Overall

- Formatting is free of obvious errors (equations, latex, margins, etc..)
- Report is submitted on time
- Code base is accessible

Title Page

- Include this checklist as first page (LaTeX also available)
- Descriptive, accurate title (under 20 words)
- Student name and NetID
- Date (5 points for on-time submissions)
- Lab partners listed (if applicable)

Abstract

- Purpose and objectives clearly stated
- Methods summarized concisely
- Key results highlighted with numbers
- Main conclusion stated
- Less more 200 words

Introduction

- Between two and five paragraphs
- Relevant background research provided
- Research question clearly stated

Data & Methods

- Data sources documented
- Data preprocessing steps documented
- Statistical methods described
- Experimental setup documented
- Includes conceptual model or diagram of methods

Results

- All tables numbered and labeled
- All figures numbered and labeled
- Statistical analysis included where appropriate
- Units clearly marked
- Error bars shown where appropriate
- Results described without interpretation *unless* section is labeled as “Results & Discussion”

Discussion

- Results interpreted
- Comparison to literature included as relevant
- Sources of error, uncertainty, and or statistical significance discussed
- Limitations acknowledged
- Future research suggested

Conclusion

- Key findings summarized
- Research question answered
- Broader implications stated
- Practical applications noted
- Future directions suggested

References

- Includes all references cited in text
- References are current/relevant
- References are varied as relevant (include older papers when appropriate)

Figures

- Generally interpretable from caption, labels, and legend
- Quality of graphics is aesthetically pleasing
- Font sizes on axes, legends are legible

Lab 0: EOF Analysis of Global SST Patterns
EAS 6995 Deep Learning in Earth and Environmental Science

Title:
Empirical Orthogonal Function (EOF) Analysis of Global Sea Surface Temperature (SST) Variability

Author:
Yifan Luo, yl3699

Date:
February 5, 2025

Abstract

This lab applies Empirical Orthogonal Function (EOF) analysis to identify dominant spatial and temporal variability patterns in global sea surface temperature (SST) data. Key steps included preprocessing the dataset (removing seasonal cycles, latitude weighting), computing EOFs, and interpreting results in the context of climate phenomena. The first three EOFs explain 30.7%, 26.6%, and 10.5% of the variance, cumulatively accounting for 67.8% of total variability. Notably, the second principal component (PC2) exhibits a strong positive correlation with the Niño Index strongly (r = 0.93), indicating its association with El Niño-Southern Oscillation (ENSO). These results highlight the utility of EOF analysis in isolating climate modes from complex spatiotemporal data.

---

1. Introduction

In the realm of climate science, global sea surface temperature (SST) anomalies are known to drive large-scale atmospheric circulations and impact weather patterns worldwide. One of the most notable examples is the El Niño–Southern Oscillation (ENSO), an interannual mode of ocean-atmosphere variability that produces pronounced warming or cooling in the central-eastern tropical Pacific and influences rainfall, tropical cyclones, and global temperature anomalies (Trenberth, 1997). However, ENSO is not the only driver of climate variability; phenomena such as the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation also contribute to the spatiotemporal complexity of the global climate system (Deser et al., 2010).

Empirical Orthogonal Function (EOF) analysis, also referred to as Principal Component Analysis (PCA) in statistics, provides a powerful way to extract these dominant modes of variability from large datasets (von Storch & Zwiers, 1999). Widely applied to sea surface temperature (SST) datasets, EOF analysis has proven instrumental in identifying large-scale climate phenomena such as the El Niño-Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) (Huang & Xie, 2015; Werb & Rudnick, 2023). By decomposing spatiotemporal fields into orthogonal patterns, EOF analysis helps clarify which regions covary in time and how strongly certain modes contribute to the total variance.

This lab applies EOF analysis to global SST data to address the following research question: What are the dominant modes of SST variability, and how do they correlate with ENSO activity? Building on methodologies established in prior studies, I preprocess the data by removing seasonal cycles and applying latitude weighting to mitigate spatial biases. The analysis aims to (1) identify the leading EOF patterns, (2) quantify their explained variance, and (3) assess linkages between principal components (PCs) and ENSO through correlation with the Niño Index. By contextualizing results within existing literature, this work seeks to reinforce the utility of EOF analysis in climate science while contributing new insights into SST variability drivers.

2. Data & Methods

2.1 Data Sources

The primary dataset for this analysis is the COBE Sea Surface Temperature Dataset from the Japanese Meteorological Center (Ishii et al., 2005), which spans from 1980 to 2019 at monthly resolution. This dataset covers the global oceans on a regular 1.0° latitude by 1.0° longitude grid, providing 360 grid points in the zonal (east–west) direction and 180 grid points in the meridional (north–south) direction, extending from 89.5°N to 89.5°S and 0.5°E to 359.5°E. The SST is taken at sea level, and the dataset is updated monthly.

As part of the analysis, I also utilize the Niño 3.4 index obtained from the NOAA Physical Sciences Laboratory (PSL), where the Niño 3.4 region (5°N–5°S, 170°W–120°W) serves as a standard measure for monitoring the El Niño–Southern Oscillation (ENSO). The index is typically calculated by applying a 5-month running mean to the SST anomalies in that region; El Niño or La Niña conditions are defined when Niño 3.4

anomalies exceed ± 0.4 °C for six consecutive months or more. This index, therefore, provides a well-established benchmark for correlating principal component time series with known interannual tropical Pacific variability.

2.2 Preprocessing

I begin by extracting the sea surface temperature (SST) variable from the original dataset, ensuring that it is properly indexed by time, latitude, and longitude. To account for the changing geographic area associated with each latitude, I compute a latitude-dependent weight given by $W(s) = \sqrt{cos\,(\varphi)}$, where φ is the latitude in radians. These weights are expanded to three dimensions so that each spatiotemporal SST field can be multiplied point-by-point by the corresponding weight, thus placing appropriate emphasis on lower latitudes (small grid-cell areas) versus higher latitudes (larger grid-cell areas).

Next, a mean seasonal cycle or monthly climatology is calculated by grouping all SST values by month of the year and taking the mean over the full record. Subtracting this monthly climatology from the original data removes the seasonal cycle, yielding a deseasonalized SST field. At this stage, I apply the latitude weighting so that the deseasonalized data are scaled accordingly in each grid cell. Any missing values are then handled using a linear interpolation in the time dimension, thus reducing potential gaps in the dataset.

Finally, I reshape the resulting three-dimensional array (with coordinates of time, latitude, and longitude) into a two-dimensional matrix of size $(T \times S)$, where T is the total number of time steps and S is the product of the latitudinal and longitudinal grid sizes. This new matrix form is suitable for EOF computation in the subsequent steps.

2.3 EOF Computation

The first step is to construct the spatial covariance matrix, $\mathbf{C}$, from the data matrix $\mathbf{X}$ of size $(T \times S)$, where T represents the number of time steps and S the number of spatial points. This data matrix contains anomalies (after any preprocessing such as seasonal cycle removal and latitude weighting). The covariance matrix is given by $\mathbf{C} = \frac{1}{T-1}\mathbf{X^T X}$, which has dimensions $(S \times S)$. Each entry in $\mathbf{C}$ measures how variations at two spatial points co-vary over time.

To extract modes of variability, I solve the eigenvalue problem $\mathbf{C\,U = U\,\Lambda}$, where $\mathbf{U}$ is the matrix, whose columns are the eigenvectors, and $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues λ_i. Since $\mathbf{C}$ is a real and symmetric matrix, its eigenvalues are real and non-negative. To find these eigenvalues, one solves the characteristic equation $\det(\mathbf{C} - \lambda\mathbf{I}) = 0$, which yields λ_i. Each eigenvector $\mathbf{u_i}$ is then determined by the equation $(\mathbf{C} - \lambda_i\,\mathbf{I})\,\mathbf{u_i} = 0$. By convention, the eigenvalues and eigenvectors are sorted in descending order of λ_i, ensuring that the first eigenvector and eigenvalue represent the mode with the highest variance explained.

Once the eigenvector matrix $\mathbf{U}$ is obtained, I compute the principal components by multiplying the original data matrix $\mathbf{X}$ by $\mathbf{U}$, namely $\mathbf{PC = X\,U}$.

Each column of $\mathbf{PC}$ is a principal component time series, also known as a PC. The corresponding column of $\mathbf{U}$ (the eigenvector) is referred to as an empirical orthogonal function or EOF. This means that for a given mode i, $\mathrm{PC}(:, i)$ is the time evolution and $\mathbf{U}(:, i)$ is the spatial pattern of that mode.

Finally, to visualize the first EOF, one takes the first column of $\mathbf{U}$, reshapes it back onto the latitude–longitude grid of the dataset, and plots the resulting two-dimensional field. Mathematically, if the first column of $\mathbf{U}$ is $\mathbf{u_1}$, then $\mathrm{EOF1} = \mathrm{reshape}\left(\mathbf{u_1}, (n_{\mathrm{lat}}, n_{\mathrm{lon}})\right)$. Similarly, the first principal component time series is the first column of $\mathbf{PC}$, namely $\mathrm{PC1} = \mathbf{PC}(:,1)$, which can be plotted against time to show how the amplitude of the first spatial mode evolves through the dataset.

2.4 Comparison with Niño Index

To identify which principal component (PC) is most closely linked to ENSO, I used the monthly Niño 3.4 index from the NOAA Physical Sciences Laboratory (PSL). First, I ensured that both the SST dataset and the Niño 3.4 index covered the same time period (1980–2019) and had a consistent monthly time axis. Next, I computed Pearson correlation coefficients between each PC time series and the Niño 3.4 index. Because both the PCs and the Niño 3.4 index are already in anomaly form, no additional detrending or deseasonalization was applied to the index. For each of the first three PCs, I calculated the correlation coefficient r, then compared these values to determine which mode most strongly captured the ENSO signal. This procedure allowed me to quantitatively link the spatial patterns in the EOFs to a well-established measure of tropical Pacific SST variability.

3. Results

3.1 Mean SST and Anomalies

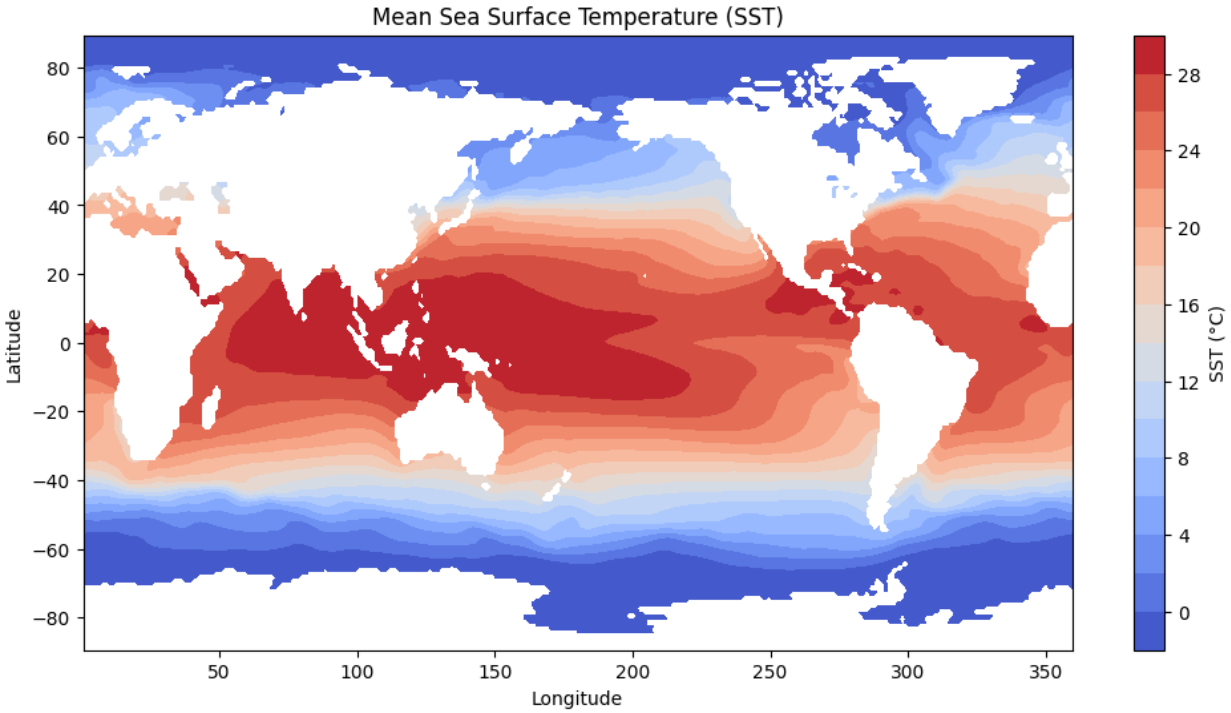

Figure 1. Global mean SST calculated over the 1980–2019 period

Figure 1 shows the long-term mean SST (°C). Warmest waters (>28 °C) appear in the tropics near the Western Pacific “Warm Pool.” High latitudes exhibit colder SST (< 5 °C), with ocean temperatures near freezing in the Arctic and Antarctic regions.

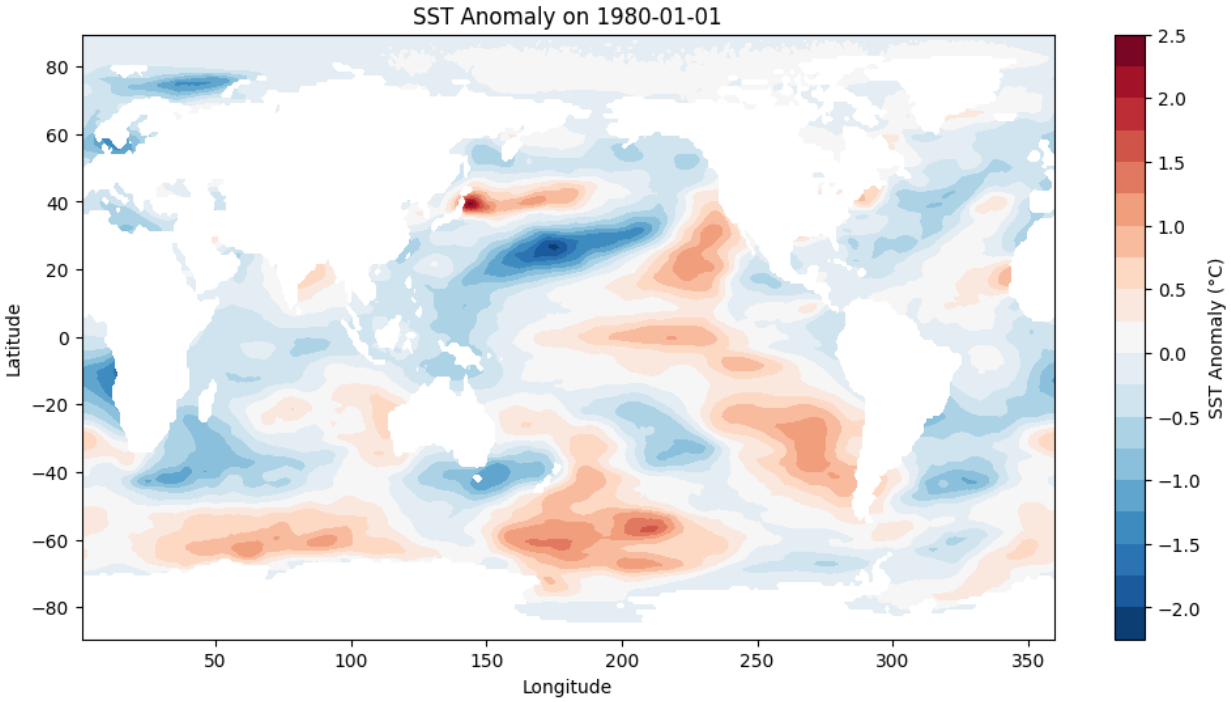

Figure 2. Global distribution of SST anomalies on January 1, 1980, relative to the long-term monthly climatology (1980–2019)

Figure 2 provides an example of an SST anomaly map on January 1, 1980. Positive anomalies (red shading) dominate some mid-latitude and subpolar areas, while negative anomalies (blue shading) appear in various ocean basins, indicating cooler-than-average conditions on this date.

3.2 EOF Spatial Patterns

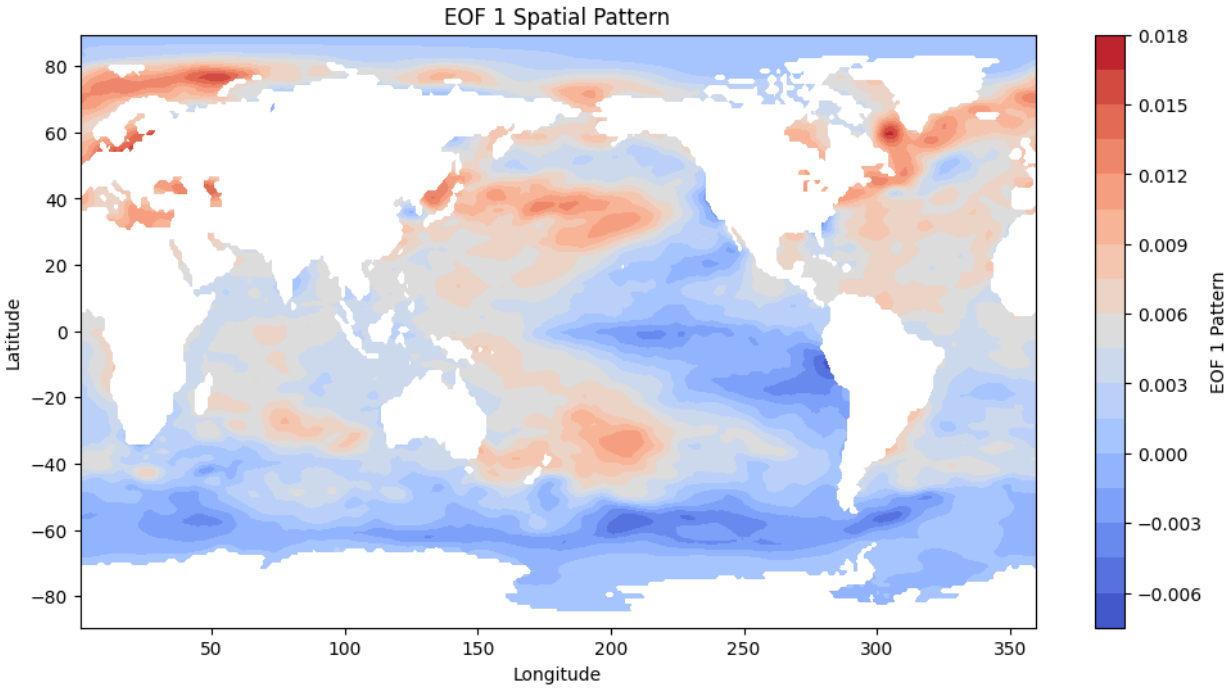

Figure 3. Spatial structure of the first empirical orthogonal function (EOF 1) of global SST anomalies, explaining approximately 30.7% of the total variance. Positive (red) and negative (blue) loadings show how different regions covary within this leading mode.

Figure 3 (EOF 1 Spatial Pattern) displays large positive loadings in subpolar regions and negative loadings in the Southern Hemisphere. This mode may reflect a broad hemispheric or global warming/cooling pattern.

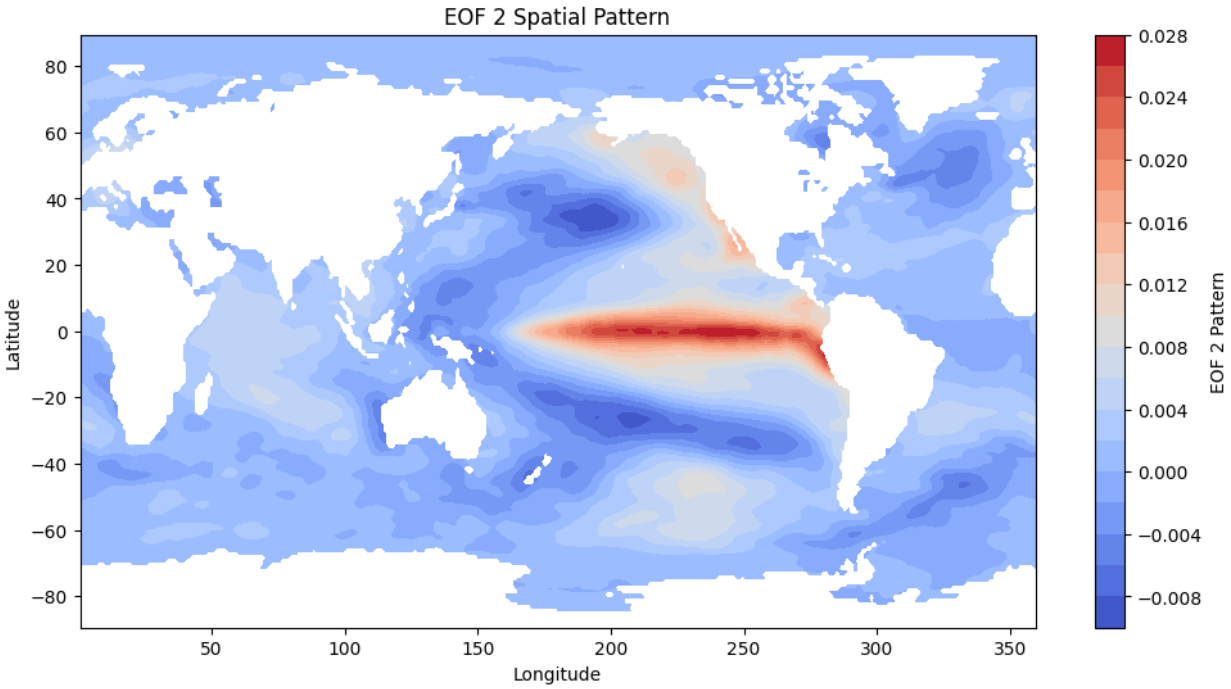

Figure 4. Spatial pattern for the second empirical orthogonal function (EOF 2) of global SST anomalies, accounting for approximately 26.6% of the total variance. The strongest positive loadings span the tropical Pacific, indicating an ENSO-like mode of variability.

Figure 4 (EOF 2 Spatial Pattern) centers in the equatorial Pacific with strong positive loadings around the Niño 3.4 region, capturing the canonical El Niño–La Niña seesaw.

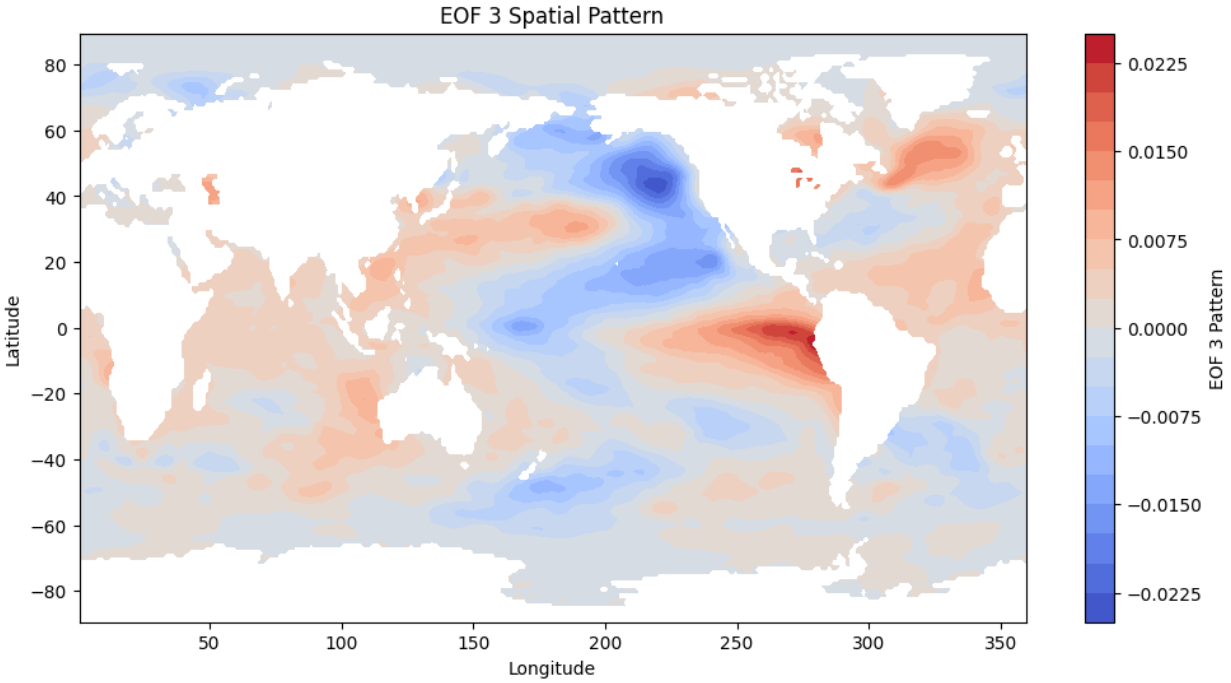

Figure 5. Spatial pattern for the third empirical orthogonal function (EOF 3) of global SST anomalies, explaining approximately 10.5% of the total variance. Prominent features include strong loadings over the high-latitude North Pacific and southwestern Atlantic

Figure 5 (EOF 3 Spatial Pattern) shows a dipole between parts of the central Pacific and the southwestern Atlantic, with additional centers in the high-latitude Northern Hemisphere, suggesting a more nuanced pattern not directly tied to ENSO events.

3.3 Scree Plot and Explained Variance

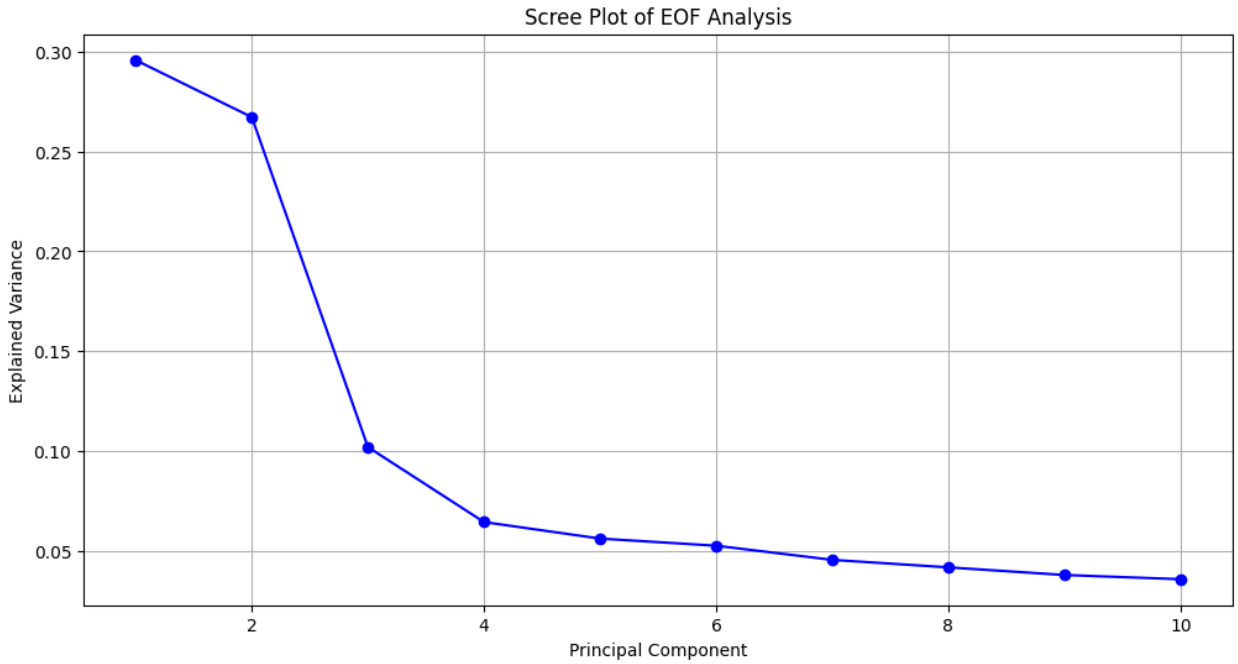

Figure 6. Scree plot showing the fraction of total variance explained by each principal component (PC). The first three modes capture the bulk of the variability, with PC1 explaining approximately 30.7%, PC2 around 26.6%, and PC3 roughly 10.5%.

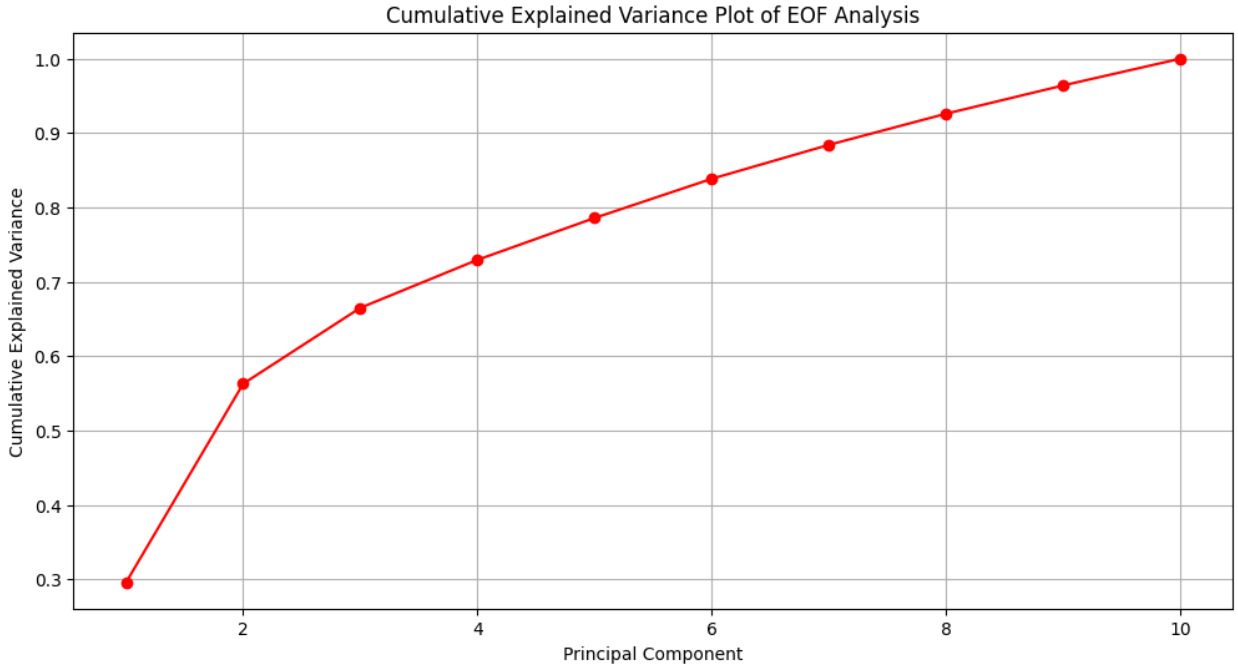

Figure 7. Cumulative explained variance for the first 10 principal components of the global SST field, showing how each additional component increases the total fraction of variance captured.

Figure 6 (Scree Plot) and **Figure 7** (Cumulative Explained Variance Plot) show the percentage of variance explained by each principal component. The first mode (PC1)

explains about **30.7%** of the variance, the second mode (PC2) adds **26.6%**, and the third mode (PC3) adds **10.5%**. By the 10th mode, the total cumulative explained variance reaches 100%.

PC	Singular Value	Eigenvalue	Explained Variance	Cumulative Variance
1	839.065002	1469.791382	0.307217	0.307217
2	780.367615	1271.343628	0.265737	0.572955
3	491.199829	503.710358	0.105286	0.678241
4	376.371002	295.730957	0.061814	0.740055
5	349.753662	255.381256	0.053380	0.793435
6	342.582581	245.016342	0.051214	0.844648
7	314.748535	206.819702	0.043230	0.887878
8	305.399231	194.715424	0.040700	0.928577
9	287.390564	172.428680	0.036041	0.964619
10	284.747131	169.271255	0.035381	1.000000

Table 1. Singular values, eigenvalues, and corresponding explained variance for the first 10 principal components of the global SST anomaly field (1980–2019).

Table 1 lists the first 10 principal components (PCs) in descending order of their explained variance. Each row shows the singular value and corresponding eigenvalue from the covariance matrix, followed by the percentage of total variance that each PC individually explains and its cumulative contribution up to that PC. The first principal component (PC1) has the largest eigenvalue, explaining about 30.7% of the total variance, while adding the second principal component (PC2) brings the cumulative variance to roughly 57.3%. By the time PC3 is included, nearly 68% of the total variance is accounted for, indicating that these three modes together describe the bulk of the dataset's spatiotemporal variability. The remaining PCs each explain progressively smaller amounts of variance, with the tenth PC bringing the cumulative variance to 100%. In practice, the rapid decline in explained variance after the first few modes suggests that the leading three PCs capture the dominant large-scale patterns, while higher-order modes may represent more localized or complex structures.

3.4 Principal Component (PC) Time Series

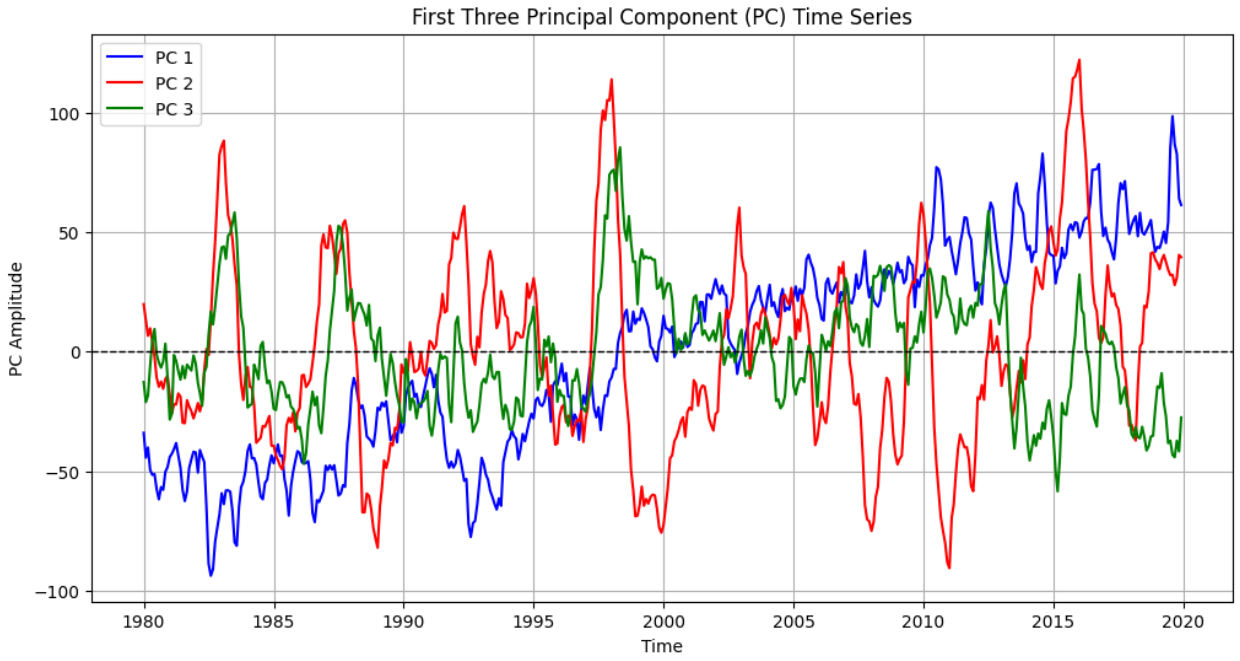

Figure 8. Time series of the first three principal components (PC1 in blue, PC2 in red, PC3 in green). These standardized PCs illustrate the temporal evolution of each EOF mode over the period 1980–2019.

Figure 8 shows the time series of the first three principal components. The red line (PC2) often shows large peaks and troughs linked to the known El Niño and La Niña episodes, while the blue line (PC1) and green line (PC3) display different, partly decadal or regional fluctuations. Notably, PC2's high amplitude around 1982–83, 1997–98, and 2015–16 coincides with major El Niño events (Golden Gate Weather Services, n.d.). By contrast, PC1 (blue) tends to vary on longer timescales, with more pronounced negative values in the early record and generally positive values after 2005, while PC3 (green) has evident swings around the mid-1990s and in the mid-2010s, suggesting a secondary or regional mode of variability.

3.5 Correlation with Niño Index

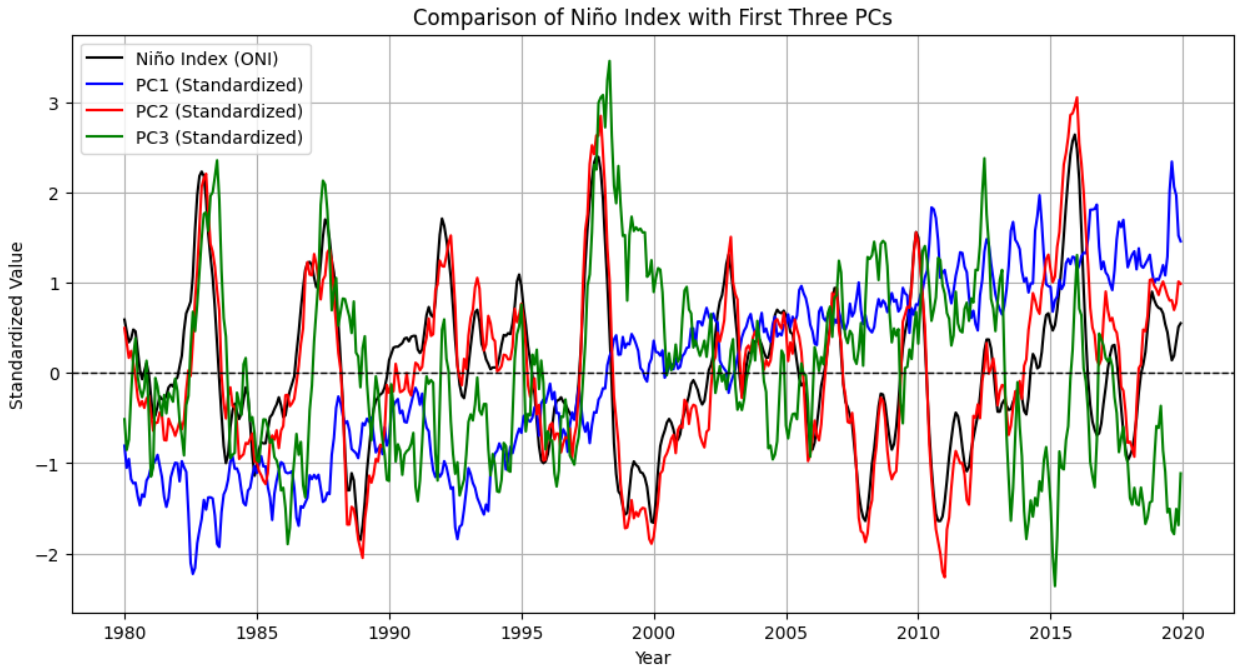

Figure 9. Comparison of the Niño 3.4 index (black) with the first three standardized principal components (PC1 in blue, PC2 in red, and PC3 in green) over the period 1980–2020. The high correspondence between PC2 and the Niño index highlights this mode's strong ENSO signal.

Figure 9 compares the standardized PCs to the Niño Index (ONI), it shows that PC2 (red) closely follows the temporal variations in the Niño 3.4 Index (black), indicating that the second mode of SST variability is strongly linked to ENSO, whereas PC1 (blue) and PC3 (green) exhibit weaker associations with the index.

PC	Correlation with Niño Index
PC1	-0.161041
PC2	0.925593
PC3	0.043799

Table 2. Pearson correlation coefficients between the first three principal components (PC1, PC2, PC3) and the Niño 3.4 Index. PC2 shows a strong positive correlation, reflecting its close association with ENSO variability.

Table 2 indicates that PC2 is strongly associated with the Niño 3.4 index ($r \approx 0.93$), confirming that the second mode largely represents ENSO-related SST variability, while PC1 and PC3 show weaker correlations.

4. Discussion

The results demonstrate that the leading modes of SST variability align with known climate patterns. In particular, EOF2's strong correlation with the Niño 3.4 index highlights its association with ENSO, consistent with previous studies (Deser et al. 2010). The first EOF captures broader spatial variability that may blend long-term trends

or additional global oscillations. Although these modes match well with expectations from the literature, some uncertainties arise from data preprocessing choices—such as the methods for deseasonalization, missing data interpolation, and latitude weighting. These steps, while standard, can introduce subtle biases in the resulting EOF patterns.

Limitations involve the reliance on a single SST dataset (Ishii et al. 2005) and the absence of formal significance testing for the eigenvalues. Extending this work with multiple datasets or comparing to other climate indices (e.g., the PDO or NAO) could help verify robustness and clarify the physical interpretations of higher-order modes. In addition, incorporating atmospheric variables might reveal more comprehensive ocean–atmosphere coupling. Future research could also examine changing behaviors of these modes across different epochs (pre-2000 vs. post-2000) to explore the potential influence of climate change on the spatial and temporal signatures of global SST variability.

5. Conclusion

This analysis revealed that the first two EOFs dominate global SST variability, with the second mode strongly tied to ENSO, as confirmed by a high correlation with the Niño 3.4 index. By demonstrating how major climate patterns appear in the dominant modes, the lab successfully addressed its main research question regarding the identification and interpretation of leading global SST variability modes.

These findings emphasize the importance of the tropical Pacific in shaping global climate fluctuations, suggesting practical applications in seasonal-to-interannual forecasting and in improving climate model validations. More broadly, the results highlight how relatively few patterns can account for a large portion of SST variance, aiding climate diagnostics and teleconnection studies. Future work might expand the dataset range, assess statistical significance for each EOF, and incorporate additional variables (e.g., sea level pressure or wind stress) to clarify the coupled nature of ocean–atmosphere dynamics and to investigate how these modes evolve under changing climate conditions.

References

- Deser, C., Alexander, M. A., Xie, S.-P., & Phillips, A. S. (2010). Sea Surface Temperature Variability: Patterns and Mechanisms. *Annual Review of Marine Science*, *2*(1), 115–143. https://doi.org/10.1146/annurev-marine-120408-151453
- Golden Gate Weather Services. (n.d.). *Oceanic Niño Index (ONI)*. Retrieved February 5, 2025, from https://ggweather.com/enso/oni.htm
- Huang, P., & Xie, S.-P. (2015). Mechanisms of change in ENSO-induced tropical Pacific rainfall variability in a warming climate. Nature Geoscience, 8(12), 922–926. https://doi.org/10.1038/ngeo2571
- Ishii, M., Shouji, A., Sugimoto, S., & Matsumoto, T. (2005). Objective Analyses of Sea-Surface Temperature and Marine Meteorological Variables for the 20th Century using ICOADS and the Kobe Collection. International Journal of Climatology, 25, 865-879. https://doi.org/10.1002/joc.1169
- NOAA Physical Sciences Laboratory (PSL). (n.d.). Niño 3.4 Index Data. ERSST.v5. National Weather Service, National Centers for Environmental Prediction. Retrieved from https://psl.noaa.gov/data/correlation/nina34.data.
- Trenberth, K. E., & Hoar, T. J. (1997). El Niño and climate change. Geophysical Research Letters, 24(23), 3057–3060. https://doi.org/10.1029/97gl03092
- von Storch, H., & Zwiers, F. W. (1999). Statistical Analysis in Climate Research. Cambridge University Press, 484 p. https://doi.org/10.1017/CBO9780511612336
- Werb, B. E., & Rudnick, D. L. (2023). Remarkable Changes in the Dominant Modes of North Pacific Sea Surface Temperature. Geophysical Research Letters, 50(4). https://doi.org/10.1029/2022gl101078