<u>ATOC5860 - Homework 3 - due March 3, 2022</u>

Please send your homework to Jen/Prof. Kay on Slack as a direct message.

Please Name Your Homework Files: "ATOC5860_HW3_LastName.pdf, .html, .ipynb"

Your submissions should include: 1) A .pdf document with responses to the questions below, 2) Your code in both .ipynb and .html format.

Show all work including the equations used (e.g., by referring to the Barnes Notes).
Write in complete, clear, and concise sentences.
Eliminate spelling/grammar mistakes.
Label all graph axes. Include units.
Report values using appropriate rounding.

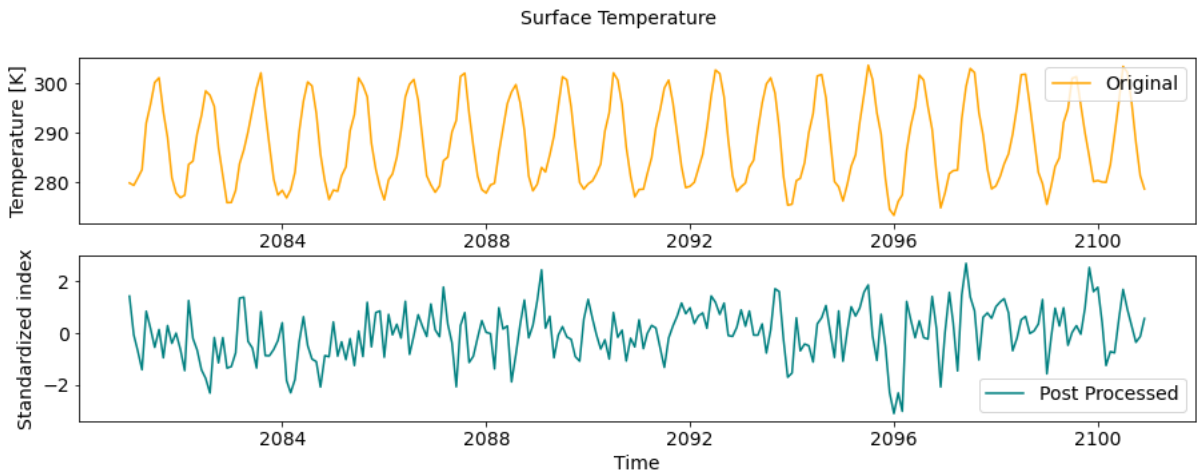
Problem I) Apply EOF-PCA analysis to the CESM Large Ensemble Monthly Surface Temperature Fields over North America (25-55 °N, 220-300 °E). (50 points total)

Your data: The monthly surface temperature field from the CESM Large Ensemble member 1. You will analyze 20 years of monthly temperatures starting in January 2081 and going through December 2100. The file is available on Canvas and is a netcdf file: b.e11.BRCP85C5CNBDRD.f09_g16.001.cam.h0.TS.208101-210012.nc

1) Prepare your data for analysis. (10 points).

> **For this dataset, this includes:**
> **a) Subset the global data to a North American Domain (25-55 °N, 220-300 °E)**
> **b) Remove the seasonal cycle**
> **c) Cosine weight the data (Account for the decrease in grid-box area as one approaches the pole (i.e. weight your data by the cosine of latitude when computing the EOFs)).**
> **d) Standardizing the data (subtract the mean, divide by the standard deviation). Hint: standardize last so that your data have a mean of 0 and a standard deviation of 1.**

First we subset the data to the north American domain with bounds WESN=[220°,300°,25°,55°]. Next, we weight the data by the square root of the cosine of latitude following Hannachi et al. (2007). This weighting considers the conversion of longitudes near the poles. The anomalous turbulent portion of surface temperature is found by subtracting the population mean from each timestamp value. We then remove the seasonal cycle by taking a moving monthly average and subtract this average from the base data for each month, respectively. This also acts as the first step for standardizing the data. Finally, the dataset is standardized by normalizing to the population standard deviation. The plots below show the original surface temperature for the full period (top) and the post-processed surface temperature for the full period (bottom) taken at an arbitrary coordinate in space.



2) Calculate the EOFs and PCs using both methods discussed in class: eigenanalysis of the covariance matrix and Singular Value Decomposition (SVD). (10 points)

Eigenanalysis Method

We first calculate the covariance matrix from Barnes Chpt. 3 Eq. 7:

$\boldsymbol{C} = \frac{1}{M}\boldsymbol{X}^T\boldsymbol{X}$

Next we calculate the eigenvalues (Λ) and eigenvectors ($\boldsymbol{E}$) from Barnes Chpt. 3 Eq. 42:

$\boldsymbol{CE} = \boldsymbol{E}\Lambda$ where individual eigenvectors ($\boldsymbol{e_i}$) correspond to eigenvalues (λ_1)

The percent of variance explained (pve) by each eigenvalue is calculated by:

$pve = 100 * \frac{|\lambda_i|}{\sum|\lambda_j|}$

SVD Method

We decompose surface temperature ($\boldsymbol{X}$) into into the produce of three matrices using Barnes Chpt. 3 Eq. 65:

$\boldsymbol{X} = \boldsymbol{U\Sigma V^T}$ where the columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{XX^T}$ and the rows of $\boldsymbol{V^T}$are eigenvectors of $\boldsymbol{X^T X}$

Here, $\boldsymbol{\Sigma}$ contains the square root of eigenvalues λ_i so we calculate percent variance explained by:

$pve = 100 * \frac{\Sigma_i^2}{\sum \Sigma_j^2}$

3) Plot the first 10 eigenvalues (scaled as the percent variance explained) in order of variance explained. How much variance do the first three EOFs explain? Add error bars following North et al. 1982 (or another method of your choice). Describe how you determined the effective degrees of freedom N* and report the value. *Reminder: You should calculate N* using the data with the seasonal cycle removed.* How many statistically significant EOFs are there? (10 points)

First, we calculate the autocorrelation from Barnes Chpt. 2 Eq. 67:

$$\gamma(t) = \frac{1}{N-\tau}\sum x'(t)\cdot x'(t+\tau)$$

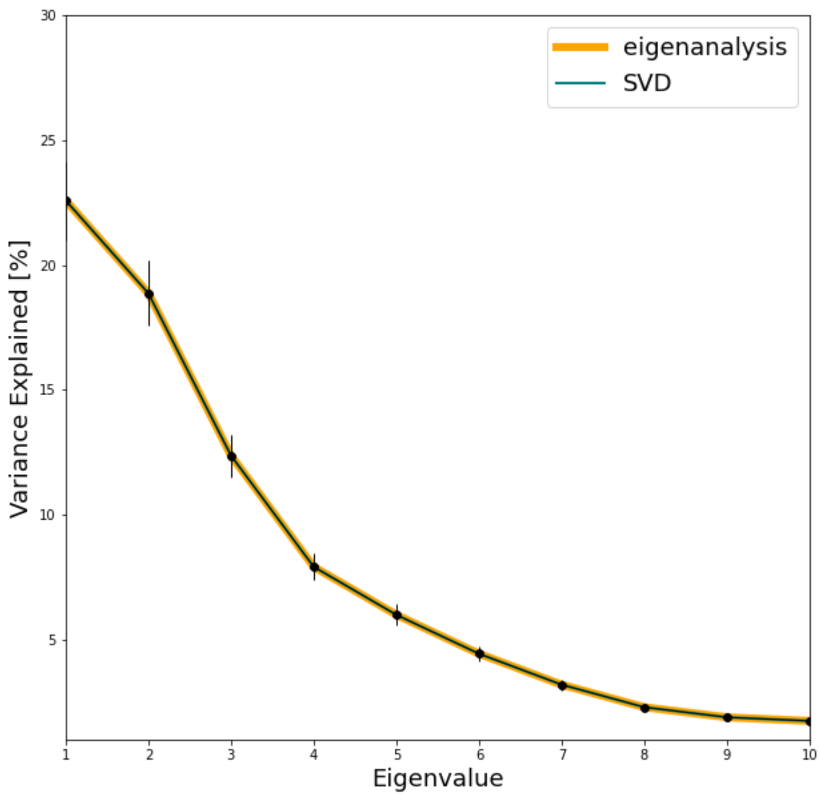
where N is the dataset size, t is the timestamp of interest, $x(t)$ is the post-processed surface temperature data at that timestep, τ is the lag of 1. This covariance is normalized to the population variance at a lag of 0. Next, the effective sample size is calculated from Barnes Chpt. 2 Eq. 90:

$$N^* = -\frac{N}{2}\ln(\gamma(t))$$

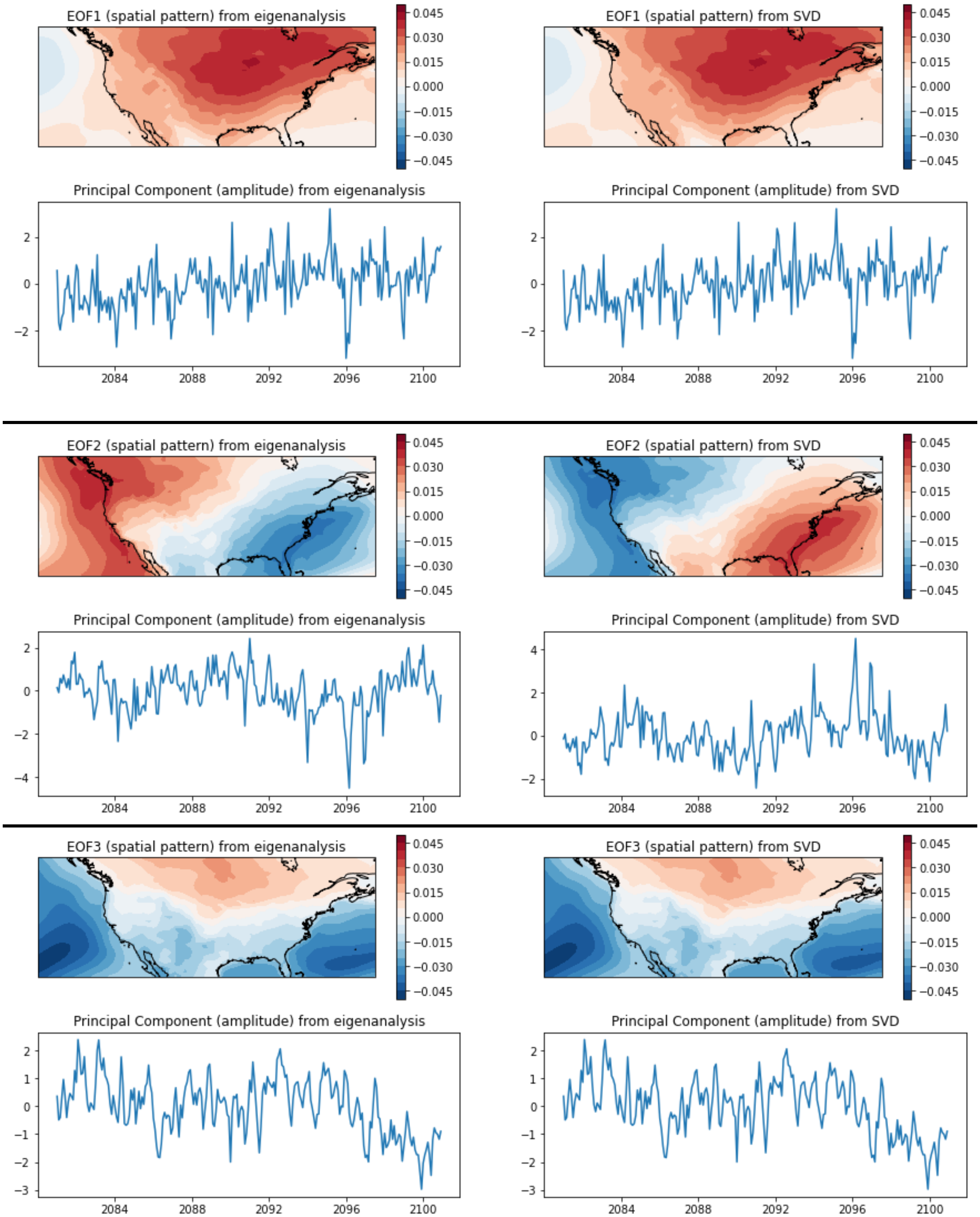
Autocorrelation and N^* values are calculated individually for each longitude/latitude pair across the full period of interest. The mean autocorrelation is 0.5. The mean N^* is 104. N^* is used to find the 95% confidence bounds on individual eigenvalues from Barnes Chpt. 3 Eq. 80:

$$\Delta\lambda = \lambda_i\sqrt{\frac{2}{N^*}}$$

From this, we keep the first three significant eigenvalues which explain 22.6%, 18.9%, and 12.4%, respectively.

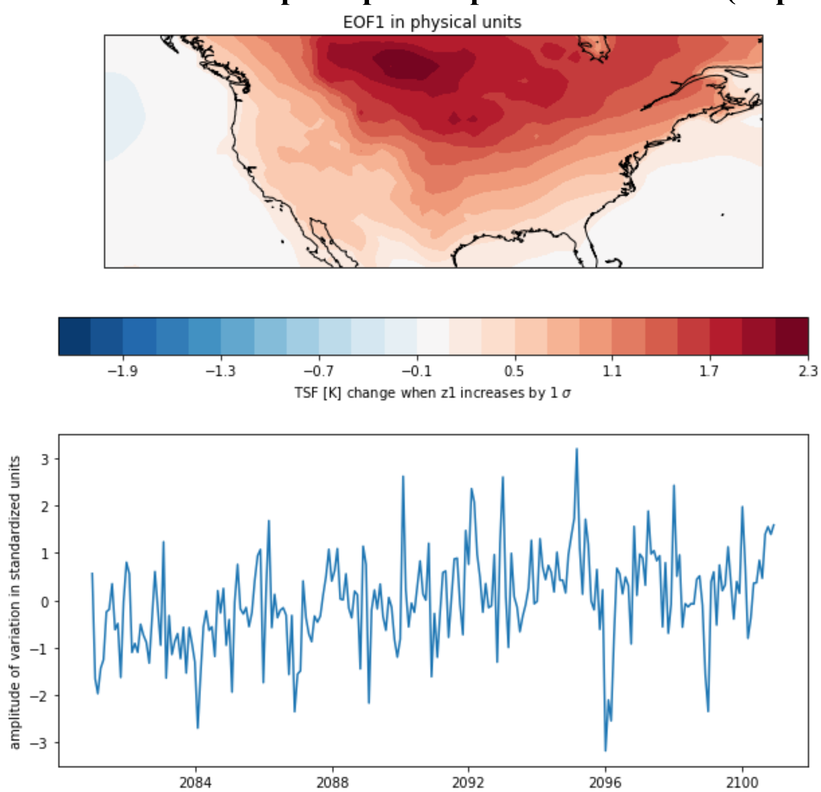


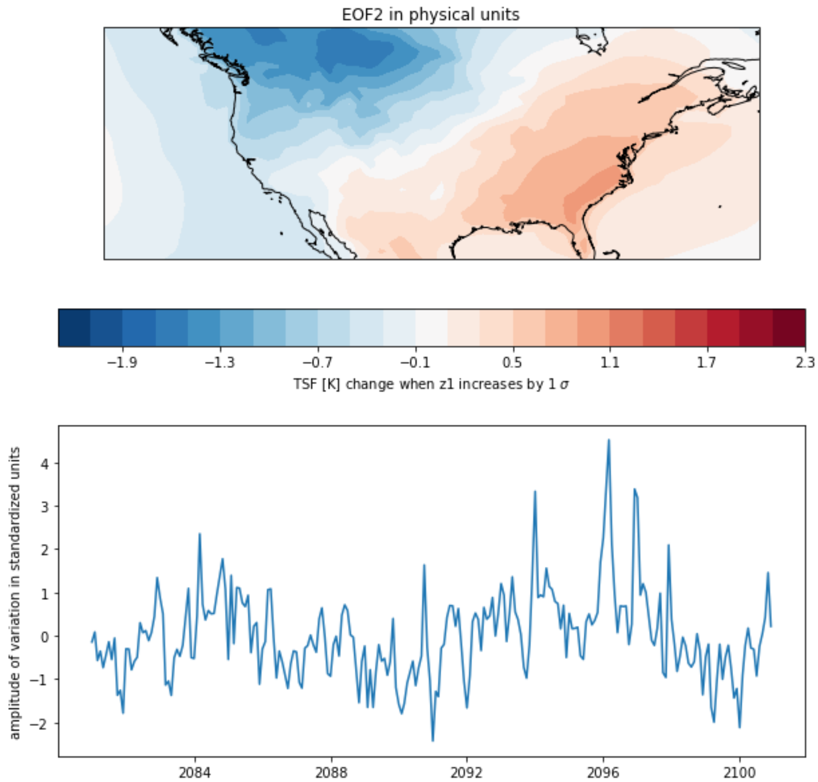
4) Plot the first three EOF patterns and PC time series (unless you want to look at more). (10 points)

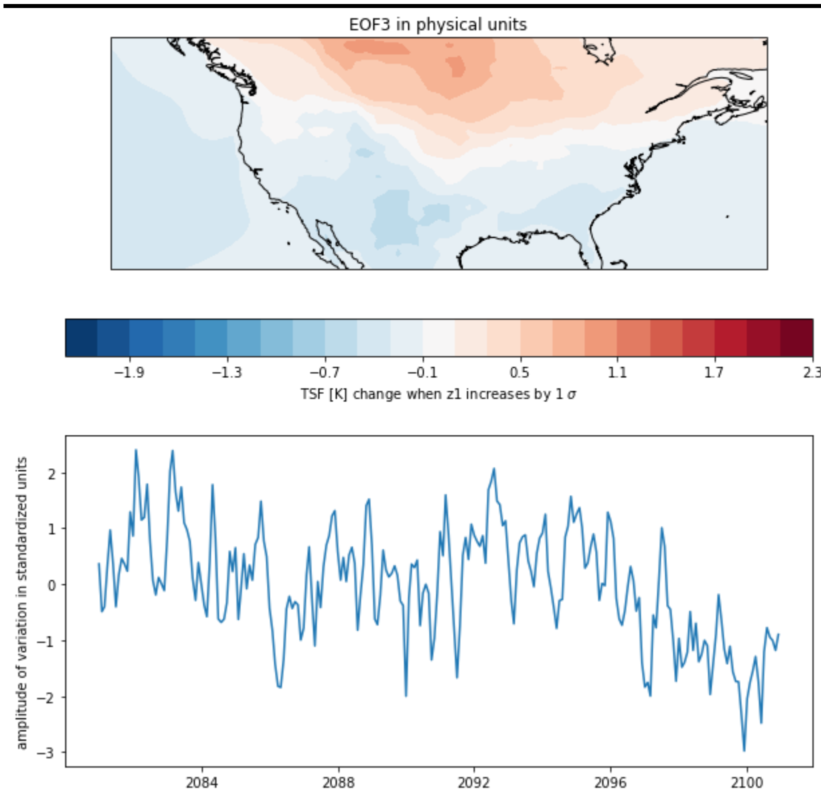


The first EOF which explains the most variance finds the strongest temperatures anomalies across the central to northeast continental United States into Canada with anomalies of opposite sign offshore of the West Coast. The second EOF finds strong temperature anomalies across the northwestern portion of the continental United States with anoamlies of the opposite sign in the southeastern portion of the United States. Using the SVD method in EOF2, the temperature gradient has flipped, although the signs of the principle components reverse to account for this flip. Thus, both of these maps are showing identical features. The third EOF finds strong temperature anomalies within the south-central portion of Canada with anomalies of opposite sign offshore, southeast of the east coast and southwest of the west coast. A slight bullseye of opposite anomaly sign exists across the border between the U.S. and Mexico.

5) Regress the data (unweighted data if applicable) onto standardized values of the 3 leading PCs. You should have one regression pattern for each PC – i.e., the EOF pattern associated with a 1 standard deviation anomaly of the PC. Plot the first three EOFs in physical units and their associated principal component timeseries. (10 points)




EOF2 in physical units
−1.9 −1.3 −0.7 −0.1 0.5 1.1 1.7 2.3
TSF [K] change when z1 increases by 1 σ
amplitude of variation in standardized units
2084 2088 2092 2096 2100


EOF3 in physical units
−1.9 −1.3 −0.7 −0.1 0.5 1.1 1.7 2.3
TSF [K] change when z1 increases by 1 σ
amplitude of variation in standardized units
2084 2088 2092 2096 2100

6) Discuss your results. Provide a physical interpretation for the first few EOFs. Which EOFs (spatial patterns) are physically significant – Which ones just look like noise? What do the PC timeseries for the EOFs tell you? Was removing the seasonal cycle from the data useful? What happens when you do not remove the seasonal cycle? (10 points). Suggested length is ~300 words.

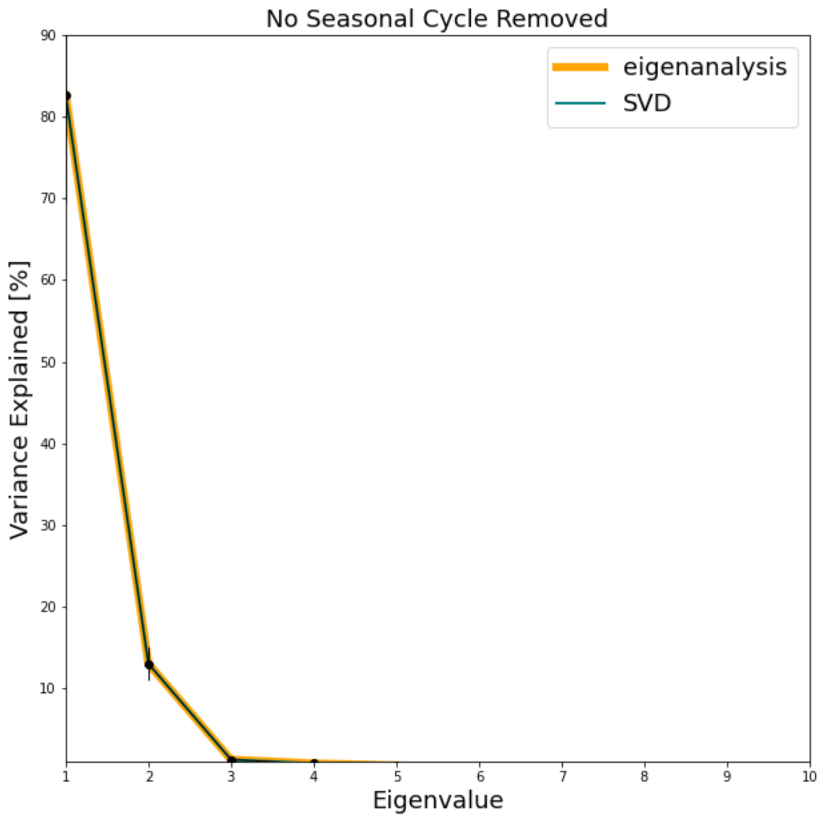
The first EOF shows a dominant mode of temperature anomaly within the mid-central United States. This region extends south towards the southern United States. These temperature anomalies are stronger over land and are more negligible offshore of both the east and west coasts. A region of anomalies with the opposite are found over the northern Pacific basin. Thus, temperature anomalies occur within the continental United States and Canada with a small region explained in the northern Pacific basin. The PC timeseries corresponding to this structure does not show a strong seasonal pattern. For example, a stationary high over the north-central U.S. may cause abnormally warm temperatures due to decreased cloud cover and positive PC values. Stationary lows that exist in the southwest U.S. can also advect warmer temperatures from the south into this region. Conversely, a strong polar vortex or frontal movement can cause cooler temperatures and negative PC values. This bullseye of strong variance occurs in the center of the continent where the influence of moisture reduces, causing larger temperature variance.

The second EOF structure shows large temperature anomalies in the northwestern portion of North America and anomalies of opposite sign across the southeastern portion of the U.S. This structure helps explain variance by introducing a diagonalized axis to represent the temperature gradient across the U.S. The reason this structure creates a basis that explains much of the variance is that it allows temperature gradients to no longer just change in the north and south direction. This mode may be dominant due to the Pacific Decadal Oscillation of the jet stream structure where stationary highs and lows in the northwest influence temperature.

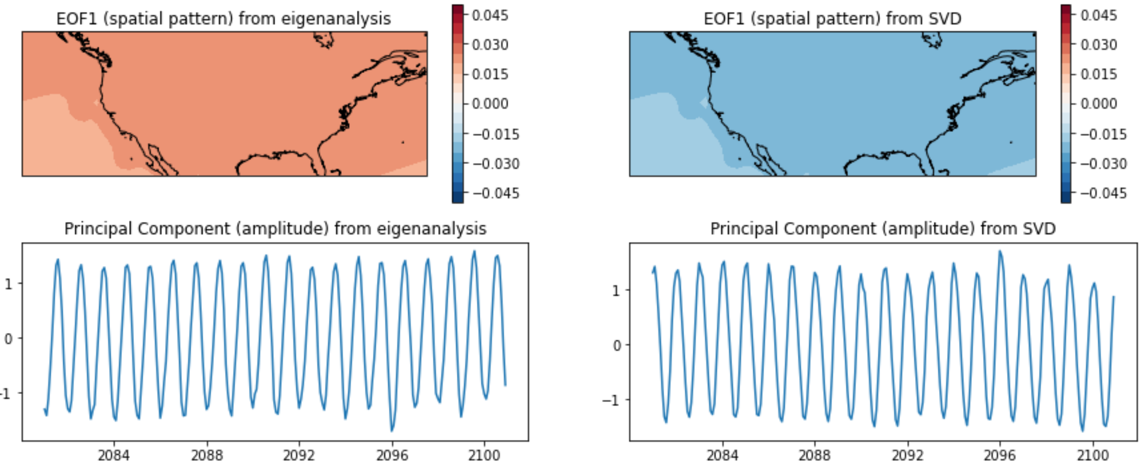
The third EOF introduces a basis that allows temperatures anomalies to change sign between the northern and southern portions of the U.S. This basis is useful for explaining frontal propagation. Many U.S. cold fronts originate in Canada and many U.S. warm fronts originate in Mexico and the Gulf of Mexico. The region ahead of a cold front, for example, typically features warmer-than-average temperature anomalies and the region aft of a cold front typically features cooler-than-average temperature anomalies. Thus, these three basis structures allow temperatures to vary both in the north-south and east-west directions and create any large-scale combination.

To me, the first EOF that appears to be noisy is EOF number 8. This EOF produces a wave-like pattern across the east coast continental shelf region that does not seem physical or relate to the Gulf Stream current. This EOF also produces a band of similar temperature anomalies that extend from the northwestern U.S. into the southeastern U.S. From the SVD EOF, it is possible that a synoptic high pressure system could increase temperature anomalies in the southwestern U.S., which helps funnel cooler air from the northwest to the southeast due to anticyclonic rotation, creating the pattern we see. However, this pattern is unlikely for negative PCs, which appear to occur relatively frequently in the PC timeseries. Using data with the seasonal cycle included, I would argue that the 2nd EOF is noisy as it hardly explains any variance in the cycle.

For this application, removing the seasonal cycle is useful. From the plot in problem 1, we see that the majority of variation across the timeseries occurs from the seasonal cycle with relatively higher surface temperatures in the summer and relatively lower surface temperatures in winter. Once the seasonal cycle is removed, small-scale temporal variations become obvious. For example, relative highs in surface temperature anomalies that occur in January of 2089 are not visible in the timeseries with the seasonal cycle present as the surface temperature is relatively low in January. When the seasonal cycle is not removed, we also see that over 80% of the variance can be explained by the first EOF.



Furthermore, the first EOF which explains the most variance provides no useful information when the seasonal cycle is not removed. The plot below shows that temperature anomalies are explained by relatively warm or cool temperatures across the entirety of North America. The resulting PCs show a seasonal trend, where negative PC values occur during warm months and positive PCs occur during cool months (left) and vice versa (right). This means that temperatures are generally warmer during warmer months and generally cooler during cooler months. Thus, this EOF analysis provides seasonal trends only and not spatial patterns.


EOF1 (spatial pattern) from eigenanalysis
EOF1 (spatial pattern) from SVD
0.045
0.030
0.015
0.000
−0.015
−0.030
−0.045
Principal Component (amplitude) from eigenanalysis
Principal Component (amplitude) from SVD
1
0
−1
2084
2088
2092
2096
2100

Problem II) Apply EOF-PCA analysis to a dataset of your choice (50 points)

a) **Provide a thorough description of your data including the reference, the variable (including names and units), the sampling frequency, etc. Describe why you think EOF-PCA analysis may provide useful information. (10 points)**

This dataset is a 3-dimensional timeseries of Weather Research and Forecasting (WRF) model output. The simulation ran for an entire year over the U.S. Outer Continental Shelf region with an output frequency of 10 minutes. There are 54 levels from the surface to a height of 50 hPa, 259 points meridionally, and 466 points zonally with a horizontal resolution of 2 km. We acquire wind direction in units of degrees at the height of 138 m only, as this is the hub height of the 12MW turbines that will be installed in coming years. Furthermore, we use output data from every 5th grid cell to prevent crashes during analysis.

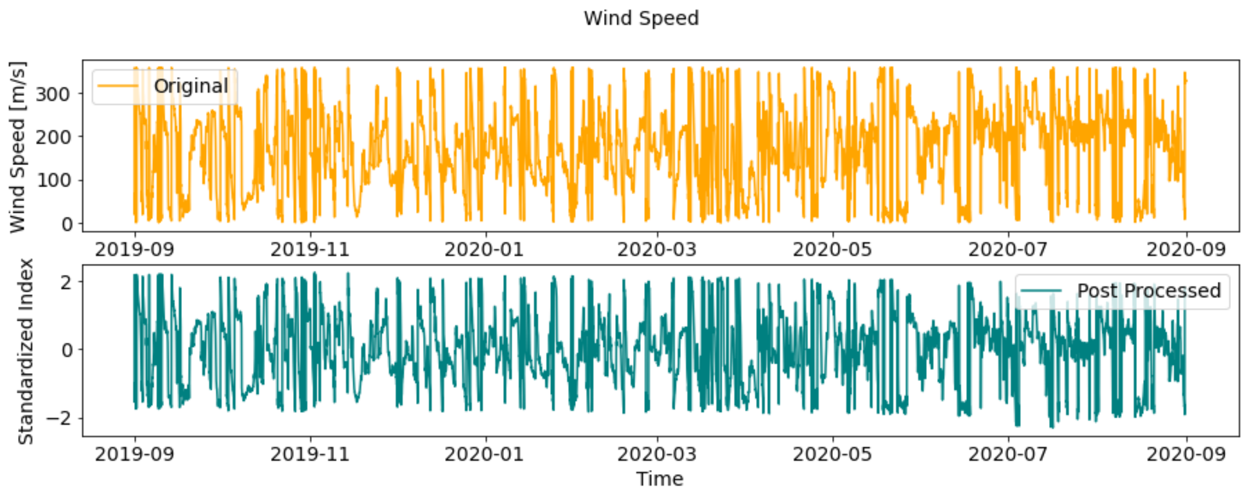
EOF-PCA analysis can provide useful information for wind farm planners. Planners want to site turbines in locations where the wind speed is fast and consistent. Areas with large wind speed variability can cause fluctuations in power output making it difficult to supply enough power to the grid to meet demand. Furthermore, wind turbines extract momentum from the wind to produce electrical power. This loss of momentum causes the wind speed to reduce downwind of each turbine, creating a "wake". So, wakes created by upwind turbines reduce the power production at downwind turbines. Although the wind direction is predominantly southwesterly in this region (Bodini et al., 2019), there may be other dominant modes of variability that can influence the propagation direction of wakes that induce power losses for wind farms.

b) **Explain what you expect to see in your EOF-PCA analysis. Do this **before** you do the analysis. What is your hypothesis? (There is not any penalty for it being wrong…!) (10 points)**

I expect that this EOF-PCA analysis will show two dominant modes of variability that differ from the mean southwesterly wind direction. I think one of these two modes will be cyclonic turning of wind speeds across the continental shelf caused by the frequent storms in the region. If one of the EOF structures contains this pattern, I will be looking out for PCs that show positive values during the cooler months each year since the midlatitudes feature more frequent storms during the winter. I think another EOF will show northwesterly wind directions caused by the diurnal and seasonal heating contrast between land and the nearby ocean.

c) **Perform EOF analysis following the steps in problem I. (20 points)**

First, we weight the data by the square root of the cosine of latitude following Hannachi et al. (2007). This weighting considers conversion at the poles. The anomalous quantities of wind direction are found by subtracting the population mean from each timestamp. We then remove the seasonal cycle by subtracting a moving monthly average for each month. However, removing the seasonal cycle has little to no impact in this case. This also acts as the first step for standardizing the data. Finally, the dataset is fully standardized by normalizing to the standard deviation. The plots below show the original wind direction timeseries for the full period (top) and the post-processed wind direction for the full period (bottom) taken at an arbitrary coordinate in space.



<u>Eigenanalysis EOF Method</u>

We first calculate the covariance matrix from Barnes Chpt. 3 Eq. 7:

$\boldsymbol{C} = \frac{1}{M}\boldsymbol{X}^T\boldsymbol{X}$

Next we calculate the eigenvalues (Λ) and eigenvectors ($\boldsymbol{E}$) from Barnes Chpt. 3 Eq. 42:

$\boldsymbol{CE} = \boldsymbol{E}\Lambda$ where individual eigenvectors ($\boldsymbol{e_i}$) correspond to eigenvalues (λ_1)

The percent of variance explained (pve) by each eigenvalue is calculated by:

$pve = 100 * \frac{|\lambda_i|}{\sum|\lambda_j|}$

<u>SVD EOF Method</u>

We decompose wind direcrion ($\boldsymbol{X}$) into into the produce of three matrices using Barnes Chpt. 3 Eq. 65:

$\boldsymbol{X} = \boldsymbol{U\Sigma V^T}$ where the columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{XX^T}$ and the rows of $\boldsymbol{V^T}$ are eigenvectors of $\boldsymbol{X^T X}$

Here, $\boldsymbol{\Sigma}$ contains the square root of eigenvalues λ_i so we calculate percent variance explained by:

$$pve = 100 * \frac{\boldsymbol{\Sigma}_{\mathrm{i}}^2}{\sum \boldsymbol{\Sigma}_{\mathrm{j}}^2}$$

Next, we calculate the autocorrelation from Barnes Chpt. 2 Eq. 67:

$$\gamma(t) = \frac{1}{N-\tau}\sum x'(t) \cdot x'(t+\tau)$$

where N is the dataset size, t is the timestamp of interest, $x(t)$ is the post-processed surface temperature data at that timestep, τ is the lag of 1. This covariance is normalized to the population variance at a lag of 0. Next, the effective sample size is calculated from Barnes Chpt. 2 Eq. 90:

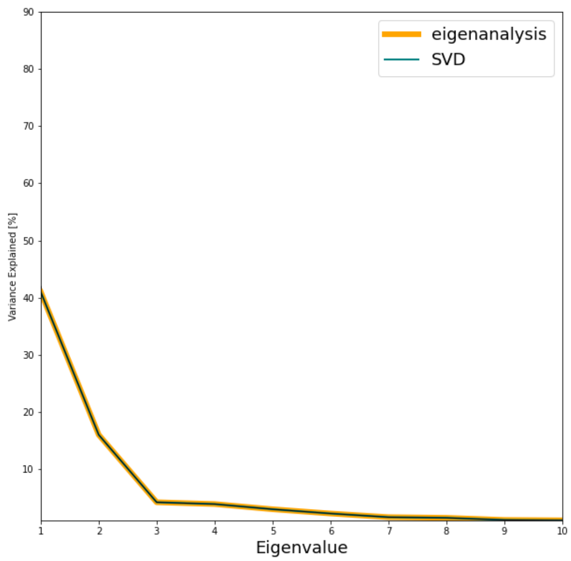
$$N^* = -\frac{N}{2}\ln\big(\gamma(t)\big)$$

Autocorrelation and N^* values are calculated individually for each longitude/latitude pair across the full period of interest. Here, the mean autocorrelation is 0.859 and the mean N^* is

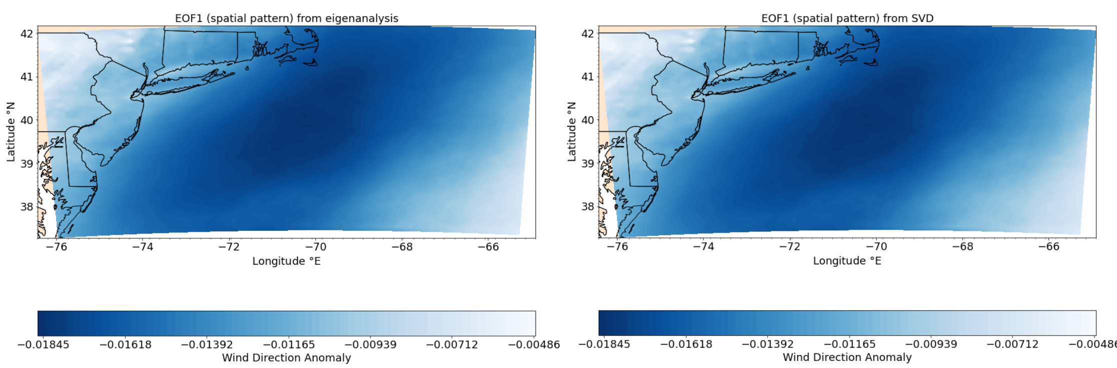
670, a 92% decrease in size. Last, we calculate the 95% confidence bounds on individual eigenvalues from Barnes Chpt. 3 Eq. 80:

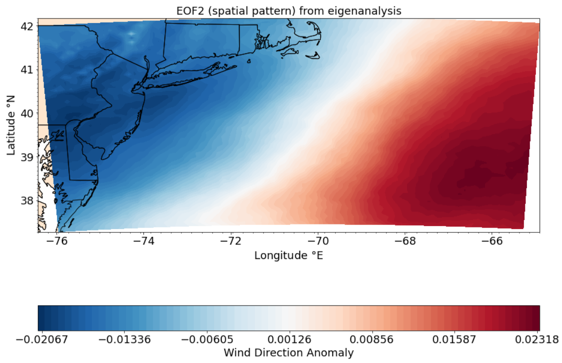
$$\Delta\lambda = \lambda_i \sqrt{\frac{2}{N^*}}$$

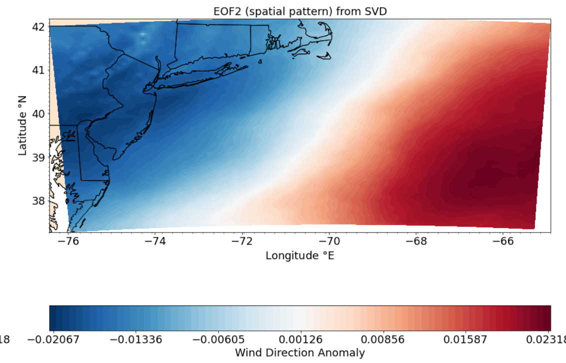
From this, the first three eigenvalues explain 41%, 16%, and 4%, respectively. The error bars are small enough that they are nearly indistinguishable. For example, the first three error eigenvalues have errors of 2.2°, 0.87°, and 0.2°, respectively.

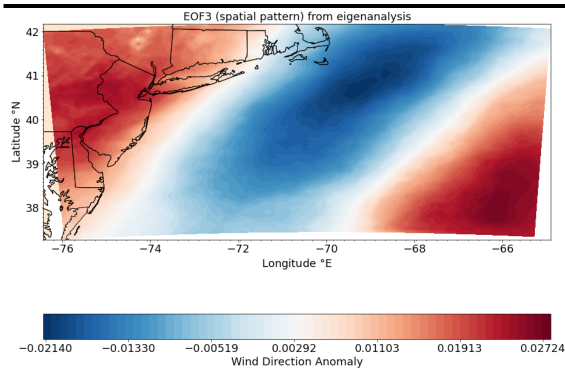


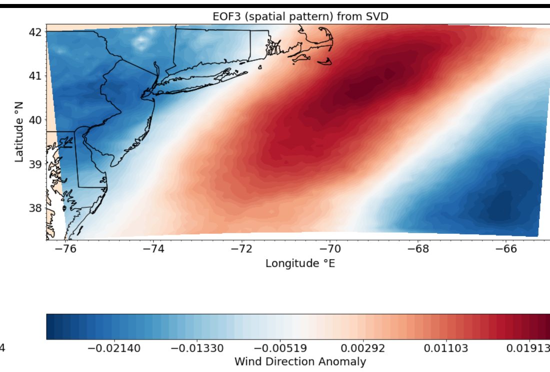
The following plots show the first three EOFs using both the eigenanalysis (left) and SVD (right) methods. PC timeseries are provided with the EOF plots in physical units.



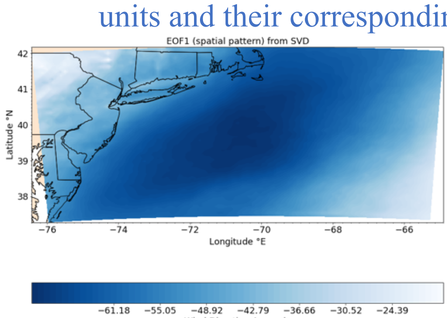


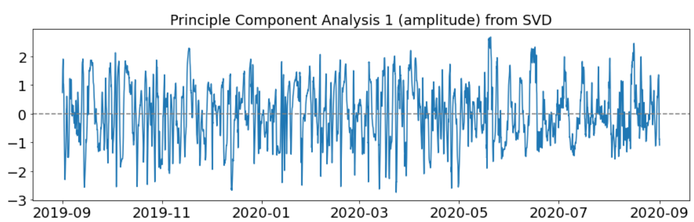


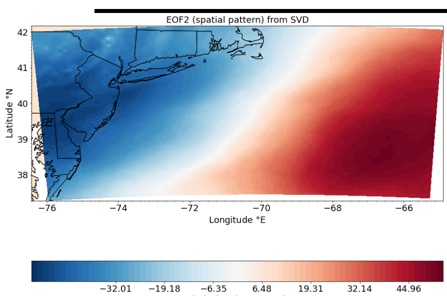


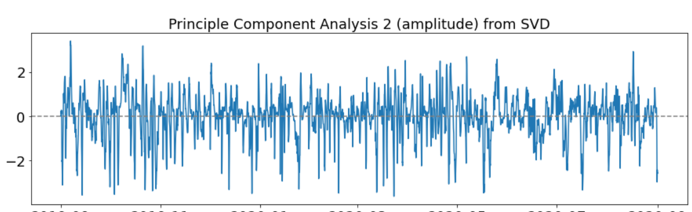


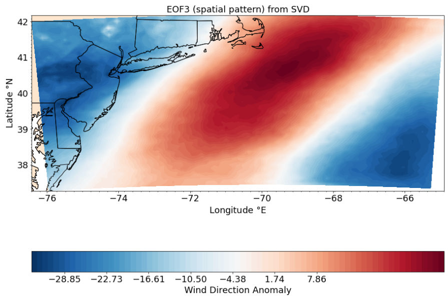
The following three plots show eigenvector structures from the SVD method in physical units and their corresponding PC timeseries on the right.

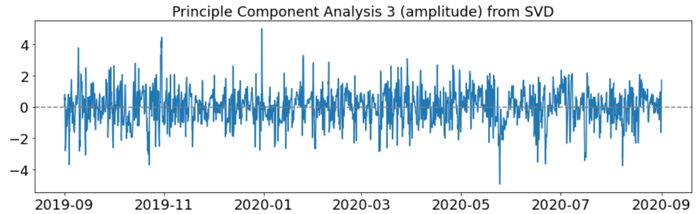




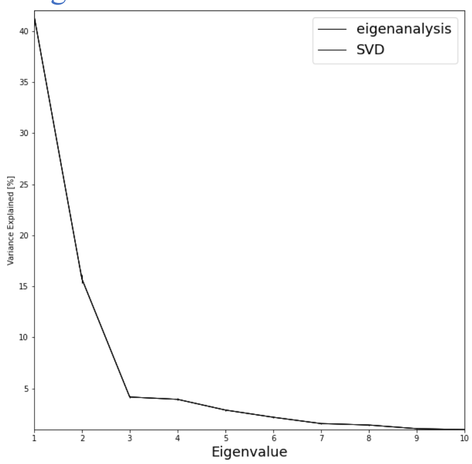


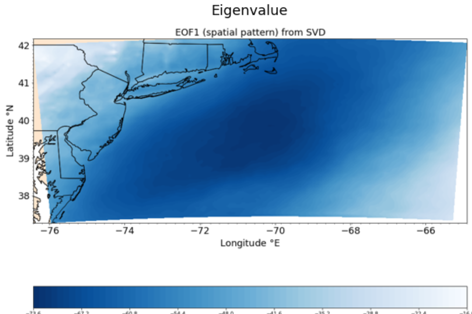


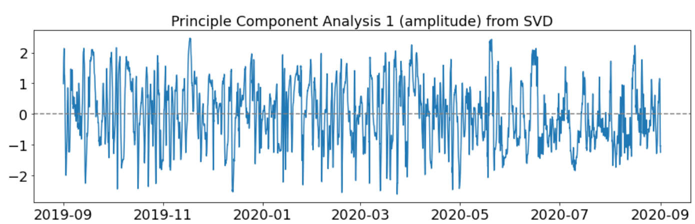




The plots below show the first 10 eigenvalues, the first EOF structure, and the first PC timeseries from the dataset without the seasonal cycle removed. Using the same methods, the autocorrelation has increased from 0.859 to 0.862 and the effective sample size has reduced from 670 to 656. The variance explained by the first three eigenvalues have not changed when rounded to the nearest integer.







d) **Describe what you found. In your case - what the EOFs and PCAs physically represent? What did you learn? (10 points)**

We offer a reminder that the typical wind direction across the Outer Atlantic Continental Shelf is southwesterly, or parallel to the shoreline. The first EOF, which explains 41% of the variance, shows that the most common anomaly relative to the mean flow is winds that veer across the near-offshore region. The amount of veer reduces onshore and in the far-offshore region. The maximum perturbation in EOF1 is almost 70°, which indicates that veering is nearly perpendicular to the shoreline. This finding is consistent with the second hypothesis that sea-breezes and land-breezes are a dominant mode of variability. The PC timeseries for EOF1 indicates that this structure is more common with positive values (flow towards land) during the warmer months and negative values (flow towards sea) during the cooler months. During the warm months, the land surface features higher temperatures than the ocean due to the lower heat capacity, inducing lower pressure and winds that flow towards shore. During the cooler months, the land surface cools more, induces higher pressure, and causes winds to flow towards the ocean. This is consistent with negative PC values during the cooler months.

EOF2 shows winds veering nearly 50° onshore, an abrupt gradient in the near-offshore region, and winds that veer 50° in the opposite direction in the far-offshore region. The PC timeseries for this structure shows more frequent positive values during the winter months and more frequent negative values during the summer months. During the winter, the jet stream translates to the south, inducing mid-latitude storms and thus more frequent lower pressure. The land surface has a higher roughness length which increases friction on wind flow. When friction increases, winds flow ageostrophically towards the center of low-pressure systems which may be why the bullseye of counterclockwise turning of winds appears over land. When low pressure systems occur over land, the pressure gradient between land and sea may be larger due to the presence of the North Atlantic Gyre. This gyre features higher relative pressure and winds that turn clockwise which explains the sharp gradient in the EOF structure. This feature is consistent with my first hypothesis.

EOF3 shows a similar structure to EOF2 although the region of clockwise veering has translated closer to shore. The PC timeseries values for this structure are more commonly positive during the winter months. Thus, this feature may be caused by the warm Gulf Stream inducing thermal wind flow due to the pressure gradient at the sea surface. Becuase the Gulf Stream is a warm current, this may cause winds to flow clockwise (towards the east) where the warmer sea surface has induced lower pressure and winds that flow counterclockwise (towards the west) on the other side of the Gulf Stream. However, as the Gulf Stream is almost always warmer than the surrounding ocean, this hypothesis does not explain negative PC values. As this structure only explains 4% of the variance, it may be explaining irregular meteorological conditions that are not vital to contributors to the dominant variance.

Finally, through exploring the impact of removing the seasonal cycle, we can conclude that it is not substantial for the analysis of wind direction. Removing the seasonal cycle reduces the autocorrelation by about 0.03 and increases the sample size by 4. These changes are negligible as the original sample size was ~54,200. Furthermore, the first EOF structure is nearly identical with the seasonal cycle removed and the seasonal cycle present. This indicates that wind direction does not follow a strong seasonal cycle, as would temperature.

Note for grading Problem II. You are analyzing your own data. Since only you know the "right answer", you will be largely graded on how well I can follow your description of the data, the methods, the results, and the conclusions. Keep your code and your explanations simple, clear, and easy to follow. Spend the time to make your code concise, clear, and well documented. Look *at your code as an opportunity to re-enforce the understanding you have gained in class as expressed through analyzing your own data.*

Note 2: FEELING BORED??? WANTING MORE???: Try implementing rotation (see Hartmann 4.10) or extended EOFs to take into account a time lag (Hannachi et. al. 2007, or Navarra/Simoncini book).

3) Future homework assignments will continue to require that you analyze a time series dataset. You will apply power spectra analysis and filtering to this dataset. A dataset with interesting variations at a wide range of timescales will be the most interesting. Please describe the dataset you plan to use here. Discuss with me or a classmate if you do not have a dataset in mind. We can help you brainstorm!! (0 points)

One issue in my research has been that the introduction of wind turbines in the model creates numerical noise. This numerical noise typically looks like high-frequency perturbations to wind speed that oscillate between fast and slow somewhat randomly. In my research, I applied hourly averaging to reduce the impacts of the noise, but I would be highly interested in applying power spectra analysis to find the period of oscillation of the noise. This may help me find a more robust method for mitigating the adverse impacts that the noise causes.

Another idea I have is that models typically require a “spinup” period where we throw out the data since it isn’t accurate yet. It is a general rule of thumb that we throw out the first 6 hours of data. Part of the reason we throw out the spinup period is that waves will reflect against the domain boundaries. Applying power spectra analysis to the gravity waves may be enlightening to figure out how long the spinup period actually needs to be.