

ATOC7500 – Application Lab #3
Empirical Orthogonal Function (EOF) Analysis

Note: This application lab requires netcdf4 and cartopy packages. Also – The 2020 python environment is provided – That one works on my computer. The 2022 python environment does not work (I think because xarray has been downgraded). Please check for updates on the python environment. I am providing these files early...

A reminder of the EOF/PCA Analysis Recipe – 5 steps

- 1) Prepare your data for analysis. Examples might include:**
 - a) sub-setting the global data to a smaller domain**
 - b) subtract the mean**
 - b) standardizing the data (divide by the standard deviation)**
 - d) cosine weighting (Account for the decrease in grid-box area as one approaches the pole (i.e. weight your data by the cosine of latitude))**
 - e) detrend the data**
 - f) remove the seasonal or diurnal cycle**
 - g) remove NaN – EOF analysis does not work with missing data.**
- 2) Calculate the EOFs and PCs using one of the two methods discussed in class:**
 - a) Eigenanalysis of the covariance matrix**
 - b) Singular Value Decomposition (SVD).**
- 3) Plot the first 10 eigenvalues (scaled as the percent variance explained) in order of variance explained. Add error bars following North et al. 1982. Describe how you determined the effective degrees of freedom N^* . How many statistically significant EOFs are there?**
- 4) Plot EOF patterns and PC timeseries (usually just the first three or so unless you want to look at more).**
- 5) Regress the data (unweighted data if applicable) onto standardize values of the 3 leading PCs. In other words, project the standardized principal component onto the original anomaly data X to get the EOF in physical units. You should have one regression pattern for each PC – i.e., the EOF pattern associated with a 1 standard deviation anomaly of the PC. *Note: The resulting patterns will be similar to the EOFs but not identical.***

Notebook #1 – EOF analysis using images of people

[ATOC5860_applicationlab3_eigenfaces.ipynb](#)

LEARNING GOALS:

- 1) Complete an EOF analysis using Singular Value Decomposition (SVD).
- 2) Provide a qualitative description of the results. What are the eigenvalues, the eigenvectors, and the principal components? What do you learn from each one about the space-time structure of your underlying dataset?

DATA and UNDERLYING SCIENCE:

In this notebook, you apply EOF analysis to a standard database for facial recognition: the At&t database.

<https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>

“Our Database of Faces, (formerly 'The ORL Database of Faces'), contains a set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department.

There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).”

The goal is to think a bit “out of the box” of Atmospheric and Oceanic Sciences about potential applications for the methods you are learning in this class for other applications.

Questions to guide your analysis of Notebook #1:

1) Execute all code without making any modifications. What do the EOFs (spatial patterns) tell you? What do the PCs tell you? How do you interpret what you are finding?

These EOFs indicate that the average result is a balding white man. I interpret this as a very biased dataset (towards white men). The eye and nose positions seem to be a dominant driver of variance. The face edges seem to be blurred.

2) Reconstruct a face. How many EOFs do you need to reconstruct a face from the database? Does it depend on the face that it used?

When reconstructing a face, you need more than 150 EOFs to accurately reconstruct the face.

3) Food for thought: The database contains 75% white men (<https://www.cl.cam.ac.uk/research/dtg/attarchive/facesataglance.html>).

How do you think this database limitation impacts the utility of the database for subjects who are not white men? What are some parallels that you might draw when analyzing atmospheric and oceanic sciences datasets? *Hint: Think about the limitations of extrapolation beyond the domain where you have data.*

This biased dataset really limits the ability of the EOFs to capture and recreate details for those who are not white men. I noticed that adding more EOFs to recreate faces led to better improvements for white men than others. This could indicate that the biased dataset allows the capturing of more white male details than those of others.

Notebook #2 – EOF analysis of Observed North Pacific Sea Surface Temperatures

[ATOC5860_applicationlab3_eof_analysis_cosineweighting_cartopy.ipynb](#)

LEARNING GOALS:

- 1) Complete an EOF analysis using the two methods discussed in class: eigenanalysis of the covariance matrix, Singular Value Decomposition (SVD).
- 2) Assess the statistical significance of the results, including estimating the effective sample size.
- 3) Provide a qualitative description of the results. What are the eigenvalue, the eigenvector, and the principal component? What do you learn from each one about the space-time structure of your underlying dataset?
- 4) Assess influence of data preparation on EOF results. What happens when you remove the seasonal cycle? What happens when you detrend? What happens when you cosine weight by latitude? What happens when you standardize your data (divide by standard deviation)? What happens when you compute anomalies?

DATA and UNDERLYING SCIENCE:

In this notebook, you will analyze observed monthly sea surface temperatures from HadISST (<http://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>). The data are in netcdf format in a file called HadISST_sst.nc. *Note that this file is ~500 MB so it might take a bit of time to download.* You will subset the data to only look at the North Pacific. Depending on how you prepare your data for analysis – you might expect to see different spatial patterns (eigenvectors) and different time series (principal components). Some things you might look for in your results are the Pacific Decadal Oscillation, “global warming”, the seasonal cycle, Depending on your data preparation – your hypothesis for what you should see in your EOF analysis should change. Note: In this dataset - land is NaN, sea ice is -999 – the notebook sets all values over land and sea ice to 0 for the EOF analysis.

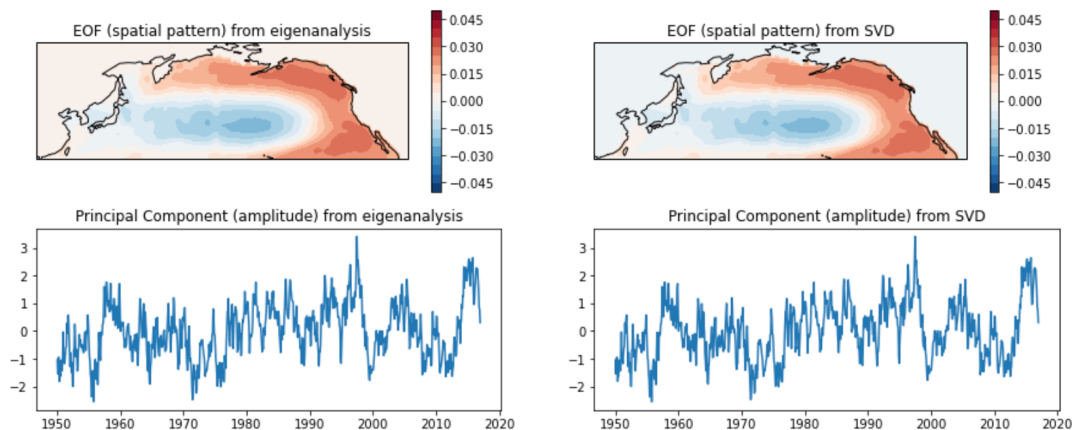
Questions to guide your analysis of Notebook #1:

1) Your first time through the notebook – Execute all code without making any modifications. Provide a physical interpretation for at least the first two EOFs and principal components (PC). What do the EOFs (spatial patterns) tell you? What do the PC time series for the EOFs tell you? What do you think of the method for estimating the effective sample size (Nstar)? Can you propose an alternative way to estimate Nstar? Do you get the same results using eigenanalysis and SVD? If you got a different sign do you think that is meaningful?.

The EOFs indicate a dipole. The first EOF has a longer frequency of variability as compared to the second EOF. The PC1 seems to have a trend.

They generate N* by getting a spatial average, but I don't think this is necessarily the best method. You could potentially standardize all data in space rather than just over time.

We got the same signs for SVD and eigenanalysis here!

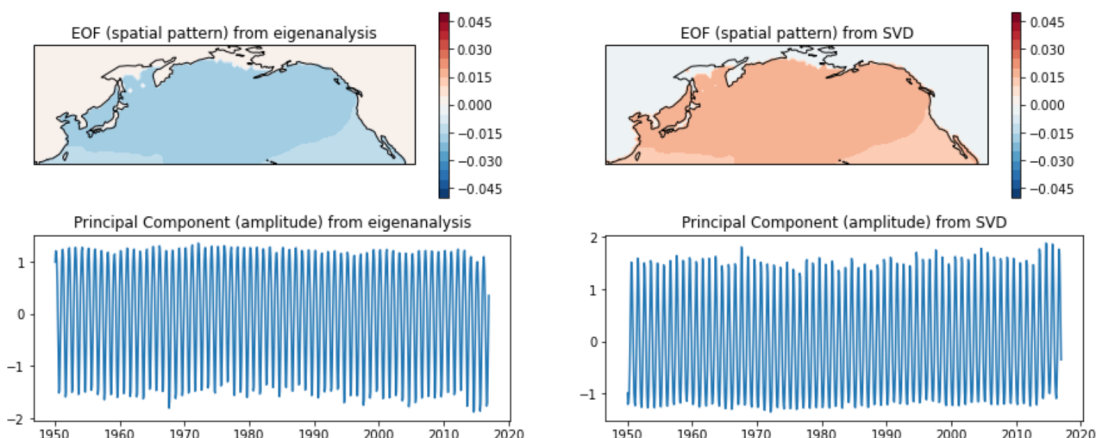


2) Save a copy of the notebook, rename it. Repeat the analysis but this time do not remove the seasonal cycle. What do you think you will see? Discuss your results with your neighbor. How do the EOFs and PC change? Was removing the seasonal cycle from the data useful? What impacts does removing the seasonal cycle have on your analysis?

The first EOF only represents the seasonal cycle. This is bad! The second EOF looks more like the PDO (what we want!). This second is much lower frequency than the first.

Removing the seasonal cycle is a good thing as it allows to see variability outside of that which we know to expect (seasonality!).

We got opposite signs here, although the magnitudes are the same on either side of zero.

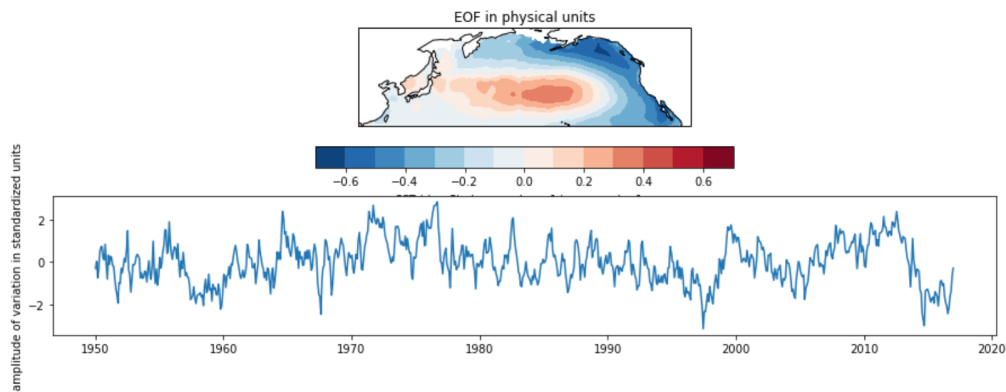


3) Save a copy of the notebook, rename it. Repeat the analysis but this time detrend the data. Discuss your results. How do the EOFs and PC change? Was

detrending the data useful? What impacts does detrending have on your analysis?

This looks like the EOF1 of the original data in that it likely represents the EOF. I think detrending is probably a good idea as climate change is dominant in most of the earth system.

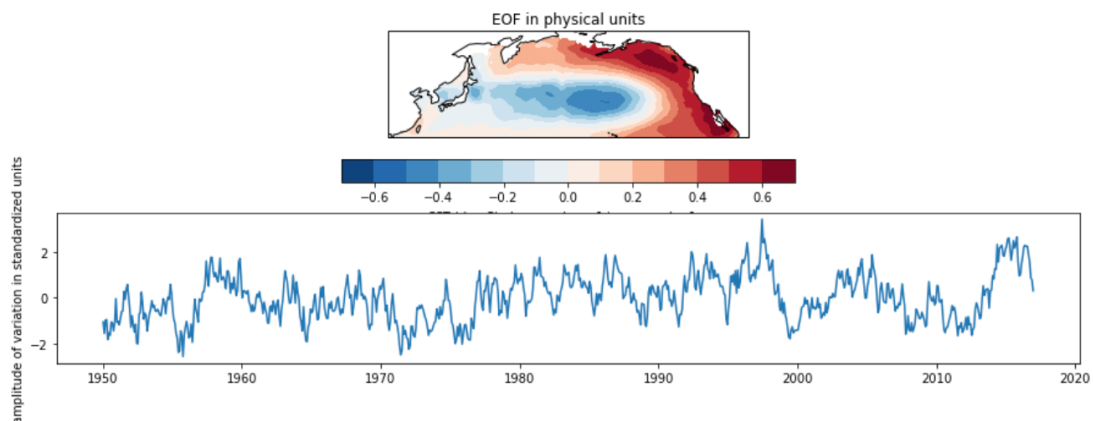
We got opposite signs here, although the magnitudes are the same on either side of zero.



4) Save a copy of the notebook, rename it. Repeat the analysis but this time do not apply the cosine weighting. Discuss your results. How do the EOFs and PC change? Was cosine weighting the data useful? What impacts does cosine weighting have on your analysis? What are examples of analyses where cosine weighting would be more/less important to do?

The cosine weighting did not change anything dramatic in the data.

We got opposite signs here, although the magnitudes are the same on either side of zero.



4) Save a copy of the notebook, rename it. Repeat the analysis but this time do not standardize the data (i.e., comment out dividing by standard deviation).

Discuss your results. How do the EOFs and PC change? Was standardizing the data useful? What impacts does standardizing the data have on your analysis?

The first mode worked about the same as before, although the second EOF is much different. Standardizing ensures that variance noticed works.

We got opposite signs here, although the magnitudes are the same on either side of zero.

