge with an understanding as to which regimes have a higher probability of transition provides the potential for a valuable forecasting tool.

www.manaraa.com

IV. SUMMARY AND CONCLUSION

This study aimed at documenting and capturing atmospheric variability over North America through the years 1979 to 2010. This was done through a machine learning tool called a Self-Organizing Map, in which 30 total regimes were created using standardized anomalies of Theta on the Dynamic Tropopause. Of the 30 total regimes, only 11 of those were found to have a correlation of 0.6 or greater to the 5 Leading PCs, indicating that these Regimes in particular are the more robust regimes. When comparing these regimes to teleconnection indices, it was found that there was even less correlation to large scale teleconnection patterns. However, some regimes hinted at some correlations to the NAO, PNA or AO indices indicating that while most of the regimes have too much internal variability to show any large-scale relationship, there are a few such as regimes 1, 5, 8, 16, and 27 that hint at some positive or negative correlation. These results agree somewhat with Vigaud et al (2018), however in their study they only had 4 regimes classified compared to our 30, suggesting that their 4 regimes naturally have dominant large scale features embedded.

In regards to predictability and transition probabilities, it was found that from 10 day lag, 20 day lag, and 30 day lag, our peak transition probability was around 10%. Starting at 10-day lead time, there was only one transition that peaked at 10%. Similarly, the 20-day lead time transition probabilities showed some peak percentages for transitions at 8.1% while the bulk of the transitions ranged between 1-5%. Apart from those higher percentages for the 10 day and 20-day lead time probabilities, the rest of the probabilities were insignificant compared to chance between the different regimes. When analyzing the 30-day lead time probabilities, larger percentages were found more frequently, with one grouping of recurrent flow also found. These probabilities were found to be statistically significant on the 95% confidence interval based on a

www.manaraa.com

bootstrap statistics test. The recurrent regime probabilities found coincide with results found by Plaut and Vautard (1993) where similar preferred paths and recurrent regime flows were found on phases of a 30-35 day oscillation.

When testing for relationships between the regimes and teleconnection patterns, it was found for the daily raw data that there was very little discernible relationship between the observed NAO, PNA, or AO indices and the observed regime. After filtering the data to only days with correlations to the 5 Leaking PCs up to 0.6, we began to see some relationships emerge. For instance, regime 16 corresponded with a positive NAO and AO index and a negative PNA index. However, even when performing the filtering, many of the remaining regimes had too much noise and variation to show any clear teleconnection relationship.

Overall, the Self-Organizing Maps provide a beneficial tool in forecasting. Through this research, we were able to utilize the frequency of each SOM to calculate the transition probabilities for each respective regime for as long as 30 days out. Combining that knowledge with the ability to associate each regime with observed extreme weather as Agel et al. (2018) showed brings on the possibility for improved guidance for extreme weather as far as 20-30 days in advance.

www.manaraa.com

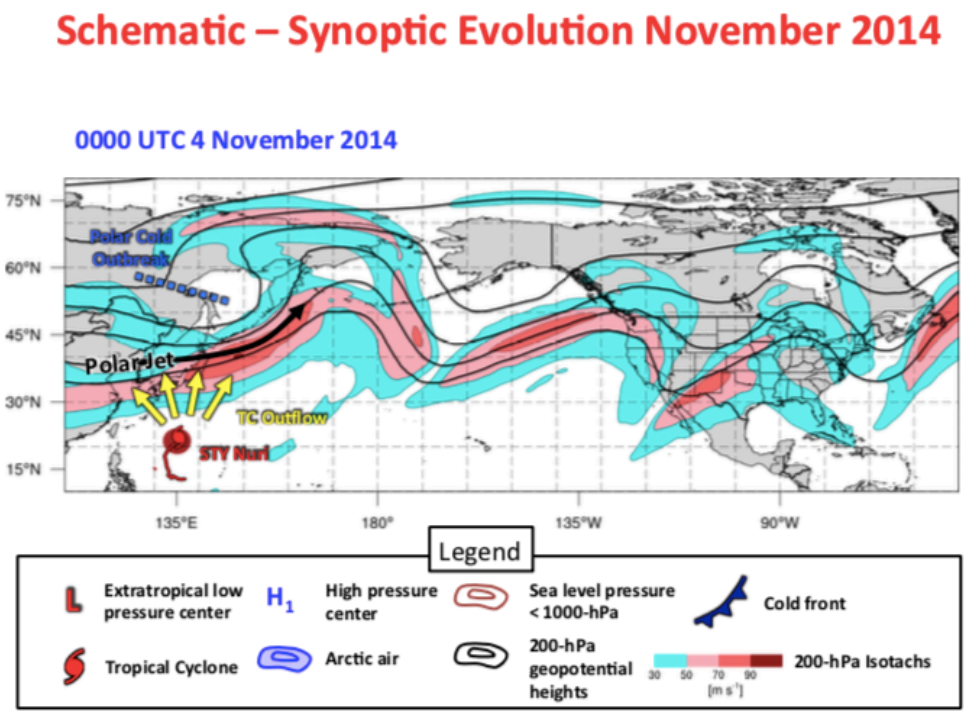


Figure 1. Schematic map of the observed large- scale flow structure across the North Pacific and North America prior to the extratropical transition of Supertyphoon Nuri at 0000 UTC 4 November 2014. Symbols as shown on the legend. Source: Bosart (2017)

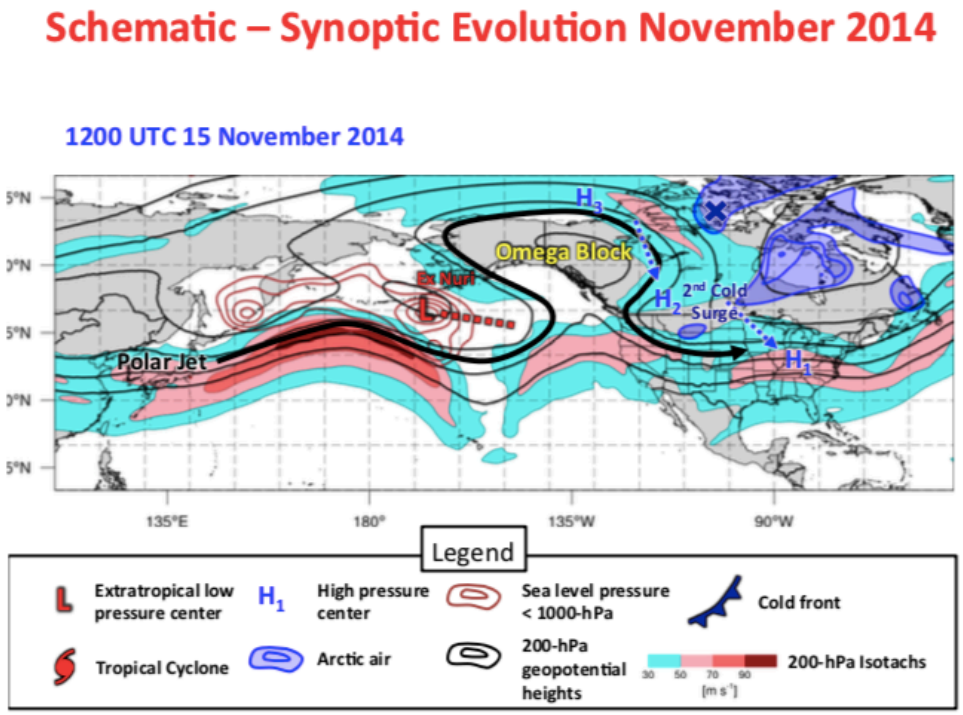


Figure 2. As in Fig. 1 except for the established high-latitude omega block at 1200 UTC 15 November 2014. Legend is as in Fig. 1. H1, H2, and H3 denote the locations of the anticyclones associated with three cold surges. Source: Bosart (2017)

www.manaraa.com

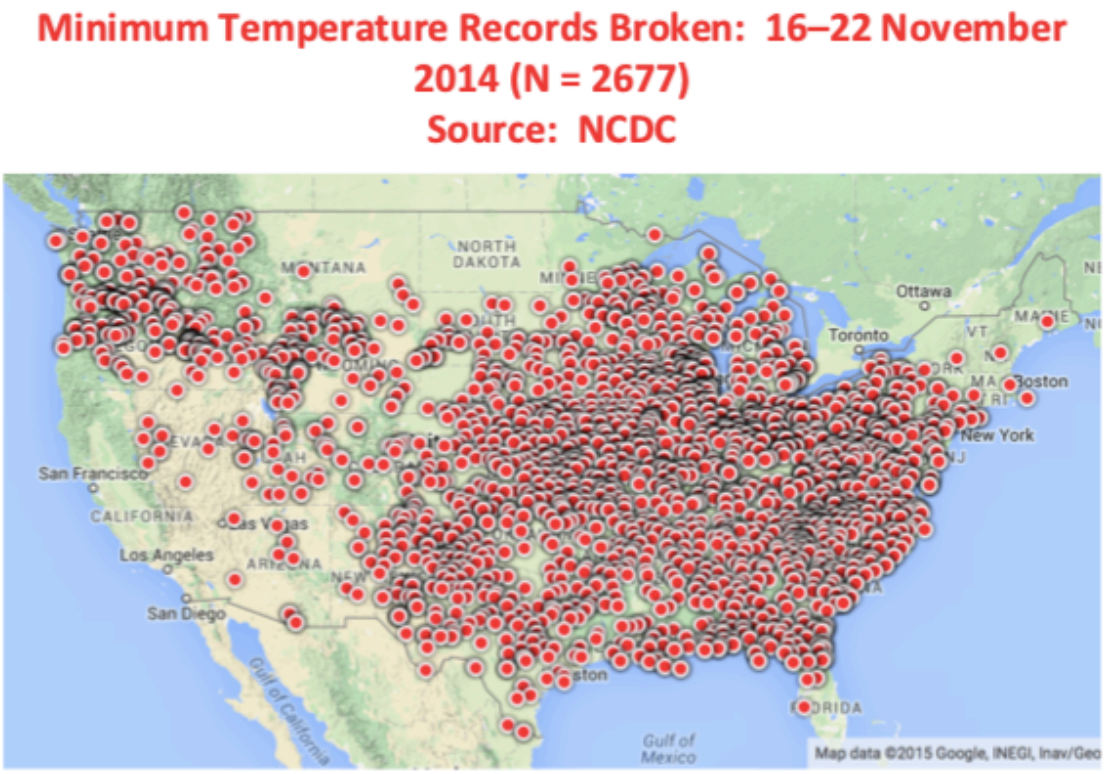


Figure 3. CONUS minimum temperature records broken between 16–22 November 2014. Source: National Climate Data Center

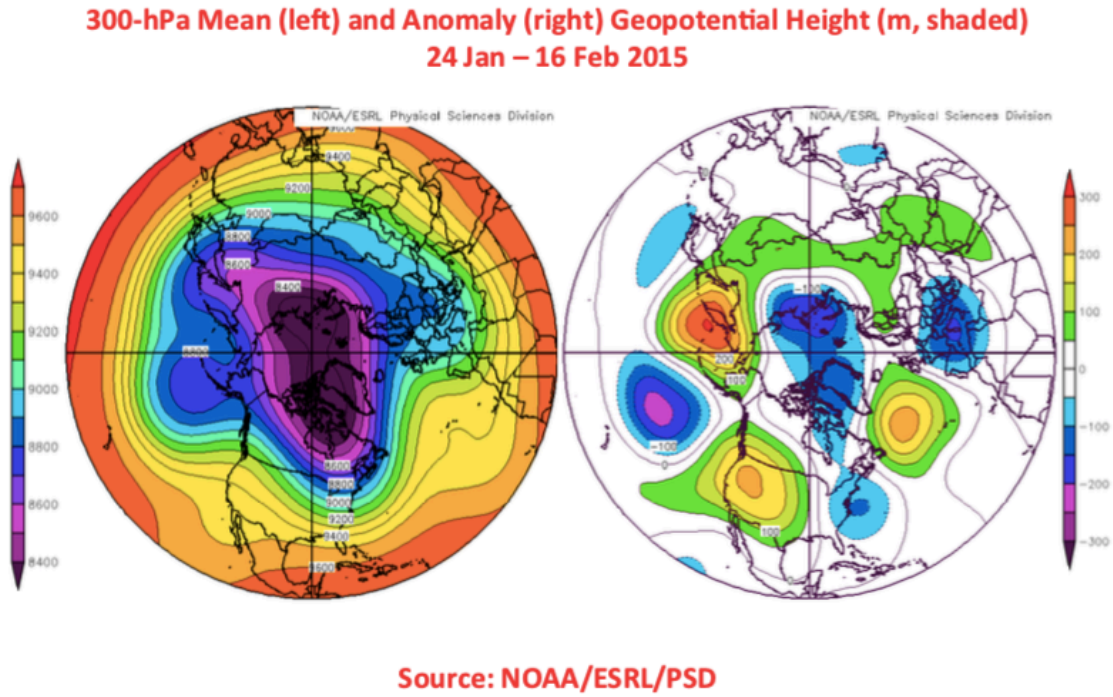


Figure 4. Northern Hemisphere 300 hPa mean geopotential height (contours in m, left) and geopotential height anomaly (shaded in m, right) for 24 January to 16 February 2015)

www.manaraa.com

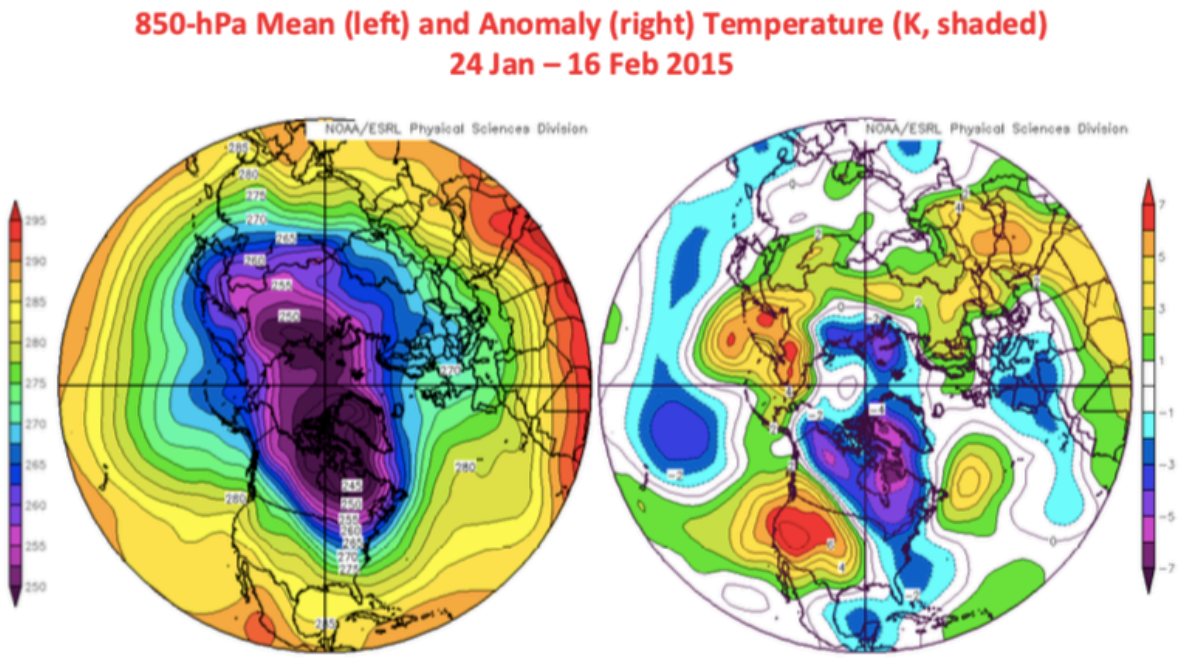


Figure 5. As seen in Figure 4 except for 850 hPa mean temperatures (contours in K, left) and temperature anomalies (shaded in K, right) for 24 January to 16 February 2015)

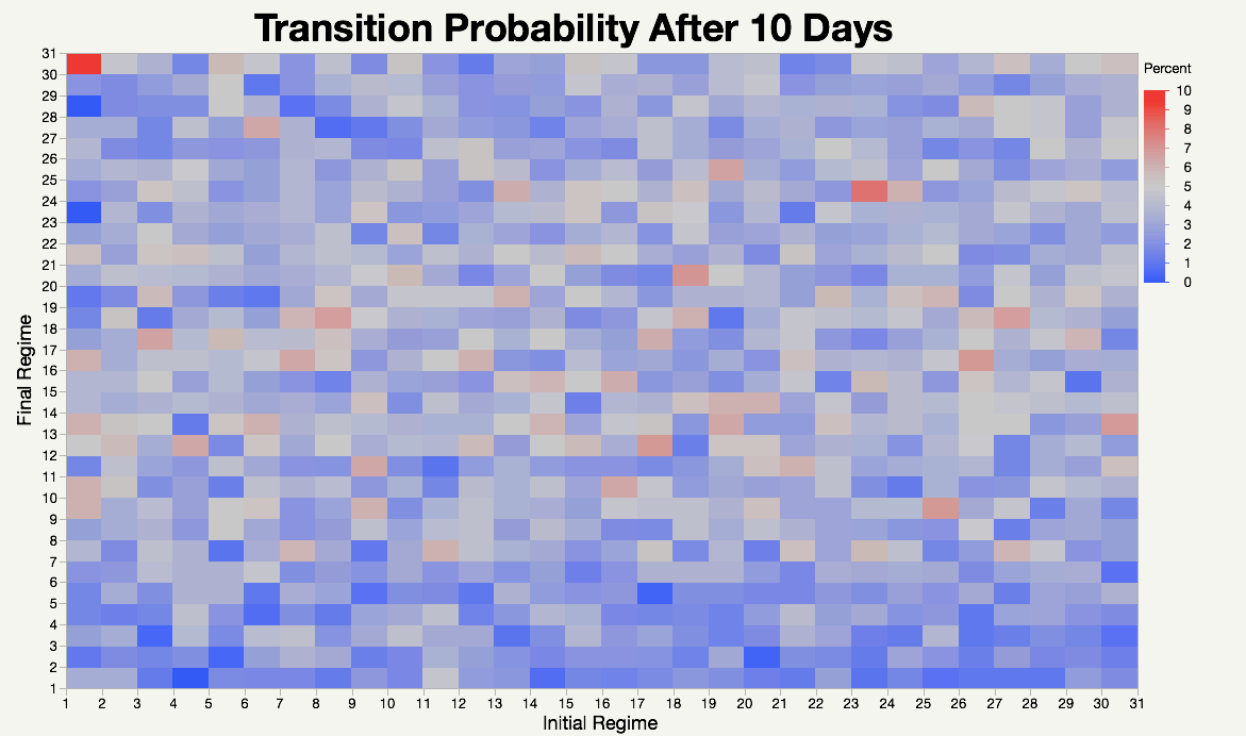


Figure 6. Transition Probabilities after 10 Days.

www.manaraa.com

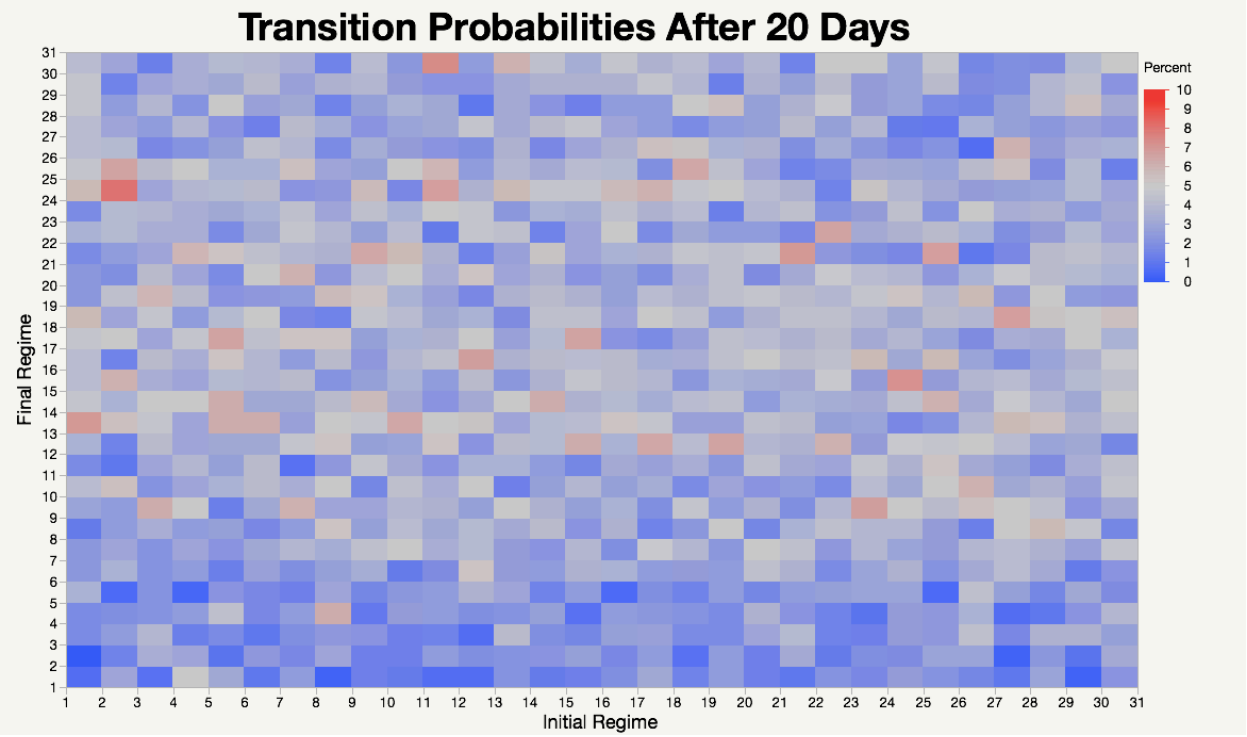


Figure 7. Transition Probabilities after 20 Days.

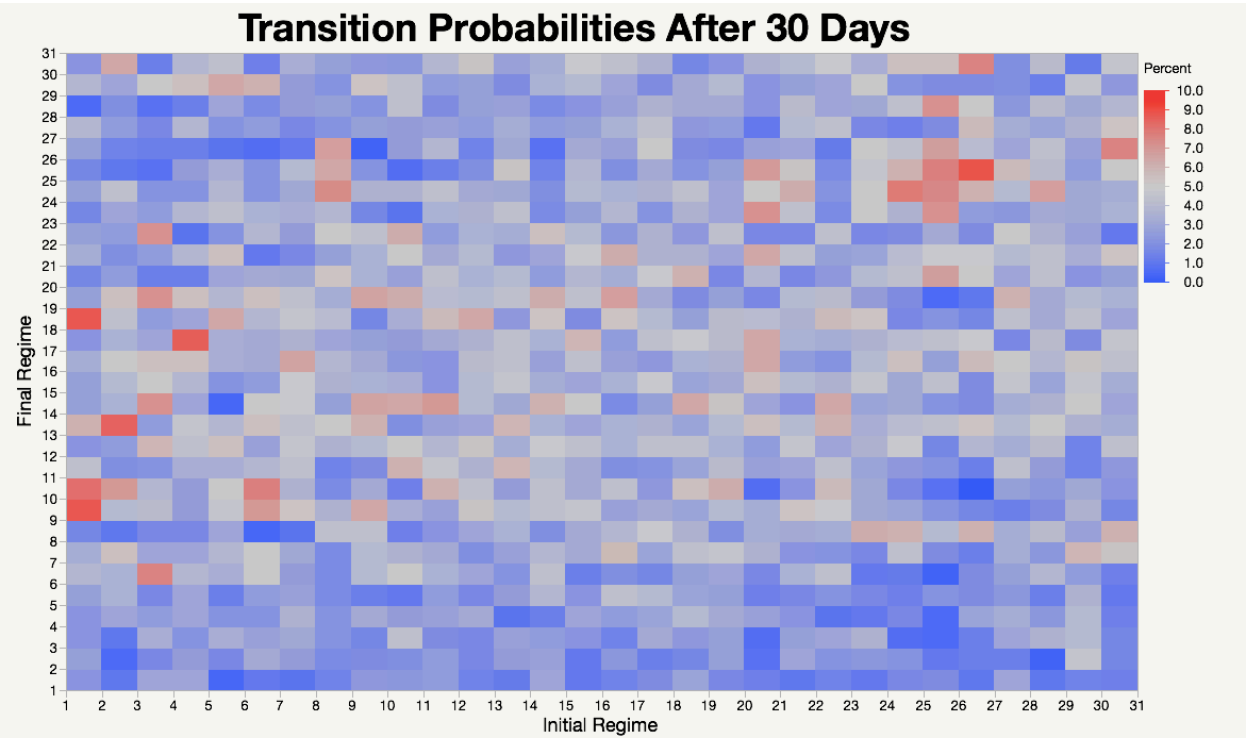


Figure 8. Transition Probabilities after 30 Days.

www.manaraa.com

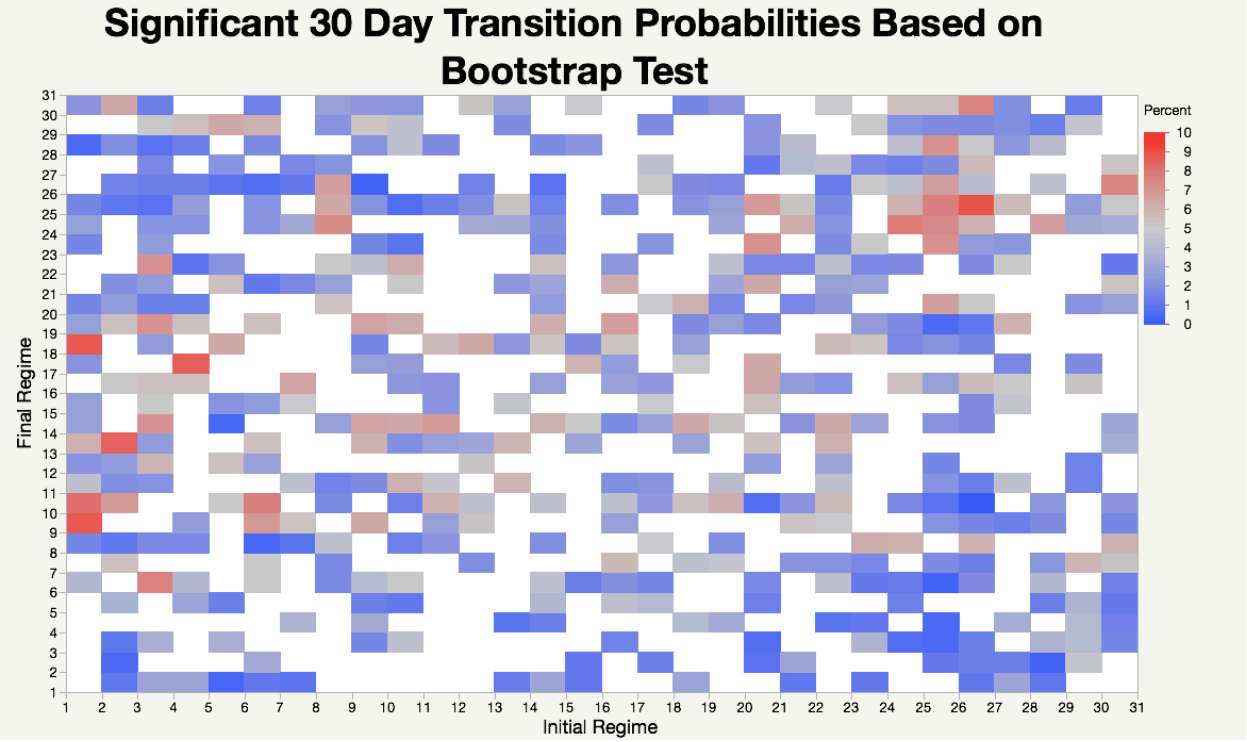


Figure 9. Statistically significant transition probabilities in shaded boxes based on Bootstrap test.

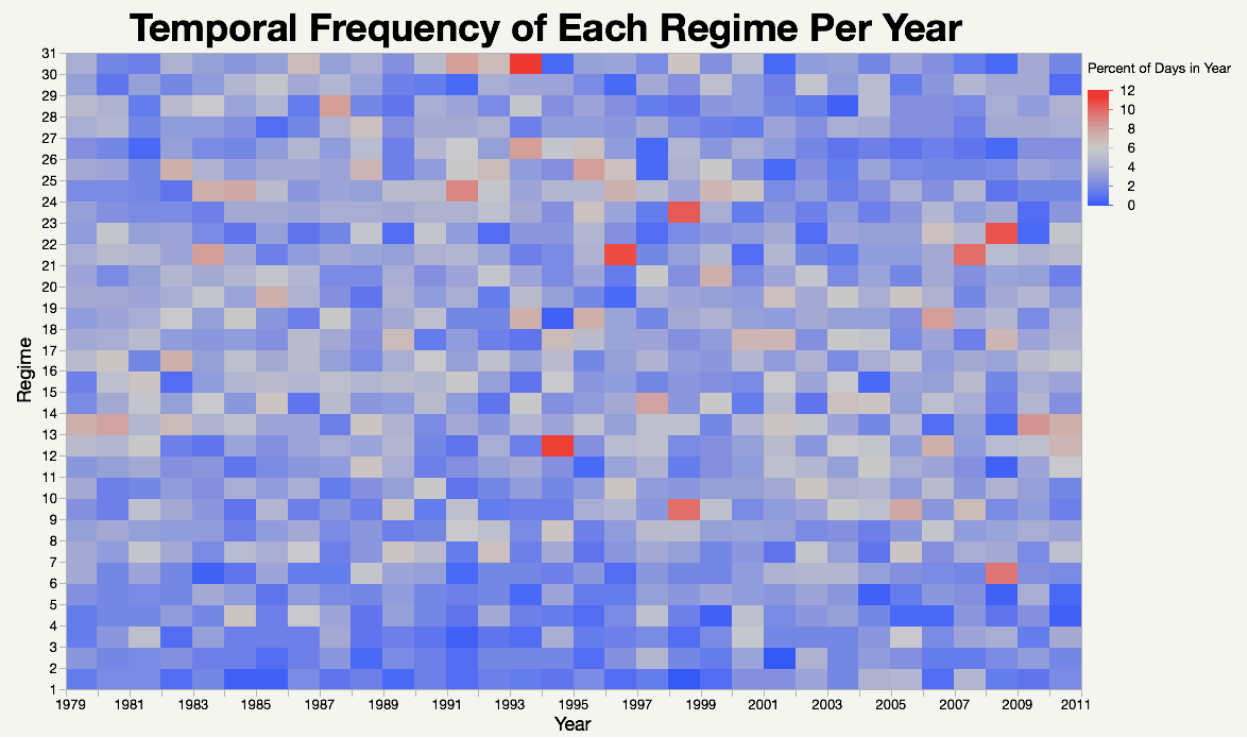


Figure 10. Temporal Frequency in percent of each regime per year.

www.manaraa.com

BiPlot

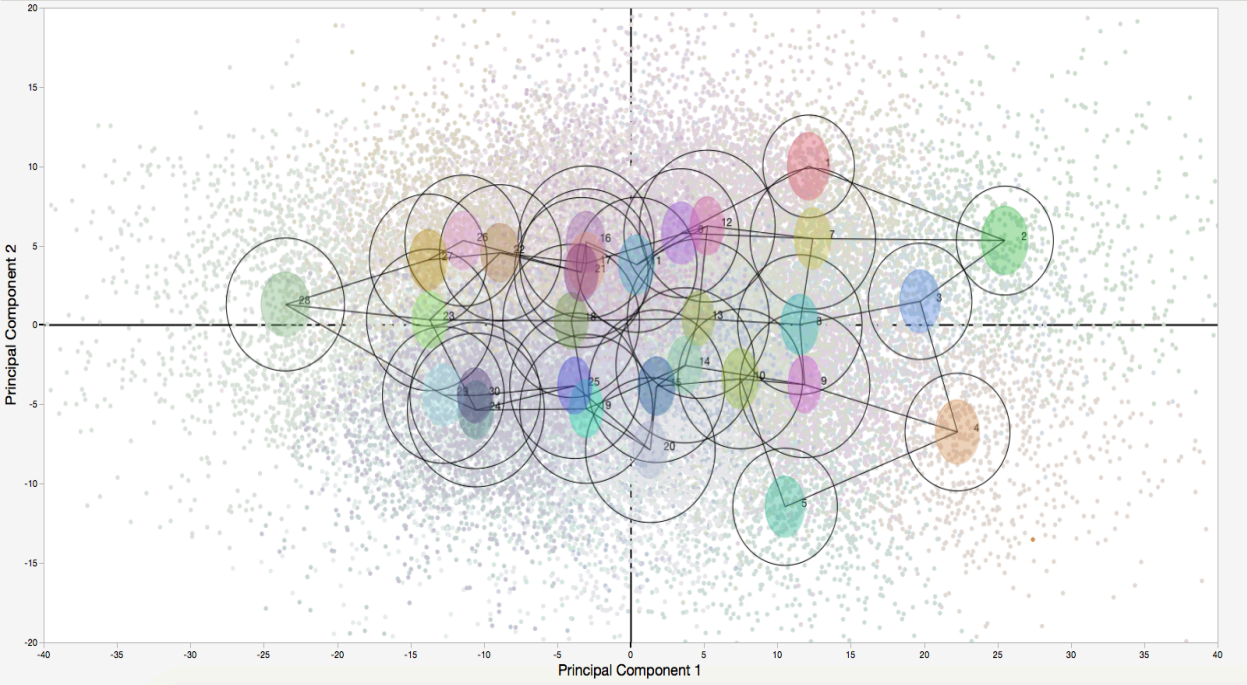
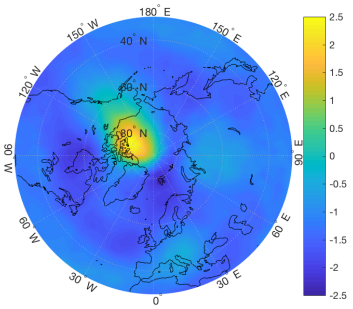
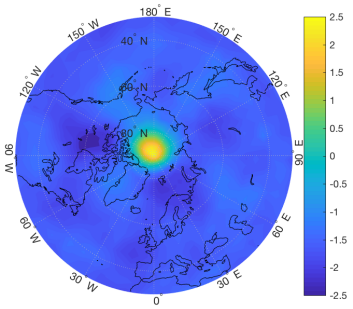


Figure 11. BiPlot showing spatial relationships between 30 total regimes.





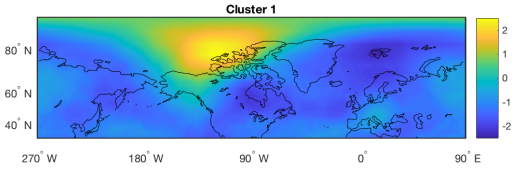


Figure 12. Regime 1 as shown by Standardized PV anomalies

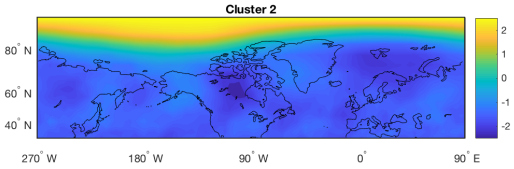
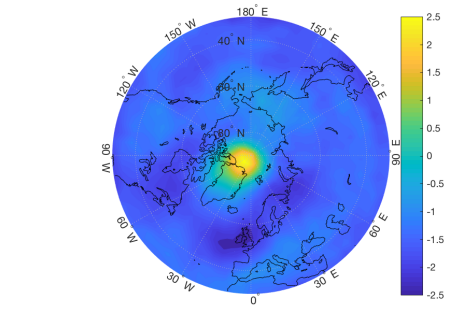
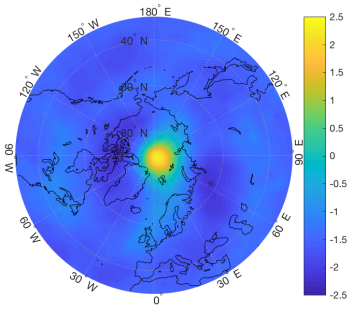


Figure 13. Regime 2 as shown by Standardized PV anomalies

www.manaraa.com





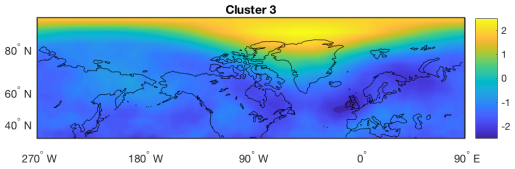


Figure 14. Regime 3 as shown by Standardized PV anomalies

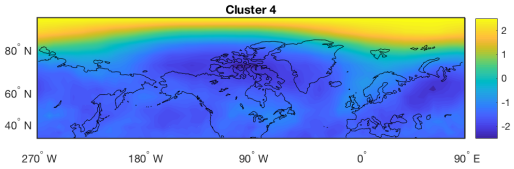
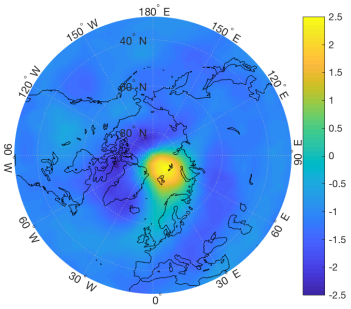
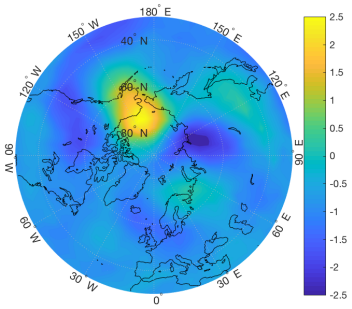


Figure 15. Regime 4 as shown by Standardized PV anomalies





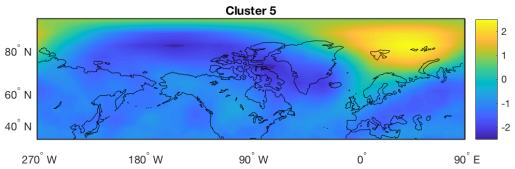


Figure 16. Regime 5 as shown by Standardized PV anomalies

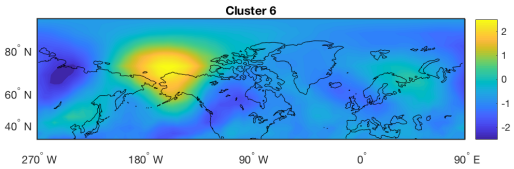
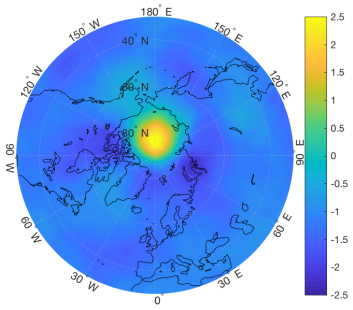


Figure 17. Regime 6 as shown by Standardized PV anomalies

www.manaraa.com



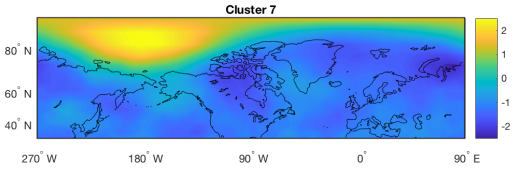
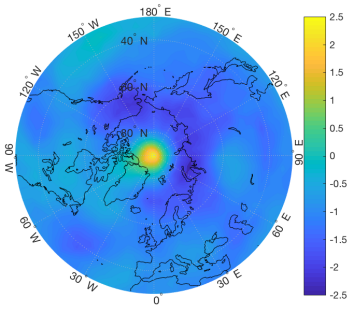


Figure 18. Regime 7 as shown by Standardized PV anomalies



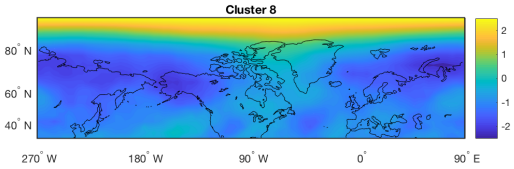
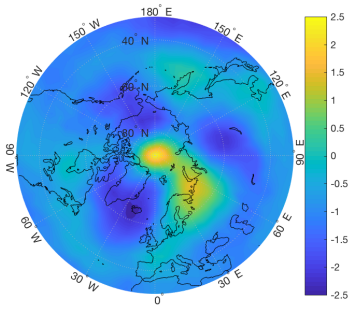


Figure 19. Regime 8 as shown by Standardized PV anomalies



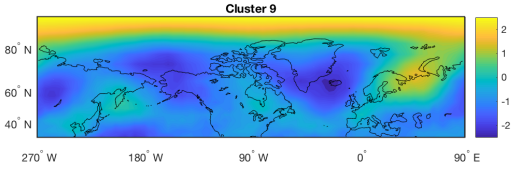
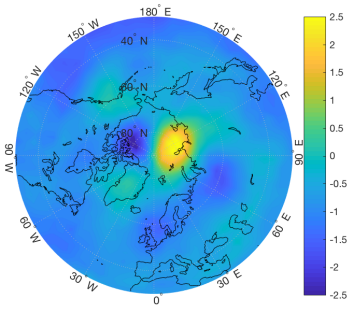


Figure 20. Regime 9 as shown by Standardized PV anomalies



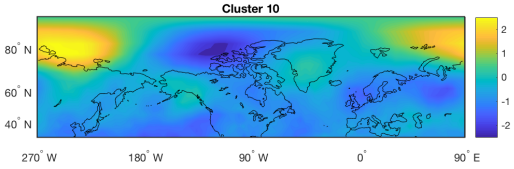
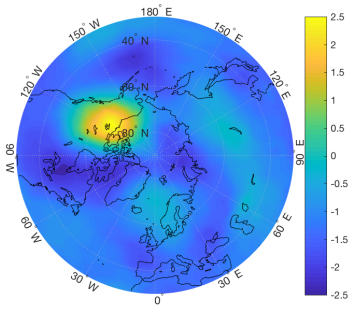


Figure 21. Regime 10 as shown by Standardized PV anomalies

www.manaraa.com



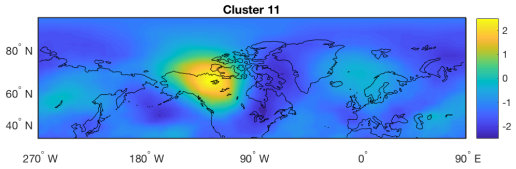
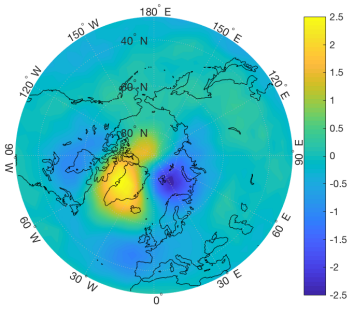


Figure 22. Regime 11 as shown by Standardized PV anomalies



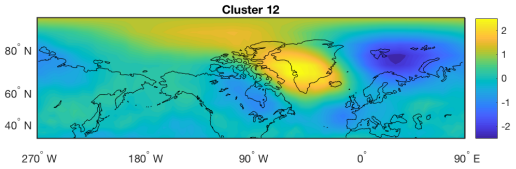
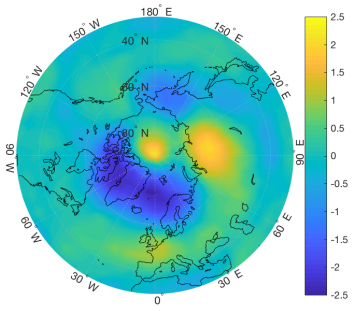


Figure 23. Regime 12 as shown by Standardized PV anomalies



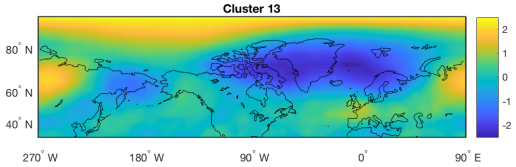
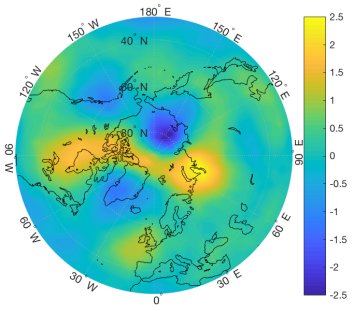


Figure 24. Regime 13 as shown by Standardized PV anomalies



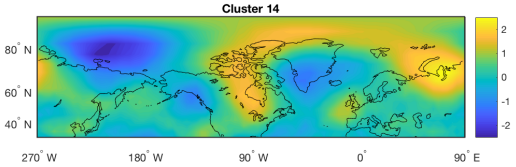
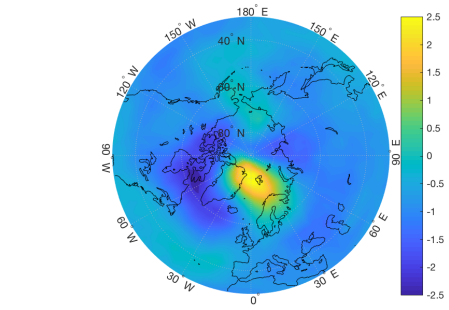


Figure 25. Regime 14 as shown by Standardized PV anomalies

www.manaraa.com



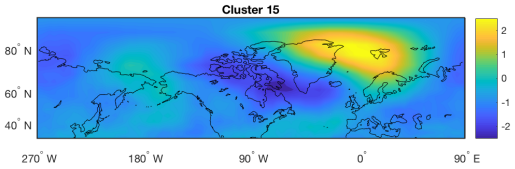
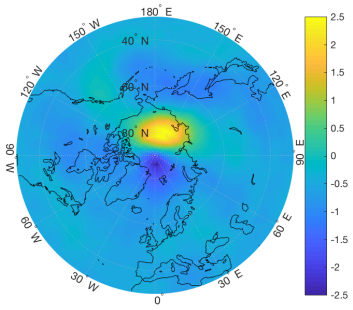


Figure 26. Regime 15 as shown by Standardized PV anomalies



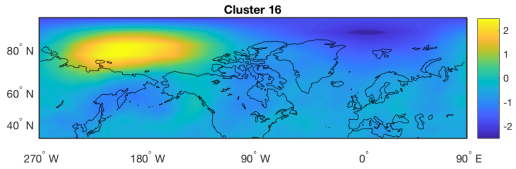
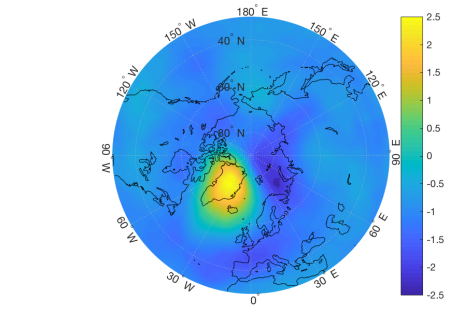


Figure 27. Regime 16 as shown by Standardized PV anomalies



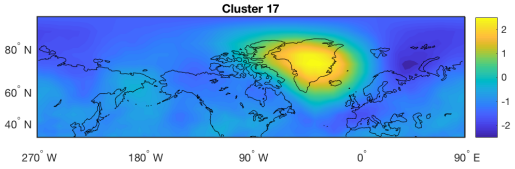
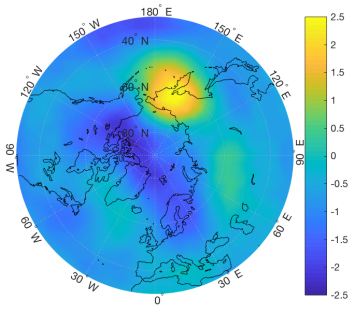


Figure 28. Regime 17 as shown by Standardized PV anomalies



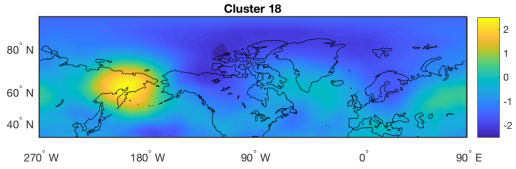


Figure 29. Regime 18 as shown by Standardized PV anomalies

www.manaraa.com

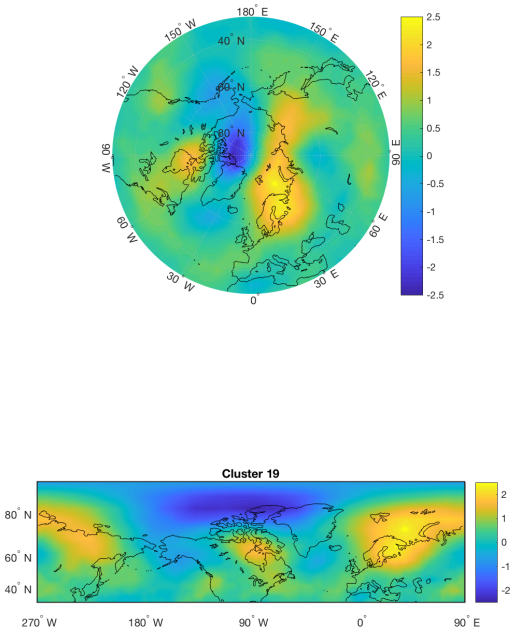


Figure 30. Regime 19 as shown by Standardized PV anomalies

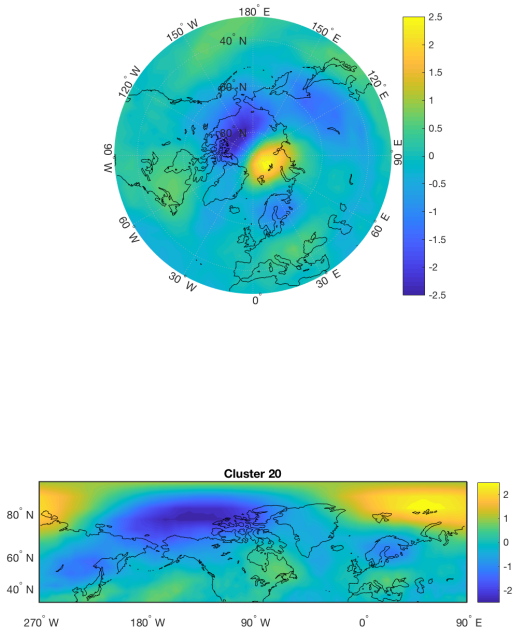


Figure 31. Regime 20 as shown by Standardized PV anomalies

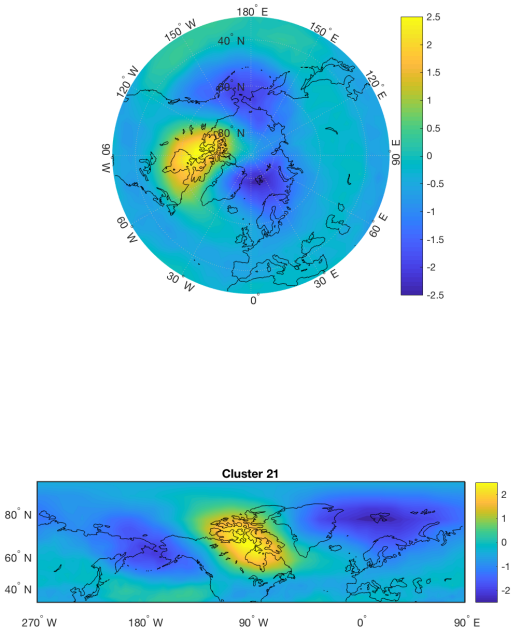


Figure 32. Regime 21 as shown by Standardized PV anomalies

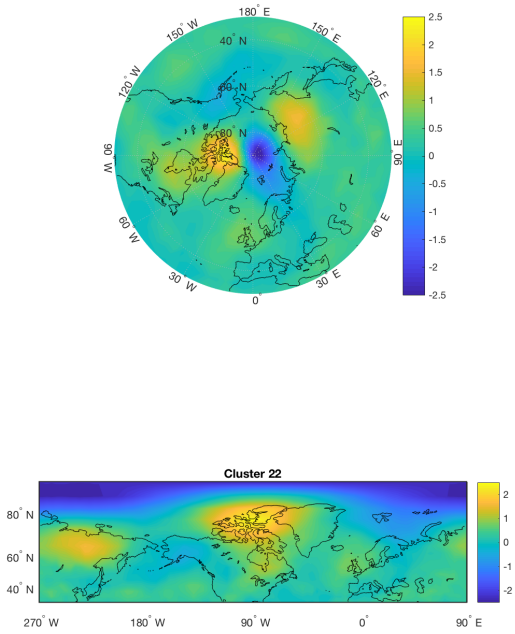
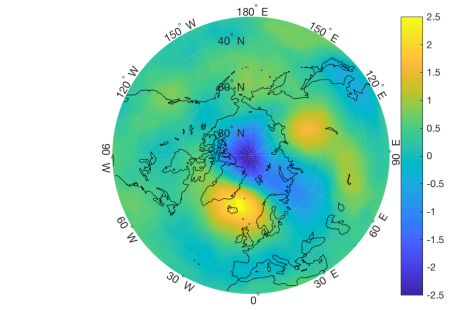


Figure 33. Regime 22 as shown by Standardized PV anomalies

www.manaraa.com



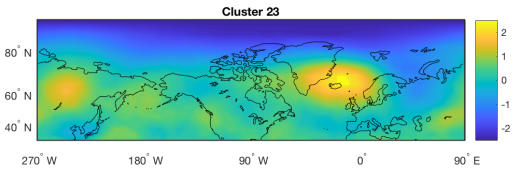
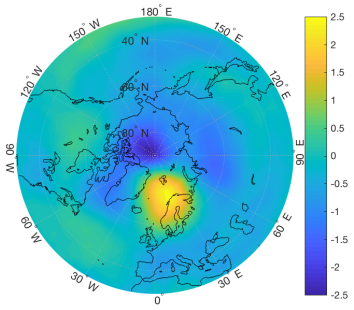


Figure 34. Regime 23 as shown by Standardized PV anomalies



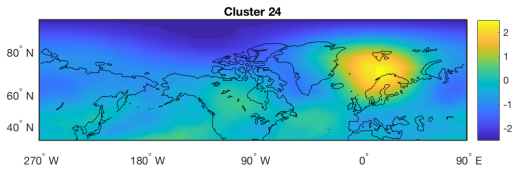
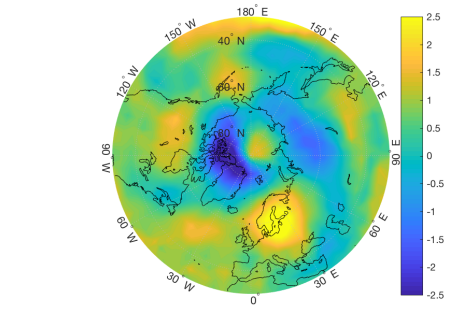


Figure 35. Regime 24 as shown by Standardized PV anomalies



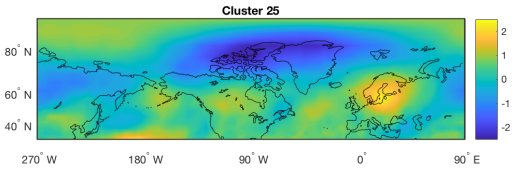
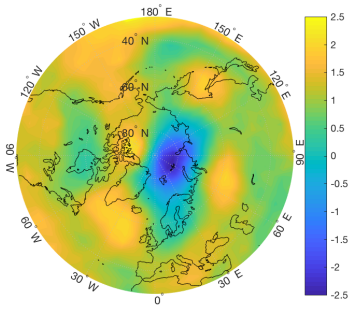


Figure 36. Regime 25 as shown by Standardized PV anomalies



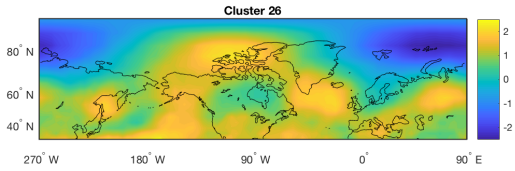
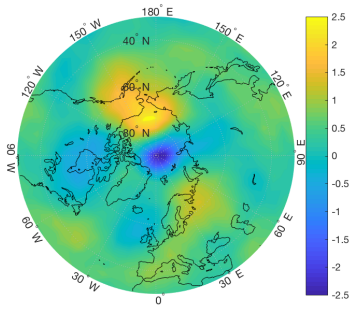


Figure 37. Regime 26 as shown by Standardized PV anomalies

www.manaraa.com



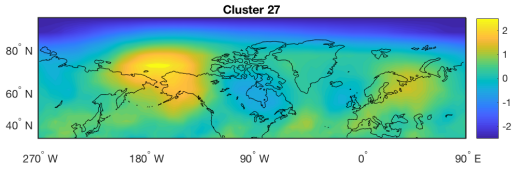
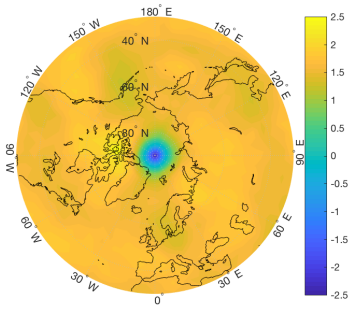


Figure 38. Regime 27 as shown by Standardized PV anomalies



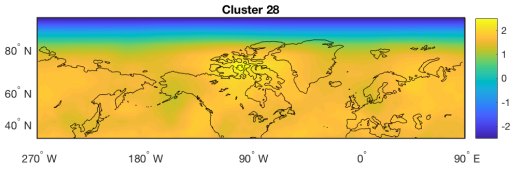
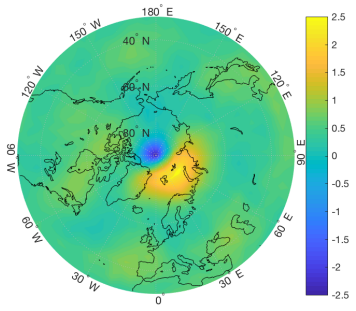


Figure 39. Regime 28 as shown by Standardized PV anomalies



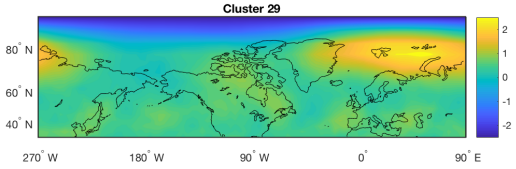
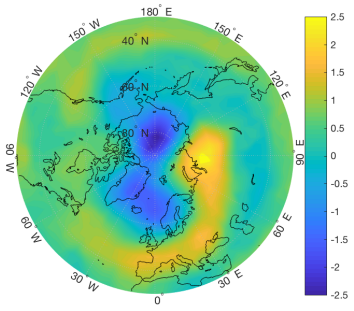


Figure 40. Regime 29 as shown by Standardized PV anomalies



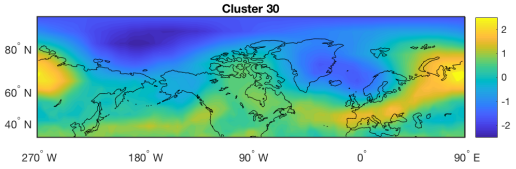


Figure 41. Regime 30 as shown by Standardized PV anomalies

www.manaraa.com

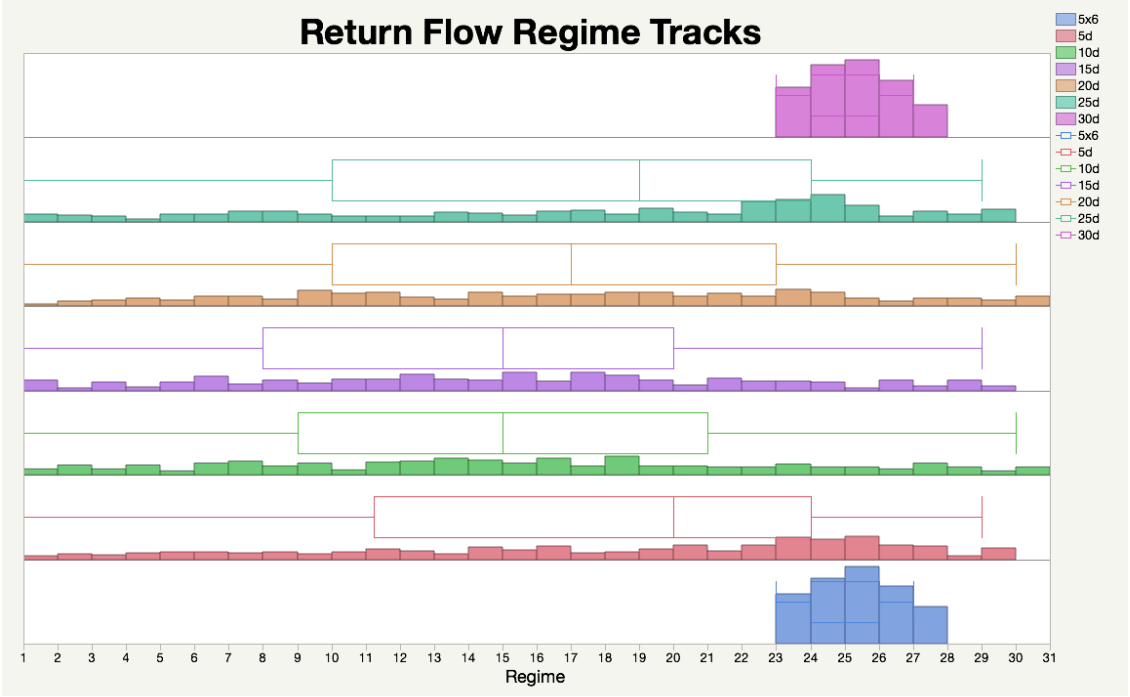


Figure 42. Return flow regime tracks. Initial regime is denoted in blue representing the starting point at the bottom row of the graph. Pink is 5 day lead time, lime green is 10 day lead time, purple is 15 day lead time, orange is 20 day lead time, dark green is 25 day lead time, and the top purple is the ending point at 30 day lead time.

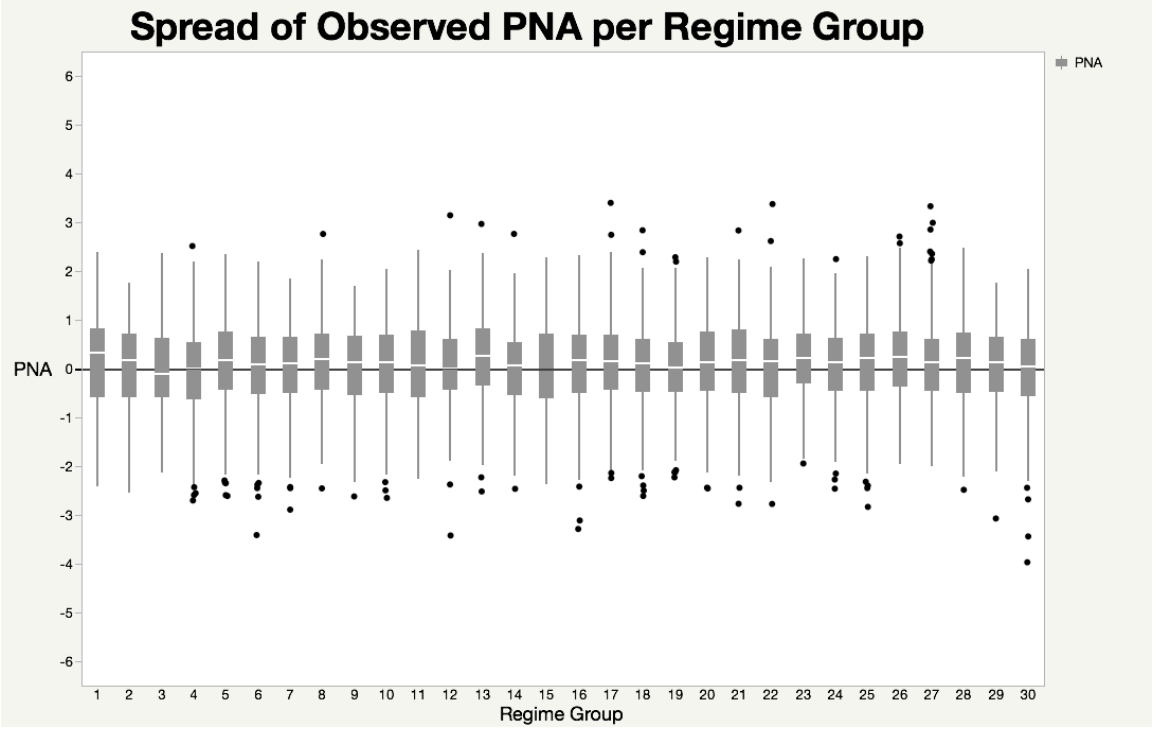


Figure 43. Spread of observed PNA index on each regime group.

www.manaraa.com

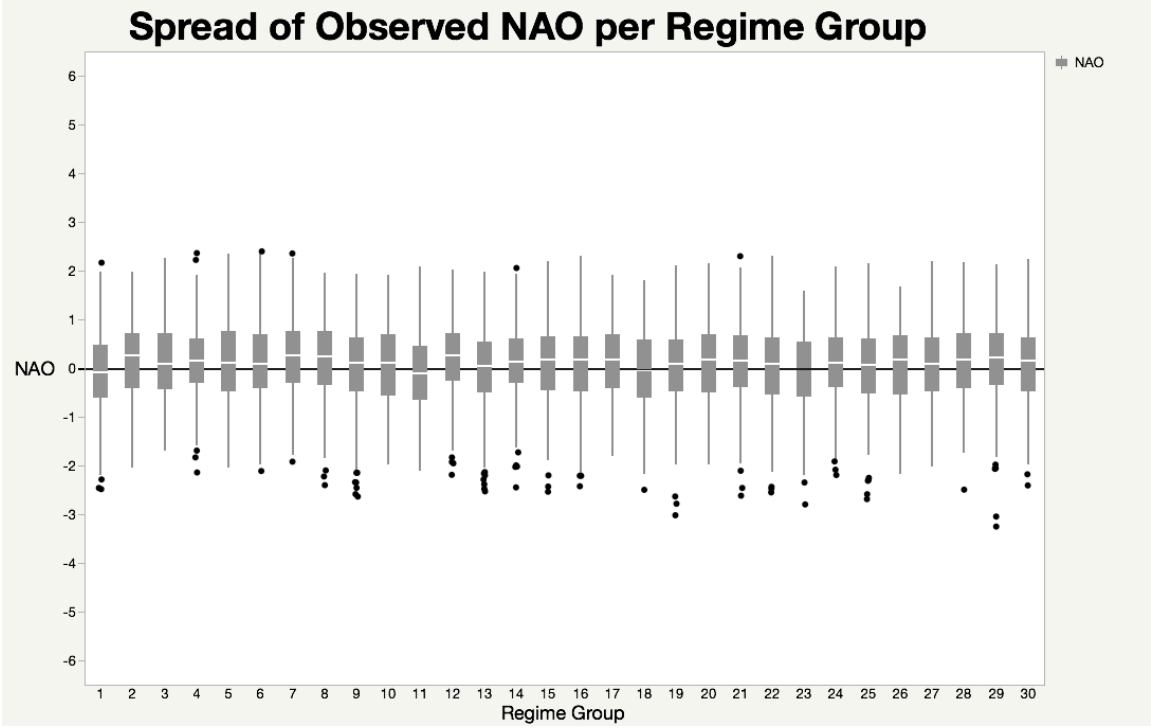


Figure 44. Spread of observed NAO index on each regime group.

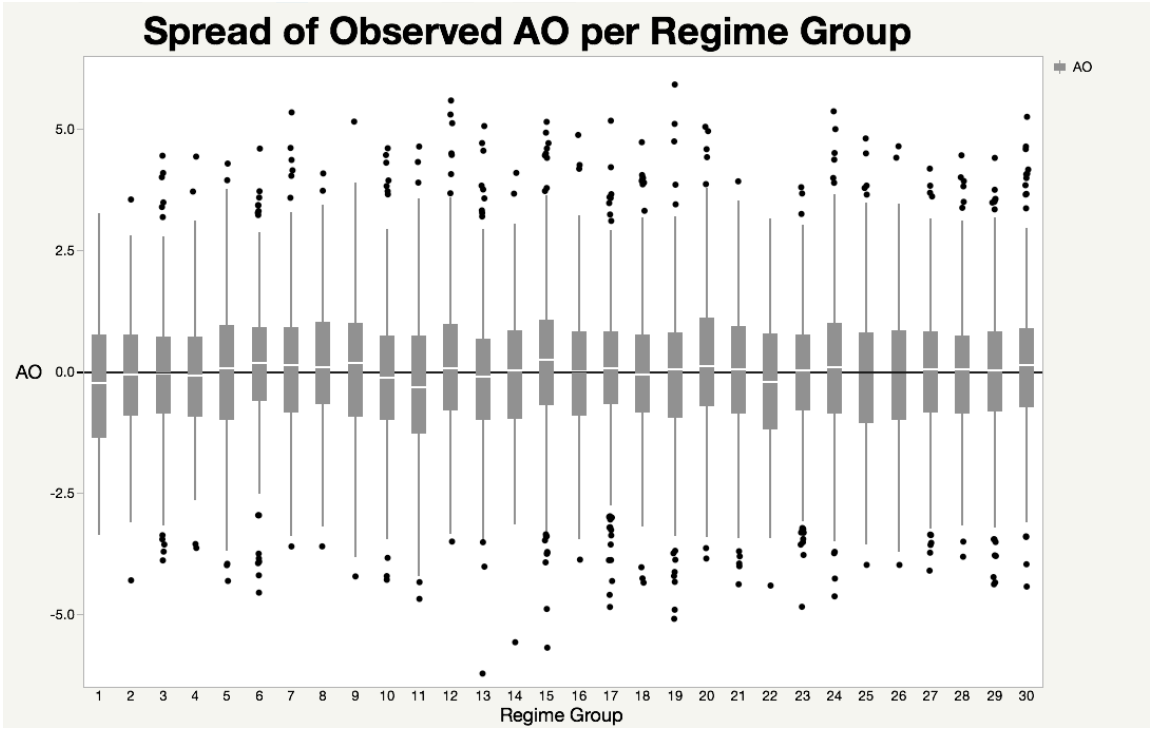


Figure 45. Spread of observed AO index on each regime group.

www.manaraa.com

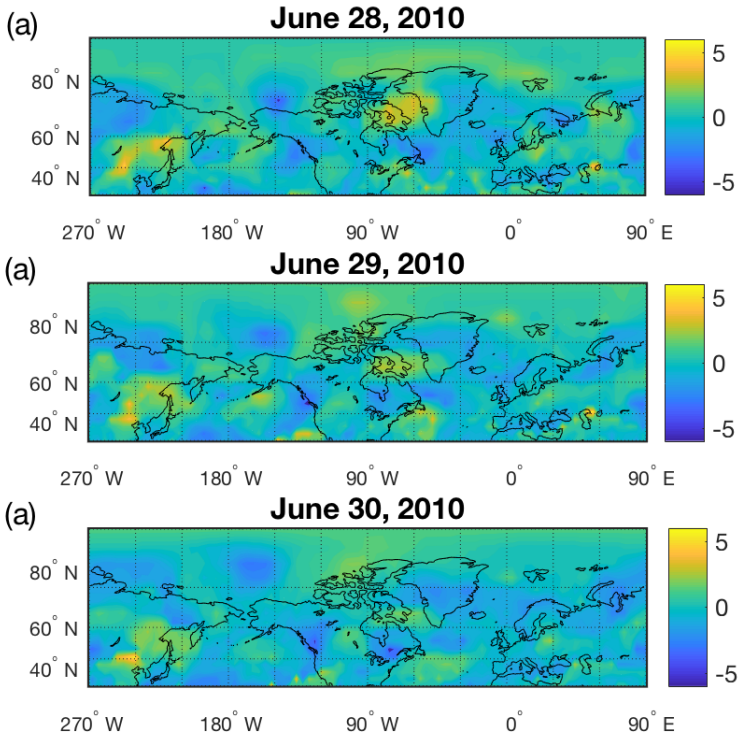


Figure 46. Standardized PV anomalies obtained from CFS Reanalysis data from 28-30 June, 2010.

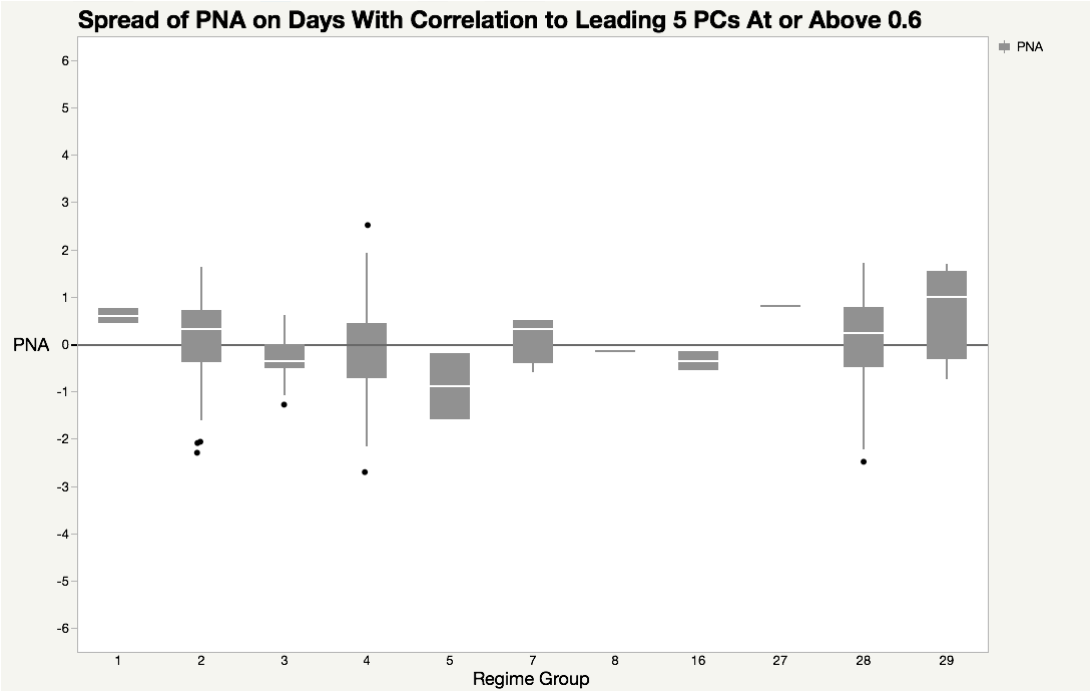


Figure 47. Spread of observed PNA on Days with individual correlation to Leading 5 PCs at or above 0.6.

www.manaraa.com

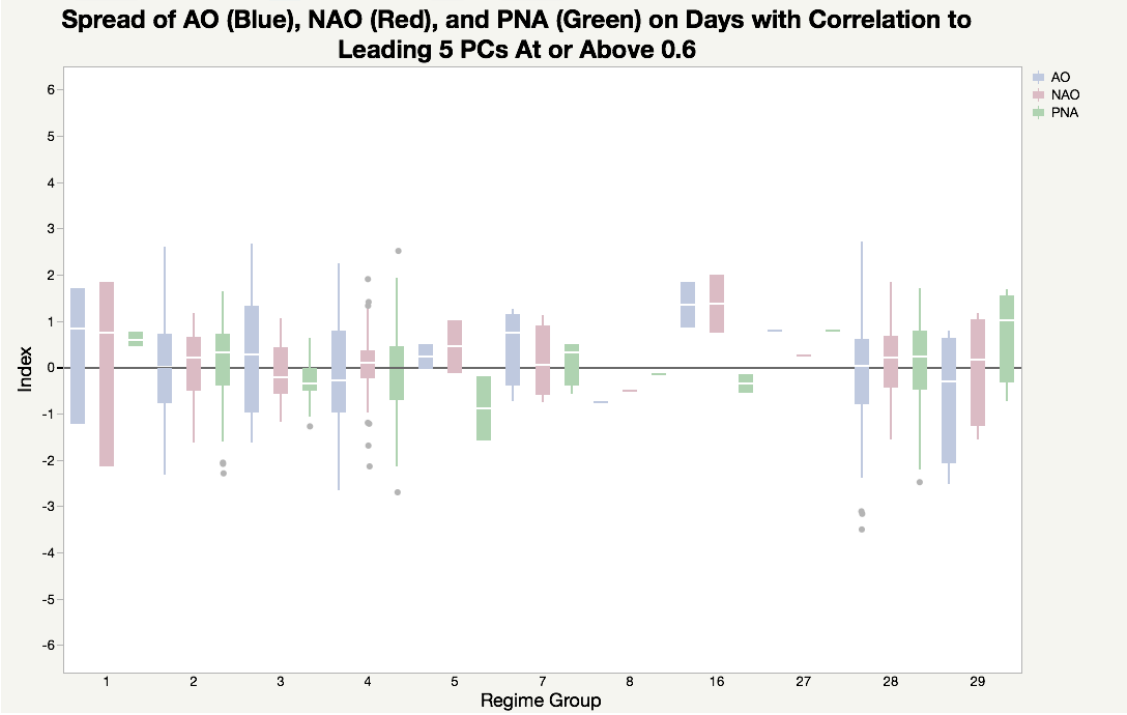


Figure 48. Spread of observed PNA (Green), NAO (Red), and AO (Blue) on in individual days with correlation to Leading 5 PCs at or above 0.6.

www.manaraa.com

REFERENCES

Agel, L., Barlow, M., Feldstein, S., Gutowski, W., 2018: Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dynamics*, **50(6)**, 1819-1839.

Anderson, J. R., and J. R. Gyakum, 1989: A diagnostic study of Pacific Basin circulation regimes as determined from extratropical cyclone tracks. *Mon. Wea. Rev.*, **117**, 2672-2686.

Bosart, L. F., 1981: The Presidents' Day Snowstorm of 18–19 February 1979: A Subsynoptic-Scale Event. *Mon. Wea. Rev.*, **109**, 1542–1566.

Bosart, L. F., and S. C. Lin, 1984: A Diagnostic Analysis of the Presidents' Day Storm Of February 1979. *Mon. Wea. Rev.*, **112**, 2148–2177.

Cassano, J. J., E. N. Cassano, M. W. Seefeldt,W. J. Gutowski Jr., and J. M. Glisan (2016), Synoptic conditions during wintertime temperature extremes in Alaska, *J. Geophys. Res. Atmos.*, **121**, 3241–3262, doi:10.1002/2015JD024404.

Cordeira, J. M., and L. F. Bosart, 2010: The Antecedent Large-Scale Conditions of the "Perfect Storms" of Late October and Early November 1991. *Mon. Wea. Rev.*, **138**, 2546–2569.

Cordeira, J. M., and L. F. Bosart, 2011: Cyclone Interactions and Evolutions during the "Perfect Storms" of Late October and Early November 1991. *Mon. Wea. Rev.*, **139**, 1683–1707.

Gervais, M., Attallah, E., Gyakum, J.R. Tremblay, B., 2016, Arctic Air Masses in a Warming World, *J. Climate*, **29**, 2359-2373.

Grumm R. H., Hart, R., 2001: Standardized Anomalies Applied to Significant Cold Season Weather Events: Preliminary Findings, *Wea. Forecasting*, **16**, 736-754, https://doi.org/10.1175/1520-0434(2001)016<0736:SAATSC>2.0.CO;2,

Gyakum, J. R., 1983a: On the evolution of the *QE II* storm. I. Synoptic aspects. *Mon. Wea. Rev.*, **111**, 1137-1155.

Gyakum, J. R., 1983b: On the evolution of the *QE II* storm. II. Dynamic and thermodynamic structure. *Mon. Wea. Rev.*, **111**, 1156-1173

Hoskins, B. J., McIntyre, M. E., Robertson, A.W., 1985: On the use a significance of isentropic potential vorticity maps, *Quarterly Journal of the Royal Meteorological Society*, **111**, 470, https://doi.org/10.1002/qj.49711147002

Lennard, C., and G. Hegerl, 2015: Relating changes in synoptic circulation to the surface rainfall response using self-organising maps. *Climate Dyn.*, **44**, 861–879, doi:10.1007/ s00382-014-2169-6

www.manaraa.com

Plaut, G., and R. Vautard, 1994: Spells of Low-Frequency Oscillations and Weather Regimes in the Northern Hemisphere, *Journal of Atmospheric Sciences*, **51**, 210-236. 10.1175/1520-0469(1994)051<0210:SOLFOA>2.0.CO;2.

Roebber, P.J., 2009: Planetary waves, cyclogenesis, and the irregular breakdown of zonal motion over the North Atlantic., *Mon. Wea. Rev.*, **13**7, 3907-3917.

Roebber, P. J., 2016: An Adaptive Bayesian Model Combination (BMC) Post Processor for the HRRR-TLE Forecast System, University of Wisconsin at Milwaukee, 8 pp.

Roebber, P. J., J.R. Gyakum, and L.F. Bosart, 2016: Improving Prediction of Large-Scale Regime Transitions, University of Wisconsin-Milwaukee, McGill University, University at Albany/SUNY, 14 pp.

Turner, J. K., J. R. Gyakum, and S. M. Milrad, 2013: A thermodynamic analysis of an intense North American arctic air mass. *Mon. Wea. Rev.*, **141**, 166-181.

Vigaud, N., A. Robertson, and M. Tippett, 2018: Predictability of recurrent weather regimes over North America during winter from submonthly reforecasts. *Mon. Wea. Rev.* doi:10.1175/MWR-D-18-0058.1, in press.

Winters, A. C., Keyser, D., Bosart, L. F., 2017: Antecedent Synoptic Environments Most Conducive to North American Polar/Subtropical Jet Superpositions. *2017 AGU Fall Meeting*, New Orleans, LA, American Geophysical Union, http://www.atmos.albany.edu/facstaff/awinters/Presentations/AGU17poster_Winters.pdf

Young, Forest W., 1999: Principal Components: BiPlots. Accessed 1 July 2018, http://forrest.psych.unc.edu/research/vista-frames/help/lecturenotes/lecture13/biplot.html.

www.manaraa.com