

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Three-dimensional empirical orthogonal functions for big climate data: Computation, visualization and interpretation

Permalink

<https://escholarship.org/uc/item/68r0z6n9>

ISBN

9798270234348

Author

Lafarga, Danielle

Publication Date

2025-12-18

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Three-dimensional empirical orthogonal functions for big climate data: Computation,
visualization and interpretation

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Computational Science

by

Danielle Lafarga

Dissertation Committee:

Distinguished Professor Samuel S. P. Shen, Co-Chair

Distinguished Professor Efi Foufoula-Georgiou, Co-Chair

Professor Kim B. Oslen

Professor Alexander Ihler

Associate Professor Yanning Shen

2025

DEDICATION

To my partner
I could not have done this without you.

Para mi familia que quiero tanto
Plebes, ¿què mitote me cuentan?

To my friends who gave me countless ‘you can do this’ speeches,
and the ones that kept me company to make sure I did.

*The dreams of those who’ve fallen! The hopes of those who’ll follow! Those two sets of
dreams weave together into a double helix! Drilling a path towards tomorrow!*

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	xii
ACKNOWLEDGMENTS	xiii
VITA	xiv
ABSTRACT OF THE DISSERTATION	xvi
1 Introduction	1
1.1 Motivation	1
1.2 Current Status of 3D Analysis on Large Climate Data	2
1.3 Research Objectives of the Following Chapters	5
2 Feasibility of Computing 3D Empirical Orthogonal Functions from the NASA JPL Ocean General Circulation Model	8
2.1 Introduction	8
2.2 Data and Methods	11
2.2.1 Temperature Data from the NASA JPL OGCM	11
2.2.2 2D Formulation of the Space-Time Data Matrix	12
2.2.3 3D Formulation of the Space-Time Data Matrix	15
2.2.4 SVD and Temporal Covariance	15
2.3 Results	20
2.3.1 Eigenvalues and Variances Explained	20
2.3.2 Physical Interpretation of EOF 1: Equatorial Upwelling	22
2.3.3 EOF 2 and the Spatial Pattern of El Niño	28
2.3.4 PCs, Trend and Detrended	33
2.3.5 Compare the 3D EOF 1 to the 2D EOF 1	37
2.4 Conclusion and Discussion	38
3 Partition Input Method for Computing 3-Dimensional Empirical Orthogonal Functions for Big Space-Time Data: An Example of High-Resolution Ocean Assimilation Output	41
3.1 Introduction	41

3.2	Data and Method	44
3.2.1	Description of Big Data: Global Ocean Physics Reanalysis	44
3.2.2	Partition Input Method: Developing an Algorithm to Handle Big Data	45
3.3	Results	57
3.3.1	Creating 3D Visualization for EOFs	57
3.3.2	Global 3D EOFs Interpretation	63
3.3.3	2D Versus 3D EOF Comparison	69
3.4	Conclusion and Discussion	84
4	Four-dimensional North Pacific Variability and Regional Impacts from a High-Resolution Ocean Reanalysis	87
4.1	Introduction	87
4.2	Data and Methods	88
4.3	Results	90
4.3.1	EOF 1 and 2	90
4.3.2	E and C Indices	94
4.3.3	Regional Expressions in the California Current System	98
4.4	Conclusion and Discussion	105
5	Development of 3D Visualization Tools for Ocean Data	108
5.1	Introduction	108
5.2	Customize Colorbars for Optimal Display of Ocean Dynamic Features	111
5.3	Customizing XSLICE for GLORYS: Creating a Visualization Interface for Three-Dimensional Data	117
5.3.1	The Cross-Section Interfaces: Implementing XSLICE for Three Spatial Dimensions	123
5.3.2	Creation of the Interface for Comparison: Plotting Multiple Figures from 3D Datasets	141
5.4	Summary and Discussion	151
	Bibliography	155
	Appendix A Partition Input Method for Computing 3-Dimensional Empirical Orthogonal Functions for Big Space-Time Data: An Example of High-Resolution Ocean Assimilation Output	168
	Appendix B Four-dimensional North Pacific Variability and Regional Impacts from a High-Resolution Ocean Reanalysis	171
	Appendix C Development of 3D Visualization Tools for Ocean Data	173

LIST OF FIGURES

	Page
2.1 Ocean water temperature from the NASA JPL OGCM output for January 1950.	12
2.2 Climatology for January and August computed using the equation above from the NASA JPL OGCM data.	17
2.3 (a) The first eigenvalue λ_1 of the temporal covariance matrix for the weighted temperature anomalies for each month from January to December. Units of the first eigenvalue are the same as the covariance matrix in equation 2.13. (b) The scree plot for January based on data p_k and q_k . (b) The scree plot for August is based on data p_k and q_k	21
2.4 EOF 1 for January and August showing the equatorial upwelling. Depths are chosen to show that the strong upwelling occurs at 150 meters and that the upwelling stops at a deeper level around 400 meters.	23
2.5 A north-south cross-section of the ocean taken at $160^\circ E$ to show at what layers upwelling occurs. Panels (a) and (b) are the entire scope of all ocean layers. Panels (c) and (d) show the top layers up to 400 meters, as this will better show where upwelling starts and stops.	24
2.6 NASA JPL OGCM temperature for January and August at 200 meters showing equatorial upwelling. Years are chosen to reflect time with strong upwelling.	26
2.7 Cross-sectional map based on the zonal mean from 0 to 360° longitude degrees to show ocean ventilation at the high latitude regions.	28
2.8 The second eigenvalue λ_2 of the temporal covariance matrix for the weighted temperature anomalies. Units of the first eigenvalue are the same as the covariance matrix in equation 2.13. January has the largest value, while August has the smallest value.	29
2.9 EOF 2 for January and August showing the ENSO pattern. Depths are chosen to show the ENSO regions in different depths from surface to 100 meters. The ENSO pattern dies out approximately at levels deeper than 150 meters.	30
2.10 Cross-section at latitude of $10^\circ S$ that shows the separation of the warm and cold anomaly regions in the Pacific.	31
2.11 A zonal cross-sectional map based on the average from latitude $10^\circ N$ to $10^\circ S$. The average in the figure are to better show a robust 3D ENSO pattern with not only the surface warm region but also a depth structure.	32
2.12 Principal components PC 1, PC 2, and PC 3 of January and August from the weighted anomalies of the NASA JPL OGCM output from 1950 to 2003.	34

2.13	The first eigenvalue $\lambda_1^{(d)}$ for each month computed from the detrended anomalies.	36
2.14	Principal components of January and August based on the detrended anomalies.	37
2.15	2D EOF 1 for the top layer based on the area-weighted anomalies from the NASA JPL OGCM. The ENSO signal is shown in this EOF 1 for the surface layer. The 2D EOFs are computed layer by layer.	38
3.1	The run time for the PIM covariance matrix calculation for different total RAM gigabytes.	54
3.2	The run time for the PIM EOF calculation for different total RAM gigabytes.	57
3.3	EOF 1 plotted for the entire globe at three depth layers (surface, 77 meters, and 222 meters). These depths were chosen to show the dynamics of El Niño's zonal dipole, and in accordance with GLORYS depths (see Table 3.1).	59
3.4	Three-dimensional visualization of EOF 1 with cross sections taken at the equator to 60°N and 160°E-60°W. The depth domain is limited to the upper 400 meters as it provides a clear display of the ENSO dynamical pattern. The right-side view cross-section shows the strong ENSO zonal dipole. The front meridional view displays the Northern Hemisphere extent of the western subsurface cold pole in the 3D ENSO signature. The top view shows ENSO's surface pattern, commonly used in the climate research community.	61
3.5	Vertical cross-sectional view of EOF 1 at 10°S, 5°S, 0°, 5°N, and 10°N in the Pacific from 160°E - 80°W. Depths go from the surface to 454 meters. These cross-sections also show the signal of ENSO's zonal dipole pattern.	62
3.6	Visualization of the first three EOFs and their PCs: 3D EOF 1 and PC 1 (left column), 3D EOF 2 and PC 2 (middle column), and 3D EOF 3 and PC 3 (right column) for the global ocean. EOFs were chosen to be displayed at four different depth layers (surface, 56 meters, 156 meters, 222 meters). These depths were chosen for EOF 1 to show the traditional El Niño, EOF 2 to show warming and central Pacific El Niño, and EOF 3 to show the meridional modes.	64
3.7	Depth-zonal view of the first three EOFs over the tropical region. The values displayed in color are taken as the average of 3D EOFs from 5°S to 5°N. The three panels in the top row, i.e., (a), (b), and (c), are from the surface to 1,941 meters, showing that strong temperature variability occurs in the top 380 meters. The three panels in the bottom row, i.e., (d), (e), and (f), are zoomed in to the layers in the top 380 meters, indicated by the black dotted line in the top row. Black contours are isotherms from boreal winter (DJF) climatology. The blue contour is the 20°C isotherm to represent the thermocline. Variability in the depths between 2,000 - 5,500 meters is very small and not displayed in this figure.	66
3.8	EOF 3 visualized for the SPM at the surface and with a zonal-depth cut at 30°S and meridional cut at 105°W. Cross-sections were chosen to bisect the warming from the SPM, and are from surface to 65 meters to interpret the subsurface structure of SPM. Grey lines in the zonal cut are at every 10 meters to align with the depth axis.	68
3.9	Same as Fig. 3.6 but for the 2D calculation. We exclude the time series as each layer outputs multiple PCs.	70

3.10	Same as Fig. 3.7 , but with the 2D calculation results.	71
3.11	Correlations at every depth between meridional modes and (a) 2D PC 2 and (b) 2D PC 3. Figure (a) displays SPMM and NPMM correlations, omitting AMM due to statistical insignificance. Figure (b) shows only SPMM for the same reason. Depths without statistically significant correlations are also omitted.	72
3.12	Correlation at every depth between 2D PC 2 and CP El Niño indices. EMI correlations were not statistically significant and are not shown. Depths without statistically significant correlations are also omitted.	73
3.13	Correlations between 3D and 2D PCs. Panel (a) displays 2D PC 2 and correlations with all 3D PCs. Panel (b) same as (a) but for 2D PC 3. Results between 2D PC 3 and 3D PC 2 are not shown due to statistical insignificance. Depths without statistically significant correlations are also omitted.	74
3.14	2D EOF 2 visualized at three depths: 92, 266, and 454 meters. Depths 92 and 266 are displayed to interpret the global warming signal throughout the layers, with the 266 meter layer also displaying a weak dipole signal in the tropical Pacific. The 454 meter depth was chosen to show a globally weak signal with no identifiable climate modes, providing context for the final correlation depth between 2D PC 2 and 3D PCs 1 and 2.	75
3.15	Correlation between two CP El Niño indices (TNI and EMI) and 2D PC 3. Panel (a) displays all statistically significant correlations. Panel (b) shows results from surface to 266 meters (red box in (a)). The first 266 meters are chosen to interpret correlations for depths where CP El Niño and the PMM experience the most variability.	78
3.16	2D EOF 3 displayed at three depths: 110, 266, and 454 meters. The 110 meter layer is displayed to interpret the PMM and CP El Niño signal. The depth 266 meters is displayed to explain the negative EMI and 3D PC 3 correlation, and 454 meters is displayed to examine the spatial pattern associated with a positive correlation between 3D PC 1 and 2D PC 3.	79
3.17	All correlation results for 2D PC 3 in the top 500 meters. AMM, 3D PC1, and 3D PC 3 show no significant correlations above 500 meters and are not displayed in this figure.	80
3.18	Visualization of 2D PC/EOF 3 and 3D PC/EOF 3 at 1,452 meters. PCs and EOFs are displayed to compare spatial patterns and time series of the depth layer with minimum correlation between 2D PC 3 and 3D PC 3 as defined by Fig. 3.13(b). The top row shows (a) 2D PC 3 and 3D PC 3. The bottom row shows the corresponding EOFs: (c) 2D EOF 3 and (d) 3D EOF 3. EOF plots use different colorbar ranges to effectively visualize cold/warm signals for comparison.	81
3.19	Same as Fig. 3.18 but for 266 meters.	82

4.1	Visualization of the first two EOFs and their PCs: 3D EOF 1 and PC 1 (left column) and 3D EOF 2 and PC 2 (right column) for the North Pacific Ocean. EOFs are displayed at four depth layers (surface, 56 meters, 156 meters, 222 meters). These depths were chosen for EOF 1 to show ENSO, and EOF 2 to show the PMM. The depth values are in accordance with depth layers defined by GLORYS.	91
4.2	Depth-zonal view of the first two EOFs over the tropical region. The values displayed in color are taken as the average of 3D EOFs from 5°S to 5°N. Both panels are from surface to 318 meters; the weak variability after 318 is not shown. Black contours are isotherms from boreal winter (DJF) climatology. The blue contour is the 20°C isotherm to represent the thermocline.	92
4.3	Same as Fig. 4.1 but for the E and C index.	95
4.4	Same as Fig. 4.2, but for the E and C index.	96
4.5	The (left column) E and (right column) C index for the U.S West Coast from surface to 222 meters. Layers are specified according to GLORYS depth layers. These depths, longitudes, and latitudes were chosen for the interpretation of EP (E index) and CP (C index) El Niño and their effect on the CCS and USWC.	99
4.5	The (left column) E and (right column) C index for the U.S West Coast from surface to 222 meters (continued).	100
4.6	Zonal cross-section cuts from the USWC of the (a) E and (b) C index. Cuts were taken from 34°N to 46°N in increments of 3° to capture the effects of both El Niño types up the coast. Longitudes were chosen to match Fig. 4.5.	102
4.7	The same Zonal cross-section cuts from the USWC of the (left column) E and (right column) C index as in Fig. 4.6, visualized in 2D for clarity. The left column is the E index, and the right column is the C index. Cuts were taken from 34°N to 46°N in increments of 3° to capture the effects of both El Niño up the coast. Panels were ordered top to bottom from north to south. Longitudes were chosen to match Fig. 4.5. Black contours in are the GLORYS boreal winter climatology	103
5.1	The two colormaps used for the custom colorbar in Chapter 2. The winter colormap is used for all negative EOF values, and the hot colormap is for all positive EOF values. The ‘under’ and ‘over’ boxes specify the colors of any values out of bounds in the colorbar, and ‘bad’ colors are for any NaN values.	113
5.2	The result of combining two colormaps to create a custom colormap. The Python code demonstrates conversion of colormap values into color arrays, vertical stacking of the arrays, and conversion back into a colormap object.	115
5.3	The colorbar implementing the colormap created in Fig. 5.2 in the range $[-.3, .3]$. The Python code generates random data, creates the figure for the visualization, applies the colormap to visualize data in a contour plot, and then applies the colormap to the colorbar.	117

5.4	(a) Application of the colorbar from Fig. 5.3 with increased color values. (b) Updated colorbar featuring new cool tones and central transition colors. EOF 3 is displayed at 186 meters for both examples to compare the visual representation of weak cooling for both colorbars. The weak cooling appears amplified in (a) due to the saturated green tones, while the same cooling in (b) exhibits reduced saturation, better representing the true value.	118
5.5	Zonal cross-section for the GODAS potential temperature at $\sim 4^\circ\text{S}$. Time was chosen as December 1997 to show temperatures for a strong El Niño.	119
5.6	Main tab for XSLICE and the cross-section menu. Each cut tab is named after the variable held fixed for the cross-section. Depth is a longitude-latitude (xy) cut, Longitude is a meridional (yz) cut, and Latitude is a zonal (xz) cut. A complete tutorial on each button can be found in [1].	120
5.7	Additional color and display features in XSLICE that can change the figure's aesthetic for export.	121
5.8	Initial interface for the depth cross-section of XSLICE for GLORYS. Above the visualization are the interactive widgets that will modify the figure. In its initial state, the interface shows the EOF 1 for the entire latitude-longitude domain at the surface. The signal captured in EOF 1 is the traditional El Niño pictured as the eastern Pacific warming.	124
5.9	The depth interface after user interaction with the depth slider. This depth shows a distinct change compared to the surface in the western Pacific, due to the emergence of the subsurface zonal dipole component of El Niño. . . .	126
5.10	Same as Fig. 5.9 with a larger colorbar range, demonstrating how signals can be amplified or weakened depending on the clip values.	127
5.11	Surface view of EOF 1 cropped to capture El Niño's eastern Pacific warming. The latitude range of 20°S - 55°N and eastern longitude boundary at 70°W were selected to follow [2]. Western longitude was chosen to align with the westernmost extent of El Niño's eastern tropical Pacific warming. Figures differ in longitude increments, where (a) is the typical increment of 1° , and (b) uses an increment of $1/12^\circ$ that matches the data's spatial resolution. . .	129
5.12	Surface view of EOF 1 cropped to show the U.S. West Coast for interpreting El Niño influence on the California Current System.	130
5.13	Same as Fig. 5.12, but for a Longitude slider that increments by $1/12^\circ$. Step size is changed to identify precise coordinates for the permanent meander at 40°N and 127°W [3].	131
5.14	Initial interface for the latitude cross-section of XSLICE for GLORYS. In its initial state, the interface shows EOF 1 for the entire longitude-depth domain at the equator. The equatorial cross-section displays the subsurface zonal dipole signal of El Niño as warming in the east and cooling in the west. . . .	132
5.15	Same as Fig. 5.14 but limited to the first 380 meters to better analyze and quantify the El Niño dipole.	134
5.16	Same as Fig. 5.15 with a larger clip value to identify the areas with stronger variability in the El Niño dipole.	135

5.17	The equatorial cross-section of EOF 3 for the first 380 meters. Clip value is set to 0.00014 to display the maximum warming in the thermocline caused by the trade wind charging mechanism of the North Pacific meridional mode.	136
5.18	Two zonal cross-sections taken at (a) 40 °N, and (b) 37°N to examine the progressive narrowing of El Niño’s warming toward the California coast.	137
5.19	Multiple zonal cross-sections visualized along the U.S. West Coast from 34° - 46°N. Each cut is chosen to illustrate the changing warming pattern along the coast to interpret El Niño’s effects on the CCS. These cross-sections include the cuts from Fig. 5.18 to demonstrate the application of the resulting cross-sections in a 3D figure.	138
5.20	Initial interface for the longitude cross-section of XSLICE for GLORYS. In its initial state, the interface shows EOF 1 for the entire latitude-depth domain at the date line. The cross-section displays the subsurface dipole signal of El Niño around the equator with warming in shallower waters than the cooling.	140
5.21	Multiple meridional cross-sections taken from 140°E - 100°W in 40° increments to display the El Niño dipole across the Pacific. Slices were chosen using the Longitude interface and are displayed in a 3D figure for easier interpretation.	141
5.22	Compare interface for the depth cross-section. The initial dimension and clip values are the same as the Depth interface of Fig. 5.8. The EOF tags below the longitude range slider specify the EOFs that are visualized, and the order they are displayed. The top two figures are 3D EOF 1 and 2, and the bottom left figure is 3D EOF 3.	142
5.23	The Compare interface after removing 3D EOF 1 and 3D EOF 3 from the EOF tags, leaving only 3D EOF 2 to be displayed. A text box is displayed next to the 3D EOF 2 tag, ready for user input. A drop-down menu is displayed, providing the EOF tags that can be added for visualization.	143
5.24	A 6-figure layout for the Compare interface. Figure placement follows tag order: 1, 2, 3, 4, 5, and 6. Figures are placed accordingly in three rows and two columns to maximize space. Layouts with 3 - 4 figures follow the same configuration but omit the bottom row.	144
5.25	Two-figure layout displaying (top) 3D EOF 1 and (bottom) 3D EOF 2 at the surface. The figures are stacked vertically to maximize space for comparison. The colorbar is positioned at the bottom for the same reason.	146
5.26	The first three 3D EOFs at 77 meters cropped to focus on the entire Pacific. The top row displays (left column) 3D EOF 2 and (right column) 3D EOF 3. Bottom row is 3D EOF 1. 3D EOF 2 and 3 are placed in the top row to explain the differences and similarities between the tropical Pacific warming from CP El Niño and the NPM, respectively. 3D EOF 1 is placed in the bottom row for separate analysis of El Niño.	148
5.27	The Compare interface during the process of comparing 3D EOFs and 2D EOFs to correct for consistent sign representation in 2D EOFs. The first three (left column) 3D EOFs and (right column) 2D EOFs are shown for the global domain. The EOFs are illustrated at 1 meter, displaying a layer with incorrect signs in 2D EOF 2 and 2D EOF 3.	149

5.28	3D EOF 3, 2D EOF 2 and 2D EOF visualized at 222 meters for comparison. EOFs are displayed at this depth to identify the NPMM in 3D EOF 3 and the split representation in 2D EOF 2 and 2D EOF 3.	151
------	---	-----

LIST OF TABLES

	Page
2.1 Depths(m) of 32 layers of the NASA JPL OGCM	12
3.1 The depth values of the 50 layers in the GLORYS model (unit: meters). Values were rounded to the first two decimal places.	44
3.2 Run time for various chunk sizes in years, total GB for the PIM computation of the covariance matrix.	53
3.3 Run time for the PIM computation of EOFs versus their total GB.	57
3.4 Meridional mode region definitions to compute their indices. Regions are defined by [4].	68
5.1 Summary of published visualization tools for climate data. Applications ap- pear in the order they are referenced in this Chapter.	112

ACKNOWLEDGMENTS

I would like to thank my PhD advisor, Samuel Shen, for his support, patience, and unwavering belief in me. Thank you also to my co-advisor Efi Foufoula-Georgiou for her insight and kindness, and my doctoral committee Kim Olsen, Alexander Ihler, and Yanning Shen, for their guidance. I would also like to thank my NOAA collaborators Mike Jacox, Mike Alexander, Tom Smith, and my NASA collaborator Tony Song. I am grateful I had such kind and knowledgeable people to guide me. A thank you to the SCIL members Thomas Bui, Max Ramirez, and David Hyon for aiding in this research. A very big thank you to SCIL member Gian Tapanan and my mother, Maria, who helped me write and proofread this dissertation. Lastly, a thank you to my friends who believed in me and my partner, who went above and beyond to support me.

Chapter 2 of this dissertation is a reprint of the material as it appears in A Feasibility Study of Three-Dimensional Empirical Orthogonal Functions From the NASA JPL Ocean General Circulation Model: Computing, Visualization and Interpretation, used with permission from Tellus A: Dynamic Meteorology and Oceanography.

This study is supported by the National Oceanic and Atmospheric Administration (NOAA) – Cooperative Science Center (CSC) for Earth System Sciences and Remote Sensing Technologies (CESSRST-II) under the Cooperative Agreement Award #**NA16SEC4810008**. I would like to thank the NOAA Office of Education, the Educational Partnership Program with Minority Serving Institutions (NOAA-EPP/MSI) and the NOAA-CESSRST-II for providing me with full fellowship support. The statements, findings, conclusions and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA.

This research was also financially supported in part by the U.S. National Science Foundation grant (Award No. IIS2324008) and the U.S. NASA Jet Propulsion Lab’s CAP program.

VITA

Danielle Lafarga

EDUCATION

Doctor of Philosophy in Computational Science **2020-2025**
San Diego State University and University of California Irvine *San Diego and Irvine, CA*

Bachelor of Science in Physics **2018**
University of California San Diego *San Diego, CA*

RESEARCH EXPERIENCE

Graduate Research Assistant **2020–Present**
San Diego State University Climate Informatics Laboratory *San Diego, California*

NOAA CESSRST-II Fellow **2020–Present**
NOAA CESSRST *San Diego, California*

REFEREED JOURNAL PUBLICATIONS

D. Lafarga, T. Bui, Y. T. Song, T. M. Smith, and S. S. P. Shen, "A feasibility study of three-dimensional empirical orthogonal functions from the NASA JPL ocean general circulation model: Computing, visualization and interpretation," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 75, no. 1, pp. 213–230, 2023. [Online]. Available: <https://doi.org/10.16993/tellusa.3223>

T. Bui, **D. Lafarga**, T. Smith, Y. Song, and S. S. P. Shen, "Calculation, visualization, and interpretation of three-dimensional atmosphere-ocean coupled empirical orthogonal functions using the reanalyses data," *Theoretical and Applied Climatology*, vol. 154, pp. 1–15, 2023. [Online]. Available: <https://doi.org/10.1007/s00704-023-04513-1>

T. Bui, M. Ramirez, **D. Lafarga**, Y. T. Song, E. Foufoula-Georgiou, and S. S. P. Shen, "XSLICE: An efficient visualization and diagnosis tool for 4-dimensional climate dynamics," *Journal of Atmospheric and Oceanic Technology*, vol. 42, no. 9, pp. 1185–1199, 2025. [Online]. Available: <https://journals.ametsoc.org/view/journals/atot/aop/JTECH-D-24-0113.1/JTECH-D-24-0113.1.xml>

D. Lafarga, G. Tapanan, and S. S. P. Shen, "Partition input method for computing 3-dimensional empirical orthogonal functions for big space-time data: an example of a high-resolution ocean assimilation output," *Journal of Computational Physics*, submitted October 2025.

D. Lafarga, M.G. Jacox, M.A. Alexander, and S. S. P. Shen, "Interpretation of 4-dimensional space time north Pacific ocean dynamics from high-resolution ocean assimilation data," *Journal of Geophysical Research Oceans*, to be submitted December 2025.

D. W. Hyon, **D. Lafarga**, and S. S. P. Shen, "Realism of 3D ocean dynamics represented by 2D empirical orthogonal functions: 3D versus 2D comparison and interpretation using Global Ocean Physics Reanalysis data," *Journal of Geophysical Research Oceans*, submitted October 2025.

M. Ramirez, **D. Lafarga**, and S. S. P. Shen, "Quantify 4-dimensional variability of El Niño–Southern Oscillation using the zonal thermocline slope over the equatorial Pacific," *Journal of Climate*, submitted October 2025.

SOFTWARE PRODUCTS

3D EOF from OGCM <https://doi.org/10.5281/zenodo.7042027>
Python algorithm that computes 3D EOFs from NASA JPL OGCM

PIM for computing 3D EOFs <https://doi.org/10.5281/zenodo.17451617>
Python algorithm that computes 3D EOFs on high-resolution Global Ocean Physics Reanalysis data using the partition input method.

ABSTRACT OF THE DISSERTATION

Three-dimensional empirical orthogonal functions for big climate data: Computation,
visualization and interpretation

By

Danielle Lafarga

Doctor of Philosophy in Computational Science

University of California, Irvine, 2025

Distinguished Professor Samuel S. P. Shen, Co-Chair

Distinguished Professor Efi Foufoula-Georgiou, Co-Chair

Understanding our ocean requires analyzing it as it naturally exists—in three dimensions. Empirical orthogonal functions (EOFs) are often used to analyze dominant patterns of climate variability. They have been used extensively to discover, quantify, and predict important climate phenomena such as El Niño-Southern Oscillation, the North Atlantic Oscillation, and the Pacific Decadal Oscillation. However, computing three-dimensional (3D) EOFs for large datasets has been computationally prohibitive due to computer memory limitations when implementing traditional methods (i.e., singular value decomposition) to obtain EOFs. These limitations have restricted EOF calculations to a single layer, a small spatial domain, or coarse resolution. This dissertation presents two approaches that enable 3D EOF computation for large climate datasets using a regular personal computer with limited memory (e.g., 16 GB RAM). The first is the temporal covariance method, and the second is the partition input method (PIM). The temporal covariance approach results in an eigenvalue problem for a smaller matrix compared to the traditional spatial covariance matrix method. This method is applied to NASA JPL ocean general circulation model data for the global domain (1° latitude-longitude resolution, 447 MB per month). PIM strategically partitions a large space-time data matrix into manageable sub-matrices while preserving 3D covariances,

and is applied to calculate 3D EOFs for the Global Ocean Physics Reanalysis (GLORYS) data ($1/12^\circ$ latitude-longitude resolution, 50 depth layers). We computed the 3D EOFs for GLORYS temperature data in two spatial domains: global (49 GB total) and North Pacific (8 GB). The resulting 3D EOFs reveal the 3D structure of important climate phenomena, including El Niño-Southern Oscillation, Pacific meridional modes, central Pacific El Niño, marine heat waves, and warming. Our North Pacific analysis provides a detailed quantification of the regional effects of eastern and central Pacific El Niño types along the U.S. West Coast. Our results are relevant to understanding impacts on the California Current System. The last part of this dissertation is on data visualization, where we customized the tool Cross-Sectional Layers with Interactive Cropping and Exportation (XSLICE) for GLORYS data. The customized tool facilitates interactive examination of depth, zonal, and meridional cross-sections of the 3D ocean. Because of its distinguishing feature of fast visualization, we routinely used XSLICE for GLORYS throughout this dissertation research to gain insights into 3D ocean dynamics, before performing detailed quantitative analysis and production of publication figures. In summary, our work has demonstrated the feasibility of solving the singular value decomposition problem for a large matrix often encountered in modern engineering and science applications. Without methods like PIM, many 3D EOF computations would be limited to reduced domains or coarse resolutions. Therefore, the numerical methods for 3D EOF computing and the data visualization tools introduced in this dissertation will help explain detailed climate dynamics.

Chapter 1

Introduction

1.1 Motivation

In climate science, principal component analysis (PCA) or empirical orthogonal functions (EOFs) are used to analyze spatial and temporal variability. Researchers have implemented these techniques to discover, quantify, and predict phenomena such as the El Niño-Southern Oscillation (ENSO) (see Table 1 in [5]), Central Pacific (CP) El Niño [6–8], Pacific and Atlantic meridional modes [9,10], and marine heat waves [11]. These studies focused on climate variability of sea surface temperature anomalies (SSTA) and excluded subsurface ocean temperature anomalies (SOTA). As a result, these surface analyses cannot capture subsurface variability and cross-layer covariances, limiting their representation of ocean dynamics.

To better represent the natural structure of ocean variability, researchers have expanded their analyses to include SOTA through equatorial cross-sections or lagged correlations with principal components from SSTA. This approach has advanced our understanding of ENSO [12–14], the California Current System (CCS) [15], the Atlantic and Pacific meridional modes [10,16,17], and the Indian Ocean dipole [18–20]. These studies provide vital information on

subsurface dynamics, underlying mechanisms, and teleconnections of these phenomena. The value of subsurface analysis is clear. Therefore, recent calculations [21–24] extended these approaches to compute EOFs across latitude, longitude, and depth simultaneously (i.e., 3D EOFs), providing a more comprehensive representation of the three-dimensional structure of ocean variability.

Three-dimensional analyses offer significant advantages, yet traditional singular value decomposition algorithms that solve for 3D EOFs are not always well adapted for large climate data. These constraints have restricted EOF analysis to small datasets by reducing calculations to a single layer, a small spatial domain, or coarse resolution. This study presents a partitioning method for calculating 3D EOFs from large climate datasets on computers with limited memory. By overcoming these computational memory barriers, our method makes 3D analysis accessible to the climate community.

1.2 Current Status of 3D Analysis on Large Climate Data

EOFs in climate science were traditionally computed from the eigenvectors of a spatial covariance matrix of climate anomaly data, as demonstrated in the original EOF paper [25] and the milestone papers [26, 27]. For a $N \times Y$ space-time climate anomaly matrix, A , the spatial covariance matrix, C_S , and its eigenvalue problem are as follows:

$$\begin{aligned} C_S &= AA^T \\ C_S u_k &= \lambda_k u_k, \end{aligned} \tag{1.1}$$

where u_k are the EOFs, and λ_k are the eigenvalues. A modern way of computing EOFs is via singular value decomposition (SVD) [28, 29]:

$$A = UDV^T. \quad (1.2)$$

In SVD, $U_{N \times Y}$ is the EOF matrix, $D_{Y \times Y}$ is a diagonal matrix containing the singular values, and $V_{Y \times Y}$ is the principal component (PC) matrix.

Studies have widely implemented spatial covariance and SVD methods on SSTA to calculate EOFs and PCs [6–10, 30–32]. To explore variability for multiple altitude layers, these methods were used on equatorial cross-sections [14], or layer-by-layer calculations [12, 33]. We refer to these methods as two-dimensional (2D) because each calculation applies to a single depth or cross-section. These 2D studies produced meaningful physical interpretations of leading EOFs and PCs that advanced our understanding of the distinct mechanisms, teleconnections, and dynamics of important climate phenomena. However, the 2D EOF approach fails to capture cross-layer covariances, which provide a more realistic representation of the ocean. Thus, three-dimensional (3D) EOF calculations, which calculate EOFs for all spatial dimensions simultaneously, are necessary for explaining 3D ocean dynamics.

Although extensive published research exists on 2D EOFs, 3D applications have received less attention in climate science. Computing 3D EOFs for large datasets using the traditional algorithms for SVD or spatial covariance is computationally prohibitive due to computer memory limitations. Either case requires:

- (i) Loading all data layers of the climate anomaly matrix simultaneously into computer memory for the calculation.
- (ii) Storing the resulting spatial covariance matrix or SVD matrices.

For requirement (ii), given an $N \times Y$ space-time climate anomaly matrix A , the spatial covariance matrix dimensions are $N \times N$, while SVD produces matrices $U_{N \times Y}$, $D_{Y \times Y}$, and $V_{Y \times Y}$. When the dataset contains many spatial observations (i.e., high-resolution data), (ii) becomes impossible. When the dataset size exceeds computer memory, (i) and (ii) become impossible. Given these memory requirements, 2D EOF calculations are favorable as they require much less memory than 3D calculations. Previous 3D EOF analyses circumvented these memory barriers by using small datasets through smaller spatial domains [22] or coarse resolutions [23]. This approach enables 3D EOF computing, but does not resolve computing 3D EOFs for large datasets.

A modern option is to use machine learning alternatives for EOF calculations. Autoencoders can extract essential information from a large dataset and represent it as latent spaces [34]. However, the resulting latent spaces are neither orthogonal nor naturally ordered by variance. Therefore, autoencoders need additional constraints to generate outputs as physically meaningful as EOFs. Similarly, tensor decomposition can compute latent spaces on big data but faces the same interpretability issues as autoencoders. While both methods are feasible for climate data [35–37], they cannot substitute traditional 3D EOF analysis. Our approach draws inspiration from partition methods in tensor decomposition [38] to overcome the large data memory barrier while preserving traditional EOF analysis to maintain clear physical interpretations of spatial patterns and variability.

Another approach for the big data EOF analysis uses stochastic algorithms for SVD. These algorithms iteratively calculate and update PCs by computing gradients from randomly sampled portions of the data. Since they process only data samples rather than the full dataset, these methods avoid the memory requirements in (i). Stochastic methods typically require multiple passes over the data, but recent studies [39–41] have investigated optimal gradients to accelerate convergence to the PCs, thereby reducing the number of required passes and computational time. Additionally, when only k PCs/EOFs are needed where

$k \ll Y$, these methods become even faster, and the resulting EOF and PC matrices are much smaller than those in (ii). Although stochastic algorithms offer a solution for the big data SVD, our research prefers calculating all PCs/EOFs and requires interpreting the exact PCs/EOFs from our climate data. Therefore, we use data partitioning (through chunking or tiling) and temporal covariance for our 3D EOF analysis instead of these stochastic methods.

Principal components can be computed from eigenvectors of a temporal covariance matrix (Eq. (1.3)). This method has been used in climate science since 1967 [42] and produces results identical to those from classical SVD or spatial covariance methods. The calculation of temporal covariance requires only a matrix multiplication, which allows for partitioning of the climate anomaly matrix to avoid loading the full matrix into memory, resolving constraint (i). The resulting temporal covariance matrix is based on time dimensions ($Y \times Y$) and is typically much smaller than space-based matrices, addressing the constraint from requirement (ii). Using this simple approach, this dissertation performs a 3D EOF analysis on extremely large climate data using a personal computer with limited memory.

$$\begin{aligned} C_T &= A^T A \\ C_T v_k &= \lambda_k v_k \end{aligned} \tag{1.3}$$

1.3 Research Objectives of the Following Chapters

Building upon previous subsurface analysis and recent 3D EOF advances, this dissertation demonstrates the value and feasibility of computing 3D EOFs on large ocean datasets. Our approach is a computationally accessible 3D EOF analysis that enhances our understanding of subsurface dynamics through the visualization and interpretation of temperature variabil-

ity throughout the water column. Following the introduction are four chapters. Chapter 2 demonstrates the feasibility of computing 3D EOFs using the temporal covariance approach on coarse resolution data. Chapter 3 describes the methods for computing and visualizing 3D EOFs on large data. The Chapter 4 applies the methods from Chapter 3 to the Pacific. Chapter 5 details the custom visualization tools used throughout this dissertation for figure creation and 3D ocean analysis.

Chapter 2 demonstrates the computational feasibility of 3D EOFs using NASA JPL ocean general circulation model data (1° resolution, 32 depth layers, January 1950 – December 2003). The purposes of this chapter are to: (a) establish a computational algorithm for 3D EOFs from temporal covariance, and (b) interpret climate dynamics of the leading EOF modes. The temporal covariance method used in this chapter is validated against SVD results and serves as part of the partition input method algorithm for high-resolution data.

Chapter 3 presents the partition input method (PIM) for computing temporal covariance and 3D EOFs. We apply this method to the high-resolution global ocean physics reanalysis (GLORYS) data ($1/12^\circ$ latitude-longitude resolution, 50 depth layers) for the global domain, using boreal winter temperatures from 1993/1994 to 2020/2021. The successful application of PIM enables 3D visualization and interpretation of leading variability modes.

Chapter 4 applies the PIM from the previous chapter to the high-resolution data (GLORYS) for the North Pacific. This chapter aims to quantitatively describe the 3D structure of ocean variability for ENSO, the North Pacific meridional mode, Eastern Pacific El Niño, and Central Pacific El Niño. High-resolution data enables detailed analysis of regional effects from both El Niño types, particularly along the U.S. West Coast and within the California Current System.

Chapter 5 describes the motivation, development process, functions, and applications of the visualization tool created specifically for this research. This tool, called XSLICE for

GLORYS, has three key features: (i) the ability to visually scan depth, zonal, and meridional cross-sections, (ii) manipulation of a custom colorbar to effectively represent and enhance climate patterns, and (iii) a user-friendly interface developed with a simple and adaptable Python code for ease of use throughout this study. The tool was employed on the 2D and 3D EOFs calculated from the GLORYS data in Chapter 3. The successful application of XSLICE for GLORYS proves that the tool can serve as a simple data-adaptable visualization tool for ocean dynamic detection and analysis.

Chapter 2

Feasibility of Computing 3D Empirical Orthogonal Functions from the NASA JPL Ocean General Circulation Model

2.1 Introduction

Empirical orthogonal functions (EOFs), or spatial modes, are often used to analyze general climate model (GCM) output to explore spatial patterns of climate variability or for reconstructing historical climate data [43,44]. EOFs show which spatial patterns explain the most variance in the data, and can consequently give insight into how our data behaves. Although GCMs and climate dynamics are 3-dimensional in space and 1-dimensional in time, EOF analyses have often been performed in 2D layers in space [33,45,46]. The primary purposes of this chapter are twofold: (a) to show a computational algorithm of 3D EOFs, and (b) to interpret the climate dynamics of the first two EOFs. We do so by using the output data of a NASA JPL ocean general circulation model (OGCM).

Numerically, EOFs can be computed by finding the eigenvectors of the spatial covariance matrix from the weighted anomalies. This spatial covariance method often has convenient physical interpretations for leading modes. Most EOF calculations in recent years use singular value decomposition (SVD) with packages available in different computer languages, e.g., Python, R, and MATLAB, when the size of the SVD matrix is not too big, say less than half of your computer memory [47]. Our 1-degree spatial resolution dataset is not very big, and SVD computing can be done. However, we also use temporal covariance, which partially addresses the issue of memory limitations for future datasets of higher resolutions. This memory limitation is perhaps one of many reasons that EOFs were used to be computed layer-by-layer. Our method, detailed in this chapter, is presented in terms of covariance.

Papers such as [33], [45], and [46] calculated 2D EOFs, and indicated that the strongest signal was El Niño-Southern Oscillation (ENSO). ENSO is an ocean-atmosphere climate phenomenon characterized by the warming of sea surface temperatures in the central and/or east-central equatorial Pacific [48]. The pattern is often described as a warm tongue in the eastern Pacific. Despite its aforementioned clear spatial pattern in sea surface temperature (SST), ENSO is a phenomenon involving not only the surface but also the deep ocean, up to 150 meters, to be discussed later in this paper. Therefore, 2D EOFs lack the ability to describe the interactions between surface and deeper ocean layers. The need to consider such significant deep ocean dynamics motivated the idea of computing EOFs for all layers at once. The resulting 3D EOFs show that ENSO accounts for a smaller fraction of variability than equatorial upwelling in the OGCM data. This validates the significance of calculating 3D EOFs.

This chapter will demonstrate 3D EOF calculations based on the OGCM from surface to 5,500 meters depth, with 32 depth layers, 1-degree latitude and longitude spatial resolution, and monthly temporal resolution. Our mode visualization and physical interpretation will focus on the first two EOFs. Our 3D EOF results demonstrate that (i) the 3D spatial

pattern of equatorial ocean upwelling is mainly reflected in the first EOF mode and has its most variabilities within the depth layer between 100 and 400 meters, (ii) the 3D El Niño-Southern Oscillation (ENSO) dynamic pattern is mainly reflected in the second EOF mode and is mostly confined from surface to the depth of 150 meters, and (iii) the lead eigenvalue from the 3D EOF calculation appears to contain some signal of oceanic warming.

Relevant literature already exists on (i) multiway table data analysis theory, e.g, Structuration des Tableaux A Trois Indices de la Statistique (STATIS) [49–51], and (ii) computational algorithms for big data EOFs [52]. A special case of a multiway table is a 3-way table that includes locations, parameters, and time [53]. In our case, we only look at one parameter: temperature; our locations are three-dimensional; and our time is indexed for monthly data from the NASA JPL OGCM.

In principle, multiway data analysis algorithms such as STATIS should be able to deliver the results of PCs and EOFs when the multiway data table is properly non-dimensionalized. In practice, the problem is much more complex, such as the large size of the multiway data table, and the physical meanings of the PCs and EOFs. For example, [52] contributed to the development of innovative computational algorithms to handle multi-terabytes of data for 3D EOF computing. Their paper considered the cases: (a) Climate Forecast System Reanalysis(CFSR) data with a 1/2-degree, 41 layers, and 6-hour resolution; (b) Community Atmosphere Model 5.0 (CAM5) data with 1/4-degree, 30 layers, and 6-hour resolution. Their CFSR dataset has 2.2 TB, and their CAM5 dataset has 16 TB. They concluded that most of the run time was spent on In/Out process. Their EOFs from their multi-hour data by default have no clear physical interpretations. Accordingly, their paper included neither EOF figures nor physics interpretations.

This chapter’s study shows both the successful application of our 3D EOF algorithm that utilizes a temporal covariance matrix designed for future applications on multi-terabytes of data, and physics interpretations of the 3D EOFs from the NASA JPL OGCM. Therefore,

our study has a modest purpose to show the feasibility of (i) computing 3D EOFs from a climate model output, and (ii) making physics interpretations from the point of view of ocean dynamics. Our method of weighted 3D EOF computation and our 3D graphics Python code may be useful tools for both climate professionals and students. The comprehensive code can be accessed at https://github.com/dlafarga/calc_3D_EOFs.

2.2 Data and Methods

2.2.1 Temperature Data from the NASA JPL OGCM

We use the monthly temperature data from the NASA JPL non-Boussinesq ocean general circulation model (OGCM) from January 1950 to December 2003 [54, 55]. The horizontal resolution of the JPL OGCM is 1/4-degree latitude-longitude and was later aggregated to 1-degree as in [33]. After aggregating the data, the files were separated into MATLAB files according to their month and year. Figure 2.1 shows one file of 3D ocean temperature data for January 1950.

The OGCM data encompasses the entire global ocean with land represented by a nonnumber (NaN) grid cell and temperatures as a $^{\circ}\text{C}$ grid cell. There are a total of 32 depth layers whose depth boundaries are given in Table 2.1. As an example, take the first boundary from Table 2.1 as 5. This means the top layer extends from 0 to 5 meters. Similarly, the second layer is from 5 to 10 meters, etc.

We use the JPL non-Boussinesq (mass-conserving) OGCM as it well portrays ocean dynamics. The model was based on both the S-Coordinate Rutgers University Model [56], and the Regional Ocean Model System [57, 58]. The JPL non-Boussinesq OGCM served as a resolution to the uncertainties in altimetry sea-surface-height (SSH) and ocean bottom pressure

January 1950 OGCM Temperatures[°C]

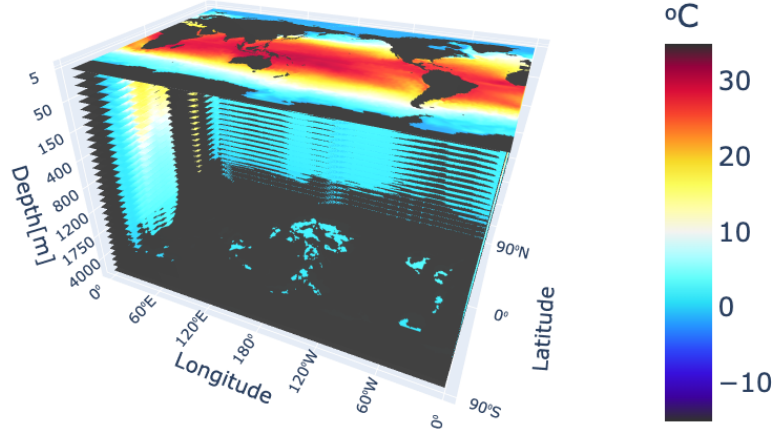


Figure 2.1: Ocean water temperature from the NASA JPL OGCM output for January 1950.

(OBP) in Boussinesq approximations [54, 55] identified by [59]. Both SSH and OBP's consistency within the model have been validated by [60] and [61]. Because the OGCM output is dependable in its representation of ocean dynamics, we find it appropriate for our EOFs.

Table 2.1: Depths(m) of 32 layers of the NASA JPL OGCM

5	10	20	30	50	75	100
125	150	200	250	300	400	500
600	700	800	900	1000	1100	1200
1300	1400	1500	1750	2000	2500	3000
3500	4000	5000	5500			

2.2.2 2D Formulation of the Space-Time Data Matrix

The OGCM output is an array written in 3D space grid boxes separated monthly for each year. Our time spans from January 1950 to December 2003. If we take January as an example, it will have 54 years from 1950 to 2003; the same can be said for February, March, ..., December. Because of seasonal variation, we compute EOFs for each month from January,

February, ..., December; therefore, we will compute 12 sets of EOFs. Let us first consider the data on a single layer with a given depth and time. We use January in the following description.

Denote the data matrix of one year and layer by $\mathbf{T}_{layer, year}$ and the final space-time data matrix by $\mathbf{X}$. The gridded temperature data ($T_{longitude, latitude}$) for the first year (y_{1950}) and first layer(l_1) form a 360×180 matrix shown below:

$$\mathbf{T}_{l_1, y_{1950}} = \begin{pmatrix} T_{0, -89.5} & T_{0, -88.5} & \dots & T_{0, 89.5} \\ T_{1, -89.5} & T_{1, -88.5} & \dots & T_{1, 89.5} \\ T_{2, -89.5} & T_{2, -88.5} & \dots & T_{2, 89.5} \\ \dots & \dots & \dots & \dots \\ T_{359, -89.5} & T_{359, -88.5} & \dots & T_{359, 89.5} \end{pmatrix}_{360 \times 180} \quad . \quad (2.1)$$

The row data of the matrix are ordered from south to north, starting at latitude -89.5° and ending at latitude 89.5° , totaling 180 columns. The column data are ordered according to longitude from east to west, starting at longitude 0° and ending at longitude 359° , totaling 360 rows. Namely, the first 180 points in the first row of the matrix represent $89.5^\circ S$ to $89.5^\circ N$ for a longitude 0° as shown by the first row of the matrix.

The next row will show the same latitudes, $89.5^\circ S$ to $89.5^\circ N$, for the next longitude 1° . This continues until longitude reaches 359° , the last row of the data matrix. Flattening this matrix by rows gives us a row vector of length $360 \times 180 = 64,800$.

$$\vec{T}_{l_1, y_{1950}} = \left(T_{0, -89.5} \quad T_{0, -88.5} \quad \dots \quad T_{359, 89.5} \right)_{1 \times 64800} \quad (2.2)$$

We transpose the $\vec{T}_{l_1, y_{1950}}$ row vector to a column vector and aggregate all the column vectors from 1950 to 2003 to form a space-time data matrix for the first layer of January. The 2D

space-time data matrix for a single layer, denoted by $\mathbf{X}_{Jan_2D}$, looks like the following:

$$\mathbf{X}_{Jan_2D} = \begin{pmatrix} \vec{T}_{l_1, y_{1950}} & \vec{T}_{l_1, y_{1951}} & \vec{T}_{l_1, y_{1952}} & \dots & \vec{T}_{l_1, y_{2003}} \end{pmatrix}_{64800 \times 54}. \quad (2.3)$$

Anomalies can be computed from this matrix and multiplied by the area weight of the grid box i , depending on its latitude in radians ϕ_i . These area weights reflect the geometric sizes of their corresponding grid boxes.

$$area_i = \sqrt{\cos \phi_i} \quad (2.4)$$

The result of the multiplication is the area-weighted space-time anomaly matrix. Conveniently, the structure of this matrix allows for the removal of land/NaN data by removing rows, as every year contains the same grid point of land data. For example, if January of 1950 had a land point at $T_{0, -89.5}$, then January 1951 - 2003 will have the same land point at $T_{0, -89.5}$. Only after the removal of land data can the area-weighted space-time anomaly matrix be fed into SVD, or any other method of computing EOFs, and produce the 2D EOFs layer by layer.

2.2.3 3D Formulation of the Space-Time Data Matrix

We can build on the idea of computing 2D EOFs to compute 3D EOFs by stacking 32 layers on top of each other, as shown below:

$$\mathbf{X}_{Jan_{3D}} = \begin{matrix} & Y_{1950} & y_{1951} & y_{1952} & \dots & y_{2003} \\ \begin{matrix} l_1 \\ l_2 \\ l_3 \\ \dots \\ l_{32} \end{matrix} & \begin{pmatrix} \vec{T}_{l_1, y_{1950}} & \vec{T}_{l_1, y_{1951}} & \vec{T}_{l_1, y_{1952}} & \dots & \vec{T}_{l_1, y_{2003}} \\ \vec{T}_{l_2, y_{1950}} & \vec{T}_{l_2, y_{1951}} & \vec{T}_{l_2, y_{1952}} & \dots & \vec{T}_{l_2, y_{2003}} \\ \vec{T}_{l_3, y_{1950}} & \vec{T}_{l_3, y_{1951}} & \vec{T}_{l_3, y_{1952}} & \dots & \vec{T}_{l_3, y_{2003}} \\ \dots & \dots & \dots & \dots & \dots \\ \vec{T}_{l_{32}, y_{1950}} & \vec{T}_{l_{32}, y_{1951}} & \vec{T}_{l_{32}, y_{1952}} & \dots & \vec{T}_{l_{32}, y_{2003}} \end{pmatrix} \end{matrix} \cdot \quad (2.5)$$

2,073,600×54

Here $2,073,600 = 64,800 \times 32$ is the total number of 3D grid boxes for the OGCM output. Following the same way, we formulate our February space-time data matrix, and so on, since our EOFs are computed for each month from January to December. This space-time data matrix for January is one of the 12 matrices used to compute 3D EOFs as outlined in the next section (see Eq. (3.3)).

2.2.4 SVD and Temporal Covariance

There are two methods described in this section: SVD and temporal covariance. We will first describe the SVD method to better understand the relationship between each method. We start by calculating our climatology. We use the entire data history from 1950 to 2003 as our climatology period. Let us use the following simple space-time data matrix notation $[T_{it}]$ to explain our calculation of climatology and anomalies:

$$X = [T_{i,y}]_{N \times M}, \quad (2.6)$$

where i is a grid box simplified from the previous latitude-longitude notation, y is still time, N is the total number of grid boxes, and M is the total number of years. Specifically, $N = 2,073,600$, and $M = 54$. This X can be the space-time data matrix for a particular month, January, February, or another month.

The climatology is computed as the time means of the data matrix, i.e., the mean for each row:

$$\bar{T}_i = \frac{1}{M} \sum_{y=1950}^{2003} T_{i,y}. \quad (2.7)$$

We compute the climatology for each month: January, February, ..., December. Thus, 12 sets of climatology are computed. Figure 2.2 shows the climatology of January and that of August.

Our anomalies are computed as follows:

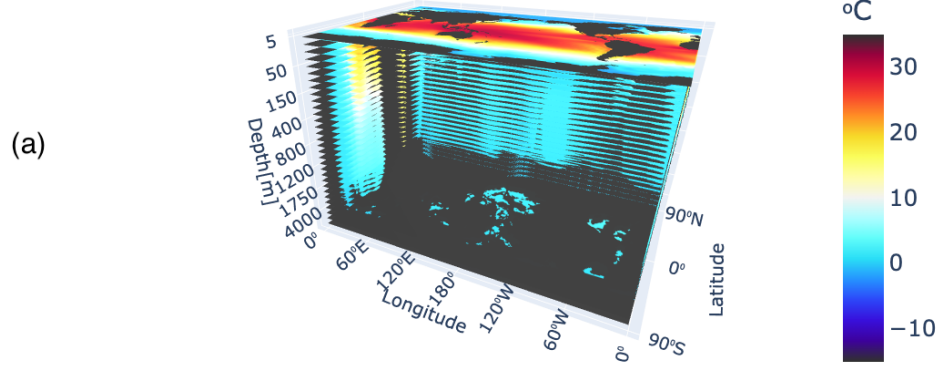
$$A = [a_{i,y}]_{N \times M} = [T_{i,y} - \bar{T}_i]_{N \times M}. \quad (2.8)$$

These anomalies are multiplied by their spatial weights to, as mentioned previously, better reflect the geometric sizes of their corresponding grid boxes. A 3D grid box in the equator region is much larger than one in the polar region. In 2D the weight is related to the area of a grid box (see Eq. (2.4)). Naturally, in our 3D case, the weight for a 3D grid box i is related to the volume of the grid box:

$$w_i = \sqrt{\Delta d_i \cos \phi_i}, \quad (2.9)$$

where Δd_i is the thickness of the layer for the grid box i and ϕ_i is the latitude of the grid box i in radians. The square root is needed due to the symmetry of the formulation of the eigenvalue problem of a covariance matrix [33].

January Climatology Temperatures[°C]



August Climatology Temperatures[°C]

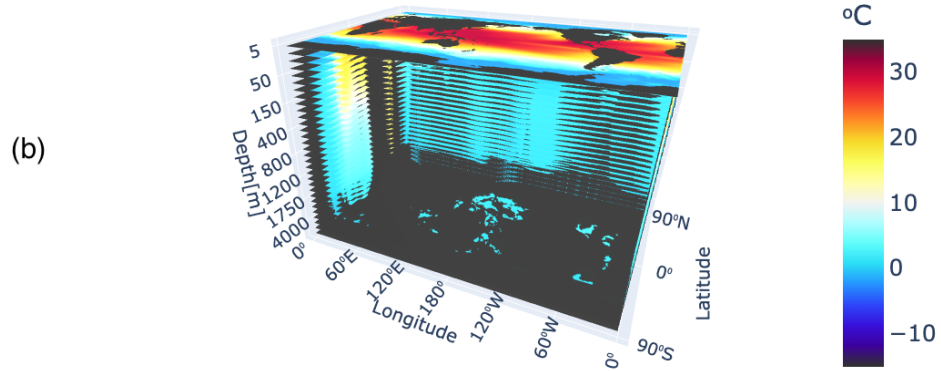


Figure 2.2: Climatology for January and August computed using the equation above from the NASA JPL OGCM data.

Often, anomalies are standardized by dividing by the standard deviation. However, in the deeper ocean layers, temperature has a very small standard deviation. Dividing by the very small standard deviation would only amplify noise and cause large errors for the EOFs in the top layers. To avoid such errors, we simply use non-standardized volume-weighted anomalies. Up to now, the NaN missing data are still in the anomaly data matrix. We remove the rows of NaN and produce a volume-weighted space-time anomaly matrix without missing data.

We denote this weighted anomaly matrix without missing data by

$$B = [b_{j,y}]_{N' \times M} , \quad (2.10)$$

where N' is the number of grid boxes with data. Then, we can feed this space-time data matrix to SVD to generate EOFs to demonstrate spatial patterns and principal components (PCs) to show temporal patterns. The SVD decomposition of B is as follows:

$$B = UDV^T, \quad (2.11)$$

where the superscript T indicates the transpose of a matrix. In this SVD formula, U is an $N' \times M$ ($N' > M$) spatial matrix. Each column vector of U is an EOF. Matrix V has dimension $M \times M$ and is the temporal matrix; each column is a PC. Matrix D is a diagonal matrix, whose elements $d_k (k = 1, 2, \dots, M)$ correspond to the standard deviations of the EOFs and PCs. The next sections will explain these EOF and PC patterns in U and V , as well as the standard deviations in D .

Although SVD is a convenient way to compute EOFs and PCs, the covariance matrix method is often helpful in interpreting the meaning of EOFs and PCs from the point of view of variance and covariance in climate dynamics. Further, this method was traditionally used in the EOF calculations before using SVD in recent years. The traditional method often uses the eigenvalue problem for a spatial covariance matrix to compute EOFs. However, in our case, the temporal dimension M is much smaller than the spatial dimension N' , hence the temporal covariance matrix is much smaller than the spatial covariance matrix. It is preferred to solve the eigenvalue problem for the temporal covariance matrix, which is much smaller than the spatial covariance matrix. The resulting eigenvectors of the temporal covariance matrix are PCs: $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_M$. The EOFs can be obtained by the projection of

the weighted anomaly matrix B to the PCs:

$$\vec{u}_k = \frac{B\vec{v}_k}{|B\vec{v}_k|}, \quad k = 1, 2, \dots, M, \quad (2.12)$$

where $|B\vec{v}_k|$ is the Euclidean norm of the vector $B\vec{v}_k$. The temporal covariance matrix, denoted as C_T , is defined as follows:

$$C_T = \left[\frac{1}{N'} \sum_{i=1}^{N'} b_{i,y} b_{i,y'} \right]_{M \times M} = \frac{1}{N'} B_{M \times N'}^T B_{N' \times M}. \quad (2.13)$$

An eigenvector $\vec{v}$ of this temporal matrix C_T is defined as a vector whose orientation is not changed after the action of C_T . Namely, the matrix action vector $C_T \vec{v}$ is parallel to the original vector. Thus, there exists a scalar, denoted by λ , so that one vector is the scalar multiplication of the other, i.e.,

$$C_T \vec{v} = \lambda \vec{v}. \quad (2.14)$$

This eigenvalue problem has M solutions whose M eigenvectors ($\vec{v}_k$, $k = 1, 2, \dots, M$) are the same PCs computed from SVD (see the rigorous mathematical proof in [47]). The eigenvalues (λ_k , $k = 1, 2, \dots, M$) are proportional to the d_k^2 , $k = 1, 2, \dots, M$ computed from SVD. Therefore, eigenvalues λ_k , $k = 1, 2, \dots, M$ correspond to variances since d_k , $k = 1, 2, \dots, M$ correspond to standard deviation.

Eigenvalues are sorted from the greatest variance to the smallest. Thus, the first eigenmode will show us the most significant pattern, as it has the most variance in climate dynamics. The last modes are usually less accurate and less important as they contain much noise and small-scale variations [62]. The details of the EOF and PC derivations can be found in [33].

We have computed the EOFs, PCs, standard deviations, and variances for each mode using both the methods of SVD and temporal covariance. We compared their results as a cross-

check and explored the physical interpretation of their meaning. These results are described in the next section.

2.3 Results

2.3.1 Eigenvalues and Variances Explained

EOFs, PCs, SVD eigenvalues, and eigenvalues of the temporal covariance matrix were computed for each month from January to December. Altogether, 12 sets of EOFs were generated. Figure 2.3(a) shows the first eigenvalue of the temporal covariance matrix for the weighted temperature anomalies for each month from January to December.

These eigenvalues represent variances of the space-time data projection onto their corresponding EOF 1 for each month. The peaks in January and August imply large variances in both months, relative to their neighboring months. Our examination of all the EOF patterns has shown that those two peaks are likely the manifestation of the relatively stronger equatorial upwelling in January and August.

Another important feature of Fig. 2.3(a) is the obvious difference between January and December. The difference is around 3% and is likely an indication of the presence of a temporal trend in the space-time temperature data. If there were no temporal trend, the annual cycle would prevail, showing a 12-month cycle instead, and hence December's variance would be similar in size to January's. Visually, this would be observed as a smoother transition between December to January. Later on, we further quantitatively verify our assumption here by solving the eigenvalue problem for the detrended data. Indeed, the annual cycle of λ_1 appears.

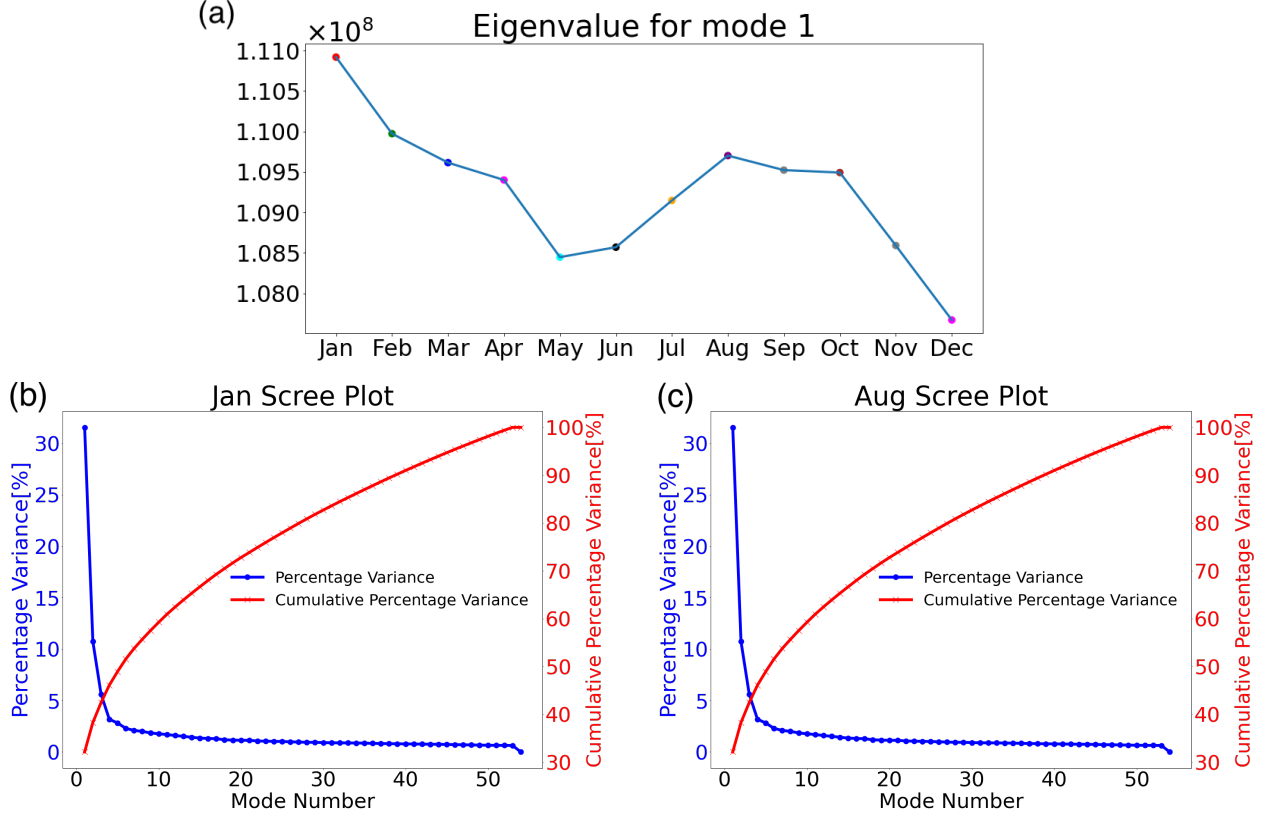


Figure 2.3: (a) The first eigenvalue λ_1 of the temporal covariance matrix for the weighted temperature anomalies for each month from January to December. Units of the first eigenvalue are the same as the covariance matrix in equation 2.13. (b) The scree plot for January based on data p_k and q_k . (c) The scree plot for August is based on data p_k and q_k .

Figures 2.3(b) and (c) show the scree plot for both peak months, January and August. The scree plot plots every eigenvalue for its respective month as a percentage variance or cumulative variance. The purpose is to quantify the variance explained by each mode and to identify how many modes are needed to explain a significant amount of variance. The blue dotted line explains the percentage variance of each EOF mode relative to the total variance, i.e.,

$$p_k = \frac{\lambda_k}{\sum_{\ell=1}^M \lambda_\ell} \times 100. \quad (2.15)$$

The red dotted line is the cumulative percentage variance relative to the total variance, defined as follows:

$$q_k = \frac{\sum_{\ell=1}^k \lambda_\ell}{\sum_{\ell=1}^M \lambda_\ell} \times 100. \quad (2.16)$$

Comparing Figs. 2.3(b) and (c), January shows about $p_1 = 30\%$ variance explained by the first mode, while August shows around 33%. Their second modes drop to 11% and 6% for January and August, respectively. Their third modes are around 5% for both January and August. Figure 2.3(b) shows that for January, eigenvalues λ_4 and λ_5 are close to each other and have a very small difference $p_4 - p_5$. EOF errors are large because they are proportional to the inverse of the difference of two neighborhood eigenvalues, i.e., $1/(\lambda_4 - \lambda_5)$, according to North’s rule of thumb [63]. This implies that the mode mixture between EOF 4 and EOF 5 is stronger, and the geometric pattern of EOF 4 may not have a meaningful physical interpretation due to the uncertainty caused by the mode mixing. For August, this mode mixing occurs from EOF 3. Therefore, in this paper, we focus on the physical interpretations of EOF 1 and EOF 2. Up to mode 2, the cumulative percentage variance q_2 reaches slightly over 40% for January and slightly below 40% for August. Meaning a large number of variances are explained by EOF 1 and EOF 2.

Considering the case of historical data reconstruction, like [33], a number of modes may be used such that 80% or 90% of the variance is represented. For our study here on the NASA JPL OCGM, it takes about 42 modes to represent 90% of cumulative variance for both January and August.

2.3.2 Physical Interpretation of EOF 1: Equatorial Upwelling

Figure 2.4 shows our 3D EOF 1 for January and August at different depth levels. We intend to use these EOF patterns to explain the ocean dynamic features associated with EOF 1, PC 1, and λ_1 . Recall that the first mode of January and August explains 30% and 33% variance, respectively. This means that a third of the cumulative variance in the data for each month is explained by the first mode alone. 2D EOF results from area-weighted anomaly data often show that the first mode is either a temporal trend or ENSO [33]. Interestingly enough, the

first mode of our 3D EOF 1 is not only related to the temporal trend but also to the vertical structure of ocean dynamics known as the equatorial upwelling.

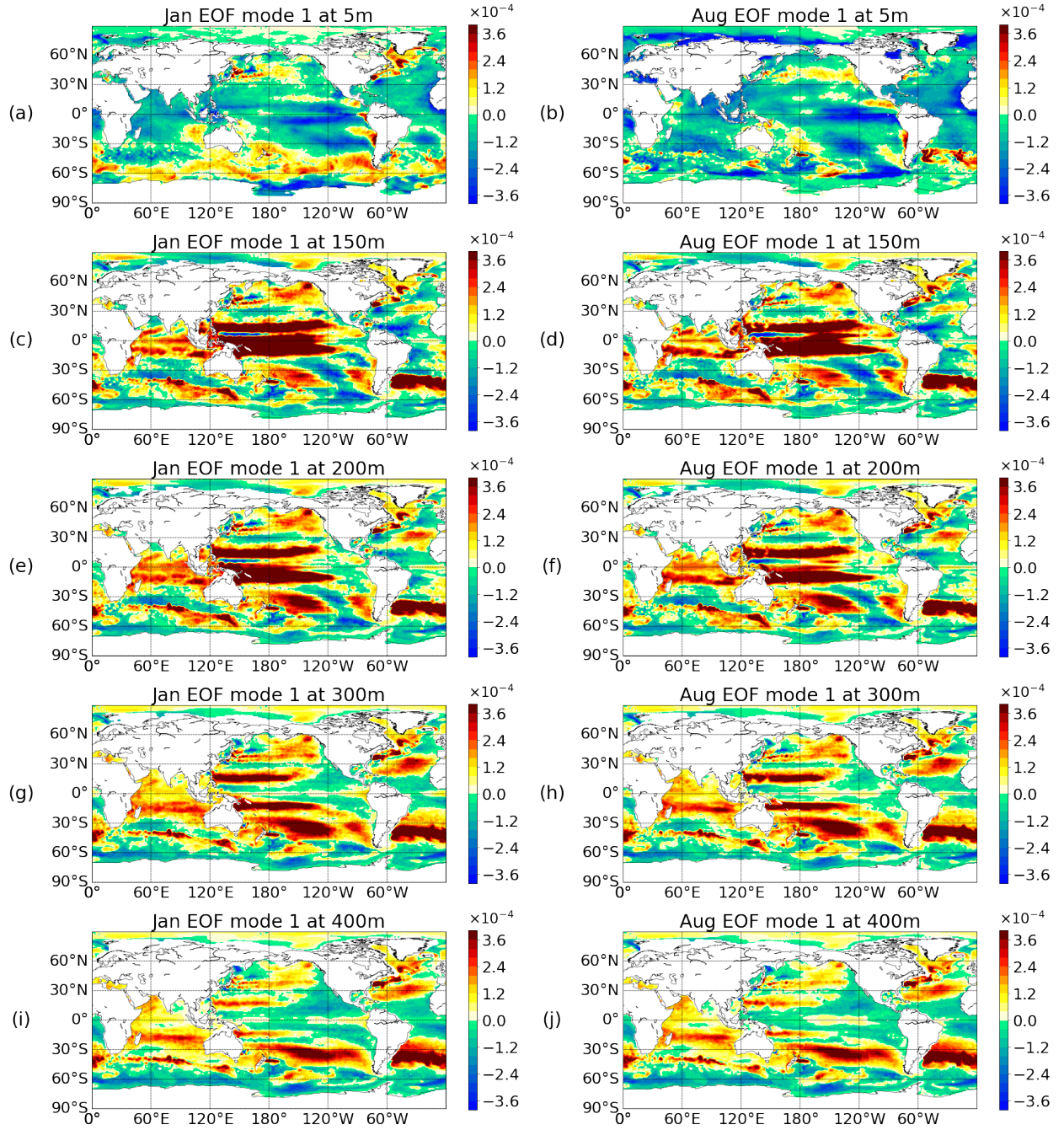


Figure 2.4: EOF 1 for January and August showing the equatorial upwelling. Depths are chosen to show that the strong upwelling occurs at 150 meters and that the upwelling stops at a deeper level around 400 meters.

Equatorial upwelling can be observed in Fig. 2.4 as the horizontal divide of cool temperatures a little north of the equator with warm temperatures straddling it. [64] and [65] define equatorial upwelling as the rising of water along the equator typically occurring from 200 meters to the surface. Near the surface, the equatorial surface waters are being deflected by wind and Ekman flow. Thus, the surface waters are pushed away from the equator, and colder deep ocean waters are pumped up, or more formally, the deep ocean water upwells [66]. Our EOF 1 provides a more precise quantitative description of the upwelling. Figures 2.4 and 2.5 show a strong upwelling starting around the depth of 100 meters that stops around 400 meters (specifically, see Figs. 2.4(c) - (h)), and Figs. 2.5(c)-(d)).

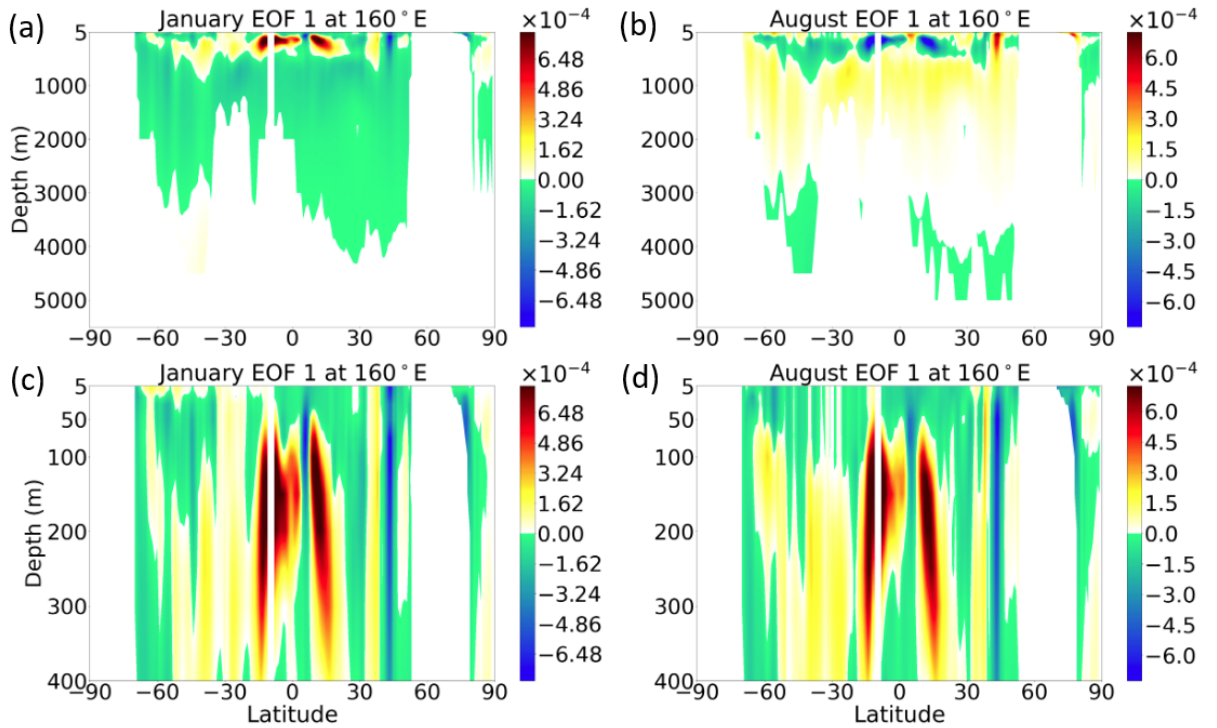


Figure 2.5: A north-south cross-section of the ocean taken at 160°E to show at what layers upwelling occurs. Panels (a) and (b) are the entire scope of all ocean layers. Panels (c) and (d) show the top layers up to 400 meters, as this will better show where upwelling starts and stops.

Despite the focus on equatorial upwelling in this section, there are other notable signals in Fig. 2.4 such as the significant warming pattern from surface to 400 meters in the Southern Ocean between 60°S and 30°S. This is a widely known result of ocean warming due to global warming (see the satellite altimetry data from [61]). This is one of many signals to

dissect from the first 3D EOF mode, but we will mainly focus on equatorial upwelling for the remainder of this subsection.

The signal strength of the equatorial upwelling process varies with depth. Looking at the surface layer for both January and August in Figs. 2.4(a)-(b), the upwelling signal seems relatively weak. January shows a small line of cooler waters a little west of 120°E to approximately 170°E . August is less obvious with a smaller diffused line of cool water between 130°E to 170°E . Although the line of cool waters is narrow and small on the map, the region still covers a large equatorial area because Fig. 2.4 is in the 2D Mercator map projection.

The upwelling region gets larger in deeper layers, particularly at 150 meters we see a very clear divide between 120°E and 180°E in Fig. 2.4. The strong upwelling continues at 200 meters for both months, but January shows a stronger signal as expected from its higher eigenvalue (see Fig. 2.3). At 300 meters, the two warmer strips off the equator begin to fade away. The warm-cool-warm sandwich pattern of the upwelling signal starts to disappear and eventually has little reminiscence at 400 meters suggesting the vertical sandwich has a depth of about 400 meters. The length in the tropical Pacific in the east-west direction stretches about 13,000 kilometers from 120°E to 120°W . The “bread pieces” of warm waters in both south and north are thicker than the thin “cheese filling” in the middle, as indicated by Figs. 2.4 and 2.5.

To show how the upwelling signal changes by depth from the perspective of cross-section, we have made a north-south cross-sectional cut along a meridional line of 160°E , shown in Fig. 2.5. To first identify upwelling in Fig. 2.5, we look for a cold section of water just north of the equator with warmer waters around it. Both January and August show cold waters at the surface between the equator and 10° north. Figures 2.5(a) and (b) suggest that upwelling stops before 500 meters as the 3D EOF 1 is quite uniform and shows little spatial variation for layers deeper than 500 meters. These observations motivated us to focus instead on depths from the surface to 400 meters where we can see more clearly a sandwich

pattern of an equatorial cold (the narrow blue region at the equator) wrapped by two pieces of off-equatorial warm waters (the red regions). At about 100 meters, the warmer sections are distinct, continuing until they reach their strongest reading at around 150 meters. The warmer patches start to become thinner at 200 meters with the southern half between 10°S and 0° almost entirely gone for August at 250 meters and January at 300 meters. There is a small region in the southern patch around 20°S that persists to 400 meters along with the northern warm patch at 20°N . The entire upwelling region is limited to a zonal bend between 20°S and 20°N . However, this is only the EOF 1 pattern, which reflects a temporally mean state. For a particular month, the equatorial upwelling region may be much larger, reaching 30°S and 30°N (Fig. 2.6).

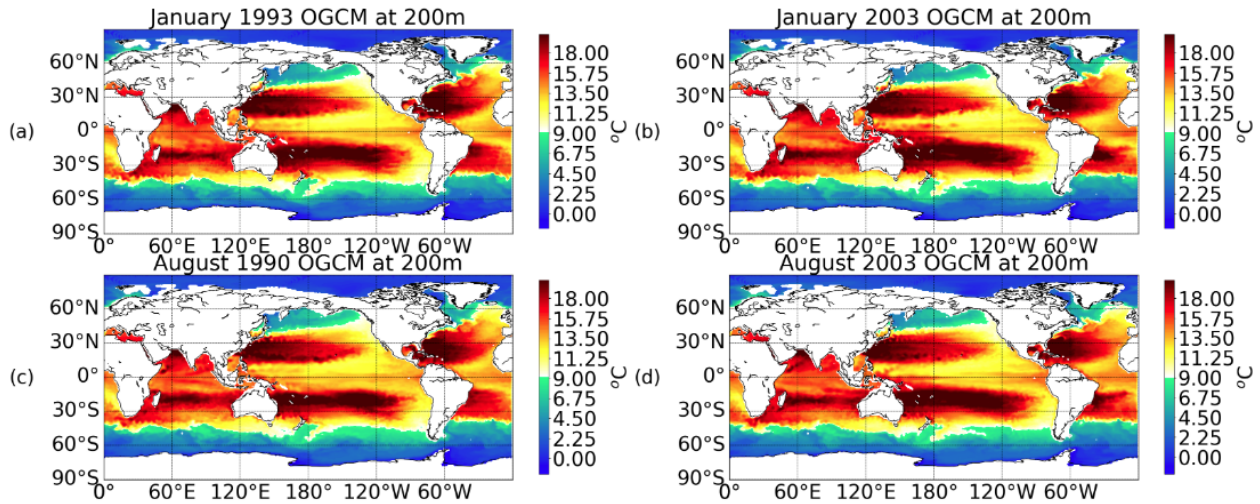


Figure 2.6: NASA JPL OGCM temperature for January and August at 200 meters showing equatorial upwelling. Years are chosen to reflect time with strong upwelling.

EOFs are just a statistical tool used to show spatial patterns. We can identify the dynamics within the spatial patterns, and quantify their variance using EOFs. Equatorial upwelling has not been widely explored in previous EOF calculations therefore it is difficult to compare these results to other calculations such as [45], [46], and [33] whose results are favorable to ENSO. However, equatorial upwelling is an observable dynamic as seen in Fig. 2.6. By visualizing the OGCM data we are able to identify what dynamics to expect of equatorial upwelling in EOF 1.

Figure 2.6 plots the OGCM output for the strongest upwelling months and years: January 1993, January 2003, August 1990, and August 2003. We can observe that upwelling for all dates occurs in both the Pacific and Atlantic oceans. The upwelling signal is the strongest in the western and the central tropical Pacific, in contrast to the ENSO warm pattern over the eastern and central tropical Pacific. The upwelling over the western tropical Pacific in January 1993 (see Fig. 2.6(a)) has a quantified warm-cold-warm sandwich pattern as 20°C - 12°C - 20°C , covering from 30°S to 35°N .

Focusing on January 1993 in Fig. 2.6, the equatorial upwelling patterns in the Atlantic are also limited to the western part of the Atlantic Ocean. The southern part of the sandwich bread piece encompasses a small area. In the Indian Ocean, the northern part of the sandwich bread piece almost does not exist. The upwelling signal is much weaker than that in the Pacific and Atlantic. The same can be said for the upwelling in all other months shown in Fig. 2.6 when examining the water temperature distribution at a depth of 200 meters.

The high-latitude green regions of Fig. 2.6 show another vertical structure of ocean dynamics: Ocean ventilation. This is the process in which surface water, recently in contact with the atmosphere, sinks, transporting carbon and oxygen to deeper ocean layers [67,68]. Regions such as the Labrador Sea are more sensitive to ocean ventilation as its cold winter air will cause surface waters to lose heat to the atmosphere, becoming cooler, denser, less buoyant, and consequently sinking deeper [69,70]. These Labrador sea waters (LSW) that have undergone such convection can be seen in Fig. 2.7.

This cross-sectional figure is a zonal mean, i.e., an average from longitude 0 to 360° . We should expect to see homogeneous values for depths up to 2500 meters at latitudes relative to the Labrador Sea, according to [71]. Observing Fig. 2.7 from 55°N to 60°N , we can identify LSW occurring from 50 meters to 3,000 meters. This range is much deeper than what was suggested from recent papers like [70] and [71].

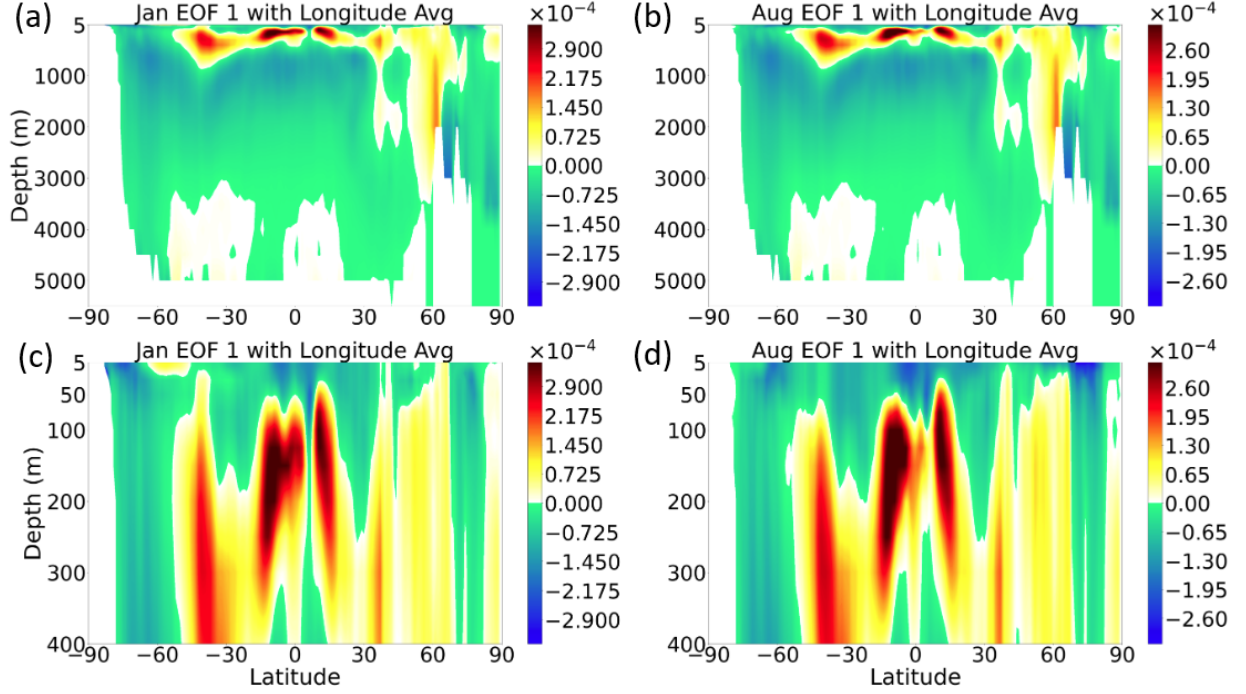


Figure 2.7: Cross-sectional map based on the zonal mean from 0 to 360° longitude degrees to show ocean ventilation at the high latitude regions.

The cross-sectional zonal mean map of Fig. 2.7 also shows equatorial upwelling. The August upwelling signal at about 100 meters depth appears to be stronger than that of January, according to Figs. 2.7(c) and (d).

2.3.3 EOF 2 and the Spatial Pattern of El Niño

Figure 2.8 shows the eigenvalue λ_2 of the temporal covariance matrix for every month. In contrast to λ_1 in Fig. 2.3, there is no longer a large jump from December to January in λ_2 . As previously mentioned, the smooth transition is due to an annual cycle.

Figure 2.9 shows the EOF 2 pattern at five selected layers: 5, 50, 100, 150, and 200 meters. The warm pattern in the eastern tropical Pacific is the well-known ENSO mode, a phenomenon that occurs mostly in the winter months of the Northern Hemisphere, as seen in [48]. The winter ENSO pattern corresponds to the large eigenvalue λ_2 in winter months.

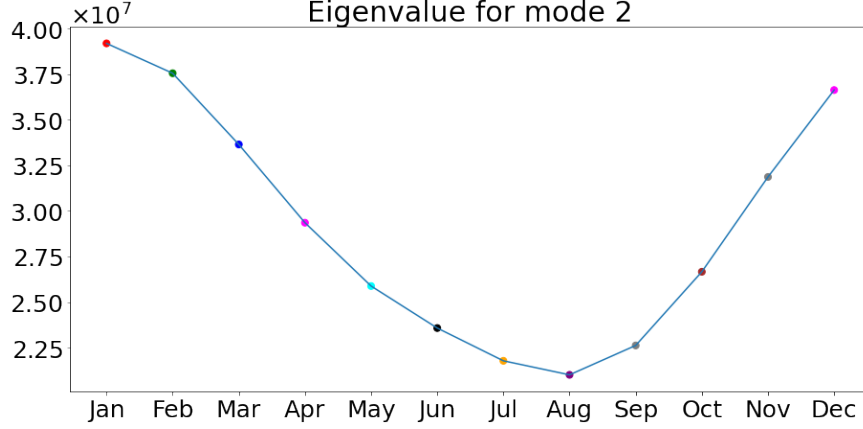


Figure 2.8: The second eigenvalue λ_2 of the temporal covariance matrix for the weighted temperature anomalies. Units of the first eigenvalue are the same as the covariance matrix in equation 2.13. January has the largest value, while August has the smallest value.

Particularly December, November, and January have the three largest λ_2 values according to Fig. 2.8. This is expected as higher eigenvalues imply larger variance and vice versa. The figure also shows the three months with the smallest λ_2 values are August, July, and September indicating there is little chance for an ENSO signal to occur in these months.

As expected from SST studies and textbooks, such as [47] and [33], Fig. 2.9 shows January has an apparent ENSO signal at its surface. The January ENSO pattern vertically exists at the surface and continues to deeper layers. The pattern starts to wither at 150 meters and is entirely dispersed at and deeper than 200 meters (see Figs. 2.9(g) and (i)).

At 100 meters (Fig. 2.9(e)), we see a large continuous blue region over the western tropical Pacific that corresponds to the cold anomalies of ENSO underneath the surface, as shown in [33] and similar to [72]. This cold anomaly stretches east-west over 7,000 km and north-south over 2,000 km, almost twice the area of the United States. This body of cold anomaly water continues to exist at 200 meters (see Fig. 2.9(i)). Our EOF 2 data shows that the body begins around 50 meters and ends around 250 meters. Also at 100 meters, we see the warm tongue in the eastern tropical Pacific split by a small strip of cool waters just north of the equator. This might be part of the signal of equatorial upwelling.

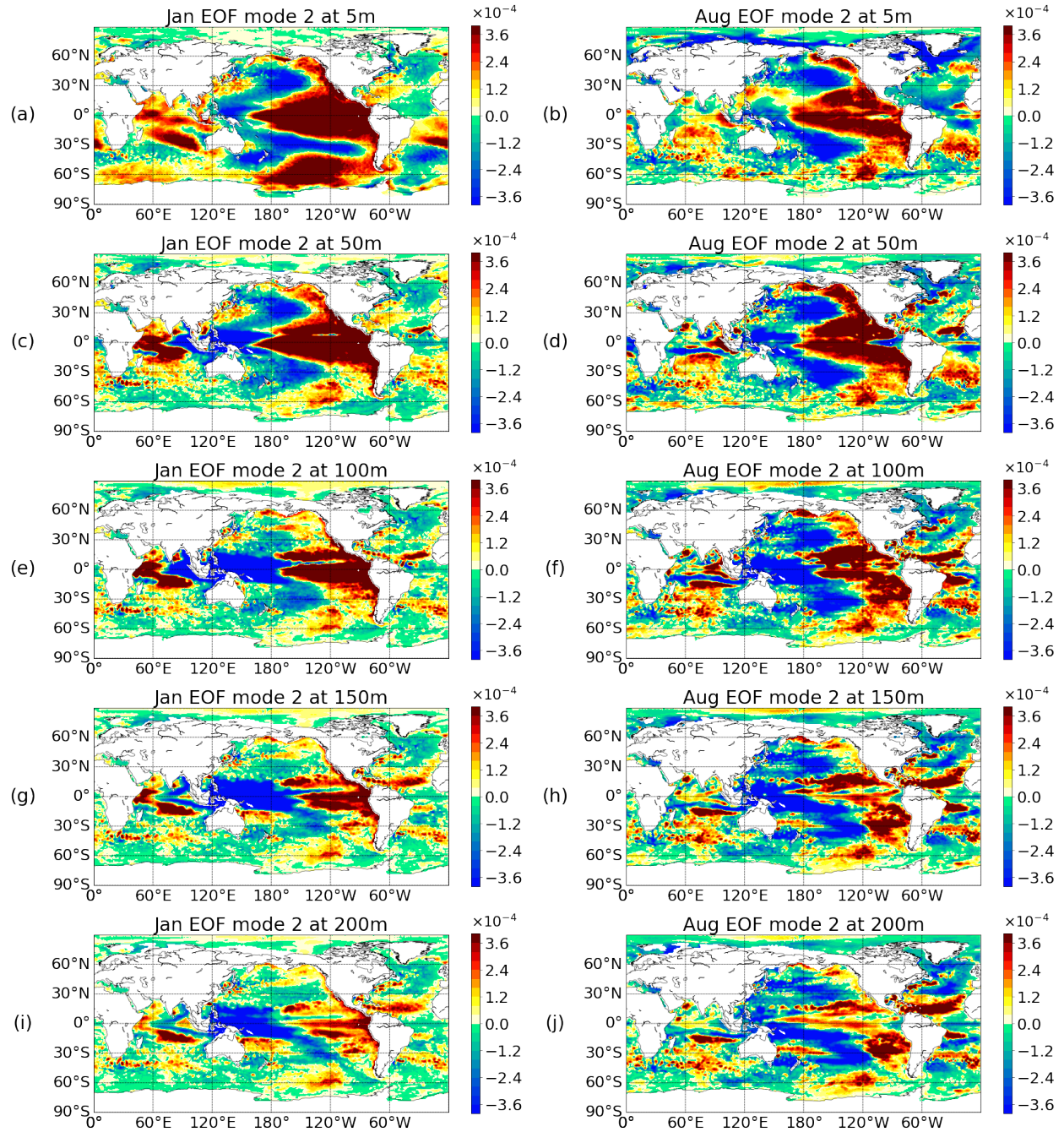


Figure 2.9: EOF 2 for January and August showing the ENSO pattern. Depths are chosen to show the ENSO regions in different depths from surface to 100 meters. The ENSO pattern dies out approximately at levels deeper than 150 meters.

Although the SST ENSO events rarely occur in August, our 3D EOF 2 for August still shows a warm tongue pattern over the eastern tropical Pacific. Despite that, if we were to compare either column in Fig. 2.9 it is evident that the ENSO signal in August is less coherent compared to January. This may mean that the 3D ENSO dynamics continuously exist throughout the year and do not completely die out in the summer.

For the August EOF 2, in the region south of the equator, the cold waters are more persistent than in January. To better understand what depth these cooler waters reach, a zonal cross-section from the surface to the bottom of the ocean (Fig. 2.10) is taken at 10°S . This cross-sectional map shows consistency in the behavior displayed by Fig. 2.9(j).

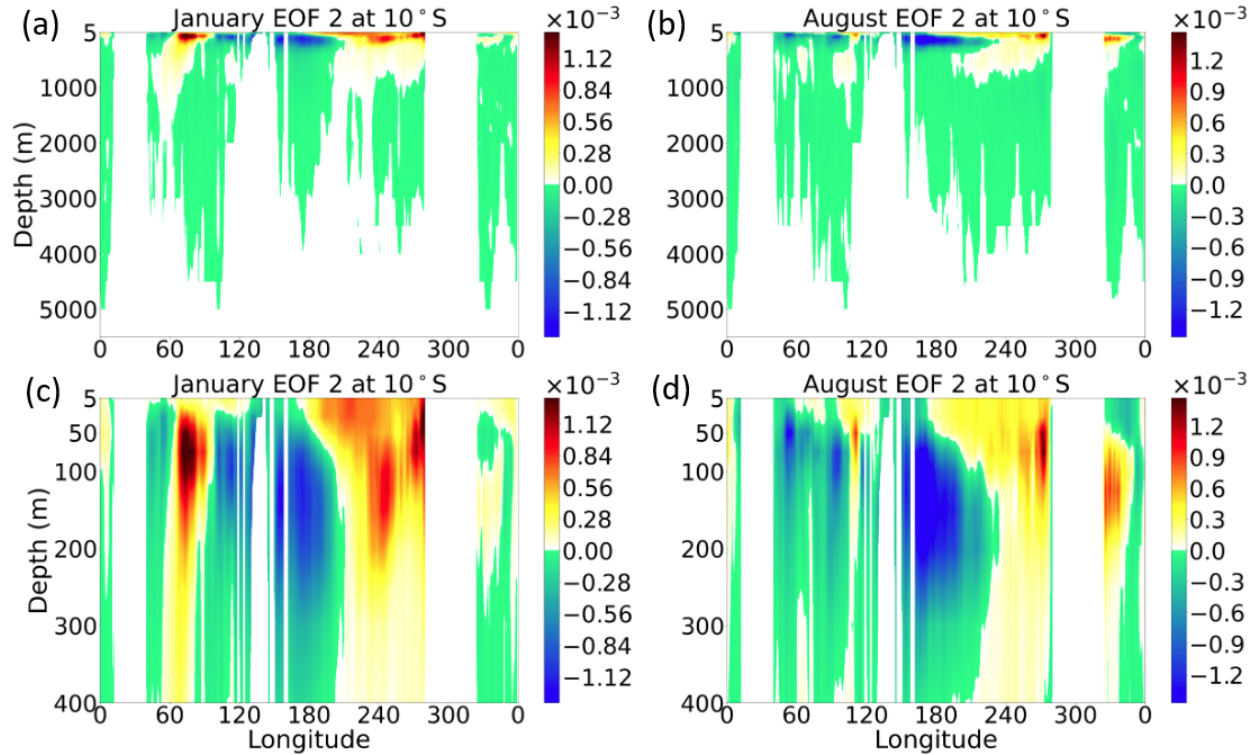


Figure 2.10: Cross-section at latitude of 10°S that shows the separation of the warm and cold anomaly regions in the Pacific.

The vertical map (Fig. 2.10) cuts through the conventional SST ENSO pattern over the tropical Pacific. Figure 2.10(c) shows that the eastern tropical Pacific's warm water body is limited to the top 150 meters, but the western tropical Pacific's cold water body approxi-

mately spans 50 to 250 meters. In January, there is also a cold-warm-cold pattern appearing in the Indian Ocean located within the top 250 meters as seen in Fig. 2.10(c). However, this Indian Ocean signal is not obvious for the August EOF 2.

To show a more robust signal of ENSO, we have plotted the meridional mean from latitude $10^{\circ}S$ to $10^{\circ}N$ in Fig. 2.11. Figure 2.11(c) shows a clear blue-red separation marked by the slanted yellow strip. The slope of the slanted yellow line is approximately -2 meters per longitude. This helps quantify the 3D structure of the El Niño dynamics, such as the thickness of the eastern tropical Pacific’s warm water to the surface, and the thickness of the western tropical Pacific’s cold water underneath the surface. January ENSO’s warm pattern ends at approximately 150 meters while August ENSO’s warm pattern ends at around 120 meters (Figs. 2.11(c) and (d)).

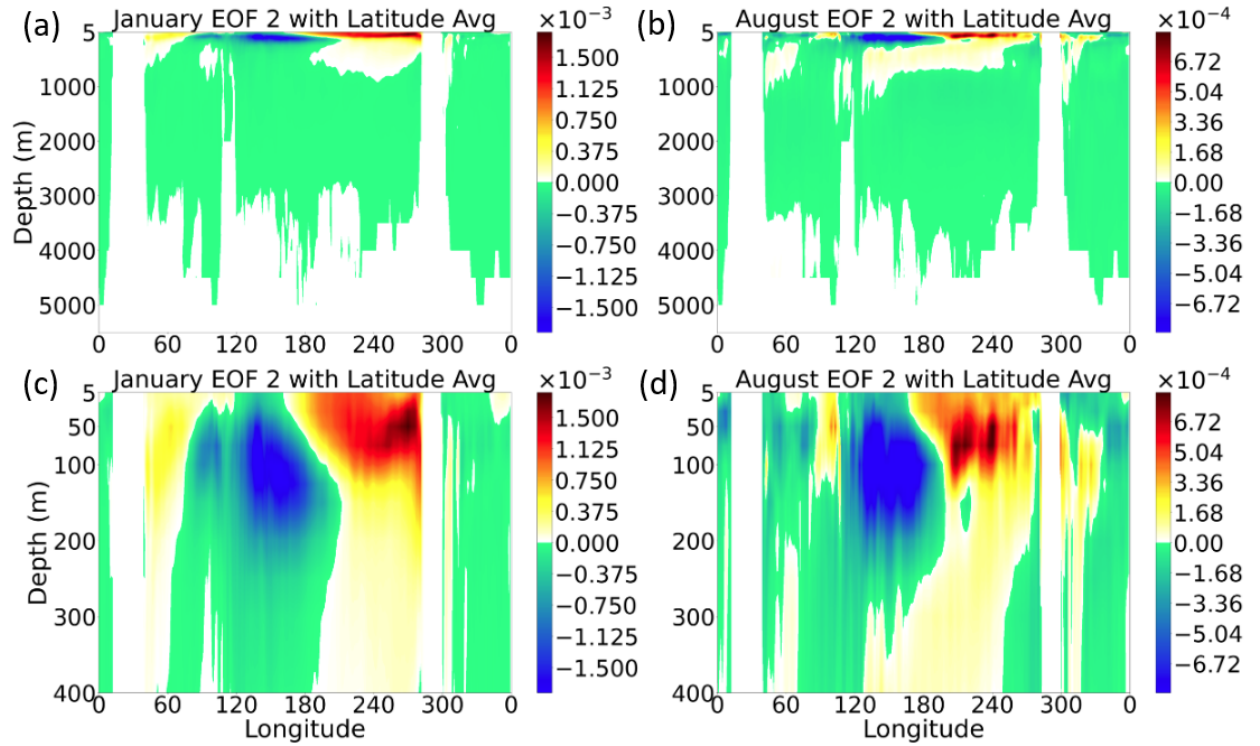


Figure 2.11: A zonal cross-sectional map based on the average from latitude $10^{\circ}N$ to $10^{\circ}S$. The average in the figure are to better show a robust 3D ENSO pattern with not only the surface warm region but also a depth structure.

2.3.4 PCs, Trend and Detrended

This sub-section describes the temporal patterns of the 3D EOF-PC analysis. Principal components(PCs) are the eigenvectors of the temporal covariance matrix and are also the column vectors of the orthogonal V matrix. PCs are on a time scale and show temporal patterns.

PC 1 shown in Figs. 2.12(a) and (b) show an upward trend of temperature anomalies. What could be the cause of this upward trend? Does the association of EOF 1 with equatorial upwelling imply that the upwelling contains some heat from global warming? If the first mode does contain some heat from global warming, how much? The detailed mechanisms of heat transfer and distribution are still to be investigated, even if the association between upwelling and global warming is true. Some literature has already appeared to address these questions. For example, [61] explored the existence of warming in the deep ocean; [73] studied some specifics of global warming on equatorial upwelling. Studies like these may serve as the basis for further quantitative investigations on whether or how upwelling and global warming are correlated.

PC 2 of January is supposed to be a time indicator of ENSO occurrence. This interpretation works well for the SST analysis on ENSO. When regarding 3D ocean, we may need to consider the time delay of heat transfer in the top 1,500 meters. The El Niño pattern in the August EOF 2 has already implied this time delay (see Fig. 2.9). It is unknown how long it takes for the heat accumulation in the eastern tropical Pacific from the surface to 150 meters and for the heat loss in the western tropical Pacific in the deep layers from about 50 to 250 meters. Figure 2.12(c) shows that the minimum of January PC 2 is about -0.32 at January 1990. According to [74] Oceanic Niño Index (ONI) based on SST, the Northern hemispheric winter of 1989 - 1990 was a neutral year, while 1988 - 1989 winter was a La Niña year, 1990 - 1991 winter was neutral too, and 1991 - 1992 winter was an El Niño year.

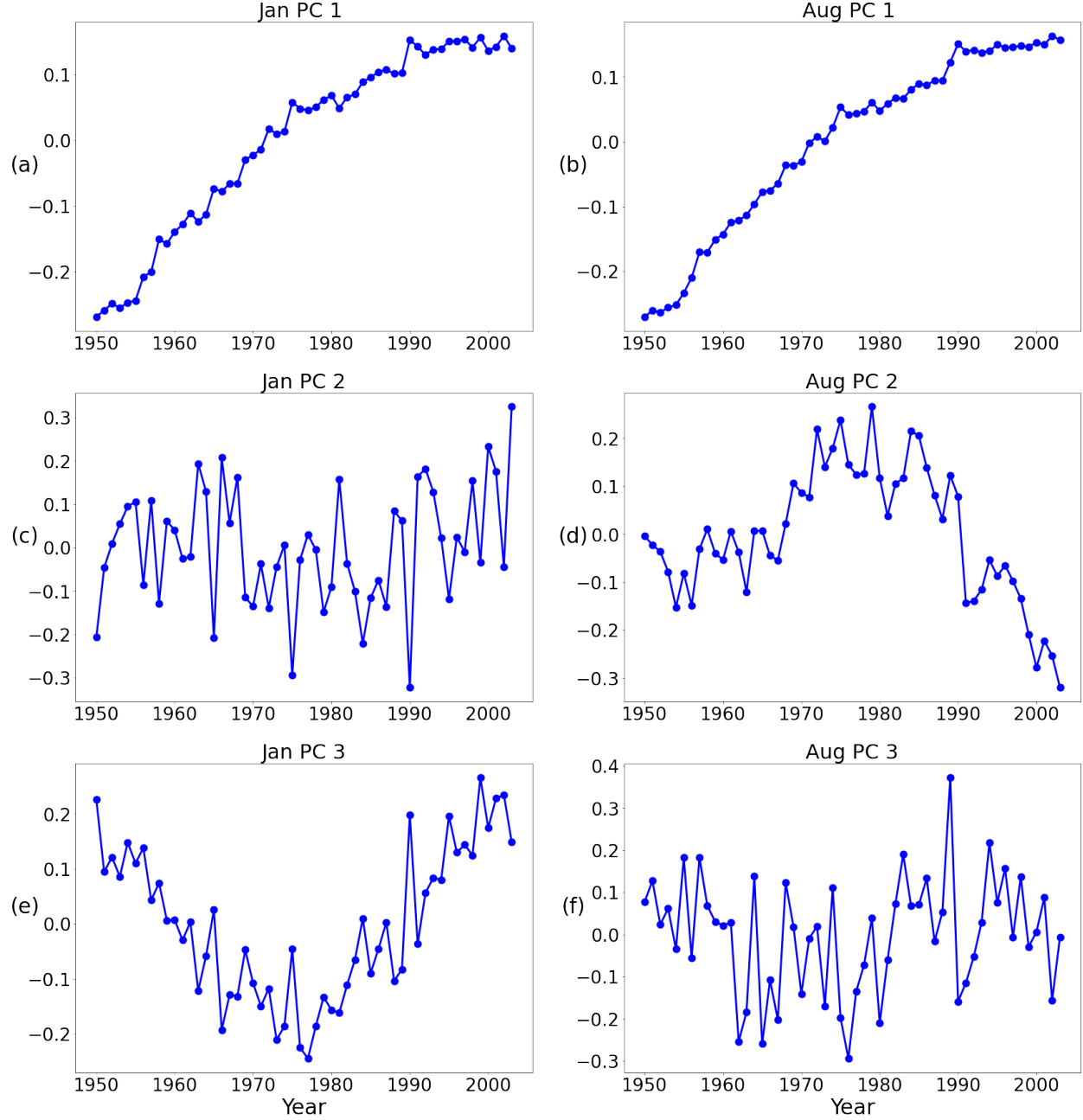


Figure 2.12: Principal components PC 1, PC 2, and PC 3 of January and August from the weighted anomalies of the NASA JPL OGCM output from 1950 to 2003.

The second minimum value of January PC 2 (Fig. 2.12(c)) is about 0.30 occurred on January 1975, which was La Niña month. The maximum January PC 2 is about 0.31 occurred in January 2003. The Northern Hemispheric winter of 2002 - 2003 had an El Niño, while the previous winter was neutral. The above analysis of the maxima and minima implies that PC

2 from the 3D covariance matrix, although associated with El Niño, is not a direct indicator of the surface El Niño time. Further analysis may be needed to interpret PC 2 and its association with El Niño.

A similar analysis may be done for the maxima and minima of August PC 2. Figure 2.12(d) shows that the global maximum of the August PC 2 is about 0.25 occurred in August 1979, which was a neutral month. The global minimum of the August PC 2 is about -0.32 occurred in August 2003, the same year, i.e., 2003, as the global maximum of the January PC 2. As discussed earlier, the Northern Hemispheric winter of 2002-2003 was in El Niño condition according to SST, but the summer of 2003 was in neutral condition.

The January PC 3 shown in Fig. 2.12(e) has a V-shape, going down from the 1950s to the 1970s, and then going up from the 1970s to the 1990s. This shape agrees with that of the Atlantic Multi-decadal Oscillation (AMO) index time series, and also that of the spatial average time series of the Northern Hemispheric average temperature, cumulative Southern Oscillation Index (SOI), and the cumulative weighted SOI1 (CSOI1) (see Figs. 4.2 and 4.3 of [47]). We limit ourselves in claiming similar temporal patterns among these parameters. Whether the similarity is supported by climate dynamics may be worth further investigation.

August's PC 3 shown in Fig. 2.12(f) has its global maximum in 1989 and several local minima in the 1960s and 1970s. This temporal pattern appears to agree with that of the North Atlantic Oscillation (NAO) [75]. Again, we present here only the similarity of the temporal patterns. The climate dynamic support for this similarity needs further investigation.

The trend in PC 1 motivated us to explore the EOF-PC analysis for the detrended anomalies. This was done by subtracting the predicted values of the simple linear regression from the anomalies. The same volume weight is applied to the detrended anomalies as described in the methodology section. Finally, the temporal covariance matrix can be calculated from the weighted de-trended anomalies to obtain EOFs, PCs, and eigenvalues.

By removing the trend, we hope to gain insight as to why there is such a large jump in λ_1 between December and January shown in Fig. 2.3. Figure 2.13 displays the resulting first eigenvalues $\lambda_1^{(d)}$ of the temporal covariance matrix from the detrended anomalies for each month from January to December.

The shape of $\lambda_1^{(d)}$ for the de-trended anomalies is almost the same as that of λ_2 before the detrending, shown in Fig. 2.8, although the values of $\lambda_1^{(d)}$ and λ_2 are not exactly the same. The large gap between December to January in λ_1 is gone in $\lambda_1^{(d)}$ (see Figs. 2.3 and 2.13 respectively), implying the trend may have caused the gap. On that account, the temporal trend in the 12 months from January to December is reflected in λ_1 , corresponding to the equatorial upwelling mode. In removing the linear trend, the ENSO mode becomes the first

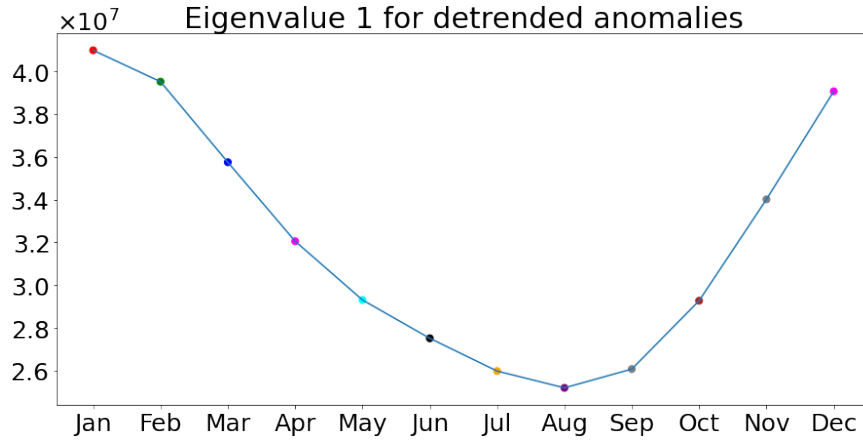


Figure 2.13: The first eigenvalue $\lambda_1^{(d)}$ for each month computed from the detrended anomalies.

mode. Comparing the PCs before detrending the anomalies (Fig. 2.12) and after detrending the anomalies (Fig. 2.14), it is clear that there is a mode shift: Detrended PC 1 corresponds to PC 2 before detrend, and detrended PC 2 corresponds to PC 3 before the detrend. This mode shift due to detrending often occurs in EOF analysis (see Chapters 9 and 10 of [47]).

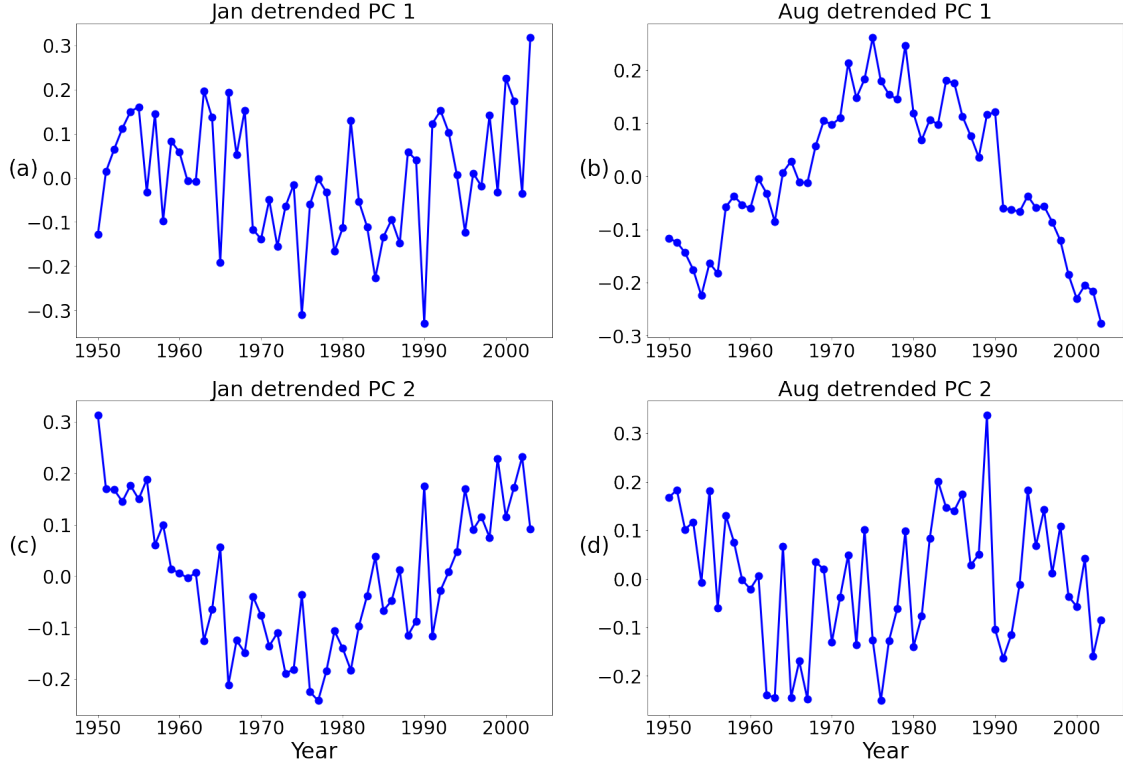


Figure 2.14: Principal components of January and August based on the detrended anomalies.

2.3.5 Compare the 3D EOF 1 to the 2D EOF 1

As a way to compare results with other publications and reinforce the value of computing 3D EOFs, we computed 2D EOFs. It is important to note that 2D EOFs are orthonormal on each layer; in contrast, 3D EOFs are orthonormal for the entire ocean domain of all layers, and not orthonormal on each layer. Figure 2.15 shows the 2D EOF 1 for the top layer based on the anomalies, not the de-trended anomalies, from the NASA JPL OGCM.

To reiterate, the 2D EOF calculation was only applied to the surface layer, so ocean dynamics interconnection with other layers is not considered. Consequently, ENSO, not the equatorial upwelling, appears in the first mode of 2D EOFs, similar to [33], [45], and [46]. The spatial pattern of Fig. 2.15's 2D EOF 1 is very similar to that of the 3D EOF 2 at the top layer shown in Fig. 2.9(a). However, the data for Fig. 2.9(a) is not normalized to one and is not orthogonal to the 3D EOF 2 at the same layer.

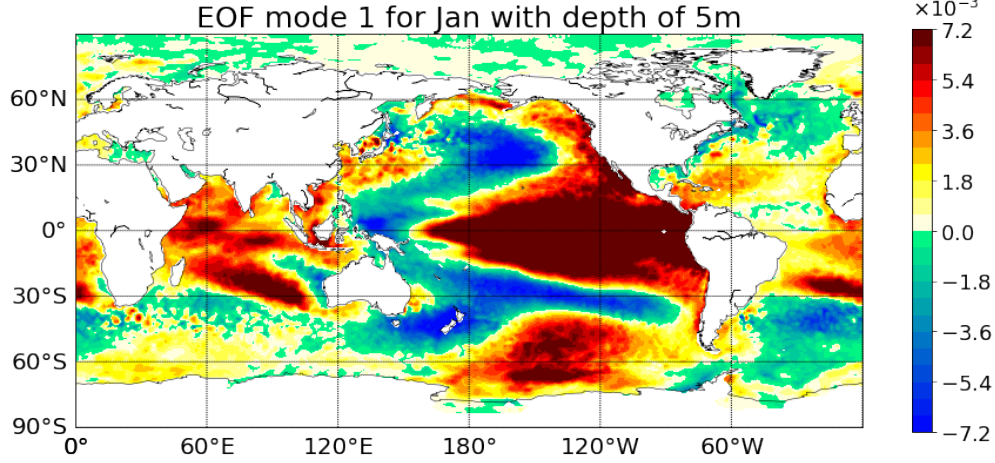


Figure 2.15: 2D EOF 1 for the top layer based on the area-weighted anomalies from the NASA JPL OGCM. The ENSO signal is shown in this EOF 1 for the surface layer. The 2D EOFs are computed layer by layer.

2.4 Conclusion and Discussion

The main purpose of this chapter is to show the method of computing 3-dimensional EOFs and interpret the physical meanings of the first two modes. To compute these EOFs, we used the temporal covariance of the volume-weighted anomalies based on the NASA JPL OGCM's monthly output of 1.0 latitude-longitude degree, 33 layers from surface to 5,500 meters depth, and a time span from 1950 to 2003. The eigenvectors of the temporal covariance are PCs. The normalized projections of the volume-weighted anomalies for the PCs form the geometric EOFs. Removing the volume factors, we obtain the physical EOFs, which are displayed in our EOF figures.

Our method to compute the January PCs starts from the formulation of the $N \times M$ space-time data matrix, with the first dimension being the total number of grid boxes and the latter being the total number of years. Anomalies are multiplied by the volume weights of each depth, resulting in the weighted anomalies. After removing land/NaN data, the new weighted anomaly matrix transpose multiplied by itself forms the temporal covariance matrix.

We also computed the EOFs and PCs using SVD for the volume-weighted anomalies. Although SVD is capable of handling our data at the 1.0-degree resolution, it cannot handle the higher resolution due to computer memory resource limitations. Thus, our temporal covariance approach is useful and will be used later on to deal with higher-resolution data. Our application of SVD here is for the purpose of verifying our temporal covariance approach.

NASA JPL’s OGCM was used before in [33] where EOFs were computed layer by layer, i.e., the two-dimensional EOF, or 2D EOF. The paper shows the ENSO signal in the first one or two EOFs, and it did not reveal the equatorial upwelling mode. This was due to the nature of a 2D calculation, which does not take into account the deep ocean dynamics interconnected among different layers and instead treats each layer in an isolated fashion. Our 3D EOFs show covariances across depth layers. Consequently, equatorial upwelling appears as 3D EOF 1 and explains a third of the data’s variance. In contrast, the ENSO mode, or 3D EOF 2, only explains 10% of the variance.

By observing the covariance among the temperatures of different layers, we were not only able to identify different ocean dynamics than the previous 2D calculations, but we were also able to pinpoint how deep each characteristic spans. Equatorial upwelling can be seen from 100 to 400 meters. ENSO is observed mostly in the upper layers until 150 meters, and at 100 meters, we can observe a large cool water patch in the western tropical Pacific Ocean.

Although through the 3D EOFs and PCs, we were able to identify some meaningful ocean dynamic patterns both in the deep ocean and surface, there is still much to explore in the resulting dynamics and the effect of detrending. Further exploration of the ocean dynamics may be made to show how the 3D EOFs are related to climate oscillation patterns, such as AMO and NAO.

Our EOFs can serve not only as a means to understand dynamics but also help in reconstructing historical climate data [33]. The 3D nature of the EOFs allows the covariance

properties between layers to be used in the data reconstruction. The covariance between points of different depths are expected to add skills to reconstruction. For example, the 3D in-situ observational data can likely support more 3D EOF modes than the 2D EOF modes in the subsurface of the ocean, where in-situ observations are scarce. Can we extend the reliable reconstruction of historical data to a longer history compared with the conventional 2D reconstruction? Can the historical in situ data support a set of 3D EOFs with a finer spatial resolution? More questions like these can be asked about the 3D EOFs with regard to their applications, computations, climate dynamics, and mathematical theory.

Chapter 3

Partition Input Method for Computing 3-Dimensional Empirical Orthogonal Functions for Big Space-Time Data: An Example of High-Resolution Ocean Assimilation Output

3.1 Introduction

In climate science, principal component analysis (PCA) or empirical orthogonal functions (EOFs) serve to identify and analyze dominant patterns of variability. They have been extensively applied to explore climate phenomena such as the El Niño-Southern Oscillation (ENSO) [5], Central Pacific (CP) ENSO [6–8], the meridional modes [9, 10], and marine heat waves [31]. These studies tend to focus on sea surface temperature anomalies, but analyses

such as [6, 76] moved beyond the surface to define the subsurface structure of leading spatial patterns through the lagged correlation between principal components (PCs) and subsurface temperature anomalies. Such research has demonstrated the importance of understanding the 3D ocean dynamics that influence the global weather and climate [5, 10, 77].

Most of the aforementioned EOF calculations used classical singular value decomposition (SVD) algorithms, which require the input data to be less than the total amount of memory for a computer used for this calculation [29]. When the computer memory is limited (e.g., 16 GB), one can compute the EOFs in a reduced spatial domain, such as computing 2D EOFs for a single layer of the ocean. Many of the earlier EOF studies on ocean data were done this way: computing EOFs layer-by-layer (2D EOFs). However, this 2D EOF approach fails to capture cross-layer covariances, which provide a more realistic representation of ocean dynamics. Thus, a 3D EOF calculation is a necessity for explaining 3D ocean dynamic patterns. The 3D EOF computation via the classical SVD method requires all layers of data to be loaded simultaneously, which often surpasses the memory of a computer. These limitations prevent the 3D EOF computations from using SVD when facing a large space-time data matrix. Consequently, some 3D EOFs computed in the past did not use a large amount of data, but were limited in smaller spatial domains [22] or coarse resolutions [23].

One method for big data SVD uses stochastic algorithms, which iteratively calculate and update PCs by computing gradients from randomly sampled portions of the data. Examples of such approaches appear in [39–41], which investigate optimal gradients to achieve both fast convergence and accurate PCs. These algorithms are convenient for large datasets because they process only data samples, avoiding the need to load the full dataset into computer memory. Additionally, they can compute a specified number of PCs and EOFs, saving memory and computing time when the full set is not needed. Although stochastic algorithms offer a solution for big data SVD, our research requires interpreting the exact 3D PCs and EOFs from the climate data, which is not ideal for stochastic algorithms.

Another alternative could be a machine learning approach: using an autoencoder to extract essential information and represent it in another space [34]. Unlike EOFs, however, the resulting latent spaces are not orthogonal nor naturally ordered by variance. Therefore, the autoencoder requires constraints in order to produce results that are physically understandable, like EOFs. In a similar vein, tensor decomposition (TD) can also be used to produce latent spaces on large data, but requires constraints for interpretability [38]. Although autoencoders and TD have been used in climate data analysis [35–37], they could not substitute a 3D EOF analysis. We have employed TD’s data partition idea prior to the decomposition and developed our partition input method (PIM) for 3D EOF calculation.

While our matrix partition is similar to that of TD [38], we retain the traditional EOF analysis to preserve the clear physical interpretations of spatial patterns and variability. The partitioning allows us to handle large amounts of data, while accounting for the local computer memory, without compromising the 3D analysis. Additionally, we employ temporal covariance to work with a smaller time-based matrix rather than a space-based one, following approaches similar to [23, 42]. By combining partitioning with temporal covariance, we successfully compute 3D EOFs from high-resolution data on limited computer memory.

The remainder of this Chapter is organized as follows. Section 3.2 describes the high-resolution eddy-resolving Global Ocean Physics Reanalysis (GLORYS) data and provides a comprehensive summary of the PIM used to compute temporal covariance and 3D EOFs. Section 3.3 details the 3D visualization techniques implemented to display the results of the first EOF, presents physical interpretations of the first three 3D EOFs, and compares 2D results computed from the same GLORYS data to 3D results. Section 3.4 contains the conclusion and discussion.

3.2 Data and Method

3.2.1 Description of Big Data: Global Ocean Physics Reanalysis

The dataset used in this research is the Global Ocean Physics Reanalysis (GLORYS) output [78], a high-resolution eddy-resolving global ocean simulation available at Copernicus Marine Environment Monitoring Service (CMEMS). It has a spatial resolution of $1/12^\circ$ latitude by $1/12^\circ$ longitude with 50 depth layers covering the entire global ocean from 80°S to 90°N and 180°E to 180°W , extending from the ocean surface to 5,727 meters (see Table 3.1 for details) [79]. For the interpretation of our calculation results for the phenomenon of El Niño and its global effects, we use the December-January-February (DJF) boreal winter mean of GLORYS ocean temperature data from 1993/1994 to 2020/2021.

Table 3.1: The depth values of the 50 layers in the GLORYS model (unit: meters). Values were rounded to the first two decimal places.

0.49	1.54	2.65	3.82	5.08
6.44	7.93	9.57	11.41	13.47
15.81	18.50	21.60	25.21	29.44
34.43	40.34	47.37	55.76	65.81
77.85	92.33	109.73	130.67	155.85
186.13	222.51	266.04	318.13	380.21
453.94	541.09	643.57	763.33	902.34
1062.44	1245.29	1452.25	1684.28	1941.89
2225.08	2533.34	2865.70	3220.82	3597.03
3992.48	4405.22	4833.29	5274.78	5727.92

When downloading the GLORYS temperature data in NetCDF format, each file represents one year and is an array in the dimensions of time, depth, latitude, and longitude. Each file has approximately 5 billion entries computed as follows: $12 \text{ months} \times 50 \text{ depth layers} \times 1/12^\circ \text{ latitude} \times 1/12^\circ \text{ longitude}$, which is equal to $12 \times 50 \times 2,041 \times 4,320 = 5,290,272,000$ entries. Reading in a single file, where each entry has 32 floating-point precision and is of 4 bytes, would require 21 GB of computer memory: $5,290,272,000 \text{ entries} \times 4 \text{ bytes} \approx 21.2 \text{ GB}$. For

a single month or season, this would only take 1.75 GB per file or year.

$$\begin{aligned}
& 50 \text{ depth layers} \times 1/12^\circ \text{ latitude} \times 1/12^\circ \text{ longitude} \\
& = 440,856,000 \text{ entries} \times 4 \text{ bytes} \\
& \approx 1.75 \text{ GB per year.}
\end{aligned} \tag{3.1}$$

Our EOF computing involves 28 boreal winter seasons from 1993/1994 to 2020/2021. Therefore, our data size will be equal to:

$$1.75 \frac{GB}{\text{file}} \times 28 \text{ files} = 49 \text{ GB.} \tag{3.2}$$

Storing the data is necessary for our EOF computations. However, it is impossible to fully load the data of 49 GB into a computer with 16 or 32 GB of memory. Our research has developed an algorithm to tackle this task, using the partition input method (PIM).

3.2.2 Partition Input Method: Developing an Algorithm to Handle Big Data

We flatten each temperature data array for every winter season to form a column vector of 440,856,000 rows. All 28 Winter seasons are stacked column-wise to form the space-time data matrix:

$$X = [T_{i,y}]_{N \times M}.$$

where $T_{i,y}$ is the temperature in the grid box i and year y , N is the total number of grid boxes, 440,856,000, and M is the total number of seasons, 28. Each grid box covers a specific volume on Earth defined by latitude, longitude, and depth. For our study, they are equivalent entries, but specify the location. For example, the first entry $i = 1$ is the grid

box at depth 0.49 meters, latitude 80°S, and longitude 0°. An example of the full space-time data matrix can be seen in Eq. (3.3).

$$\mathbf{T}_{N \times M} = \begin{matrix} & \textcolor{red}{y_{1994}} & \textcolor{red}{y_{1995}} & \dots & \textcolor{red}{y_{2021}} \\ \textcolor{blue}{i_1} & t_{1,1994} & t_{1,1995} & \dots & t_{1,2021} \\ \textcolor{blue}{i_2} & t_{2,1994} & t_{2,1995} & \dots & t_{2,2021} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \textcolor{blue}{i_N} & t_{N,1994} & t_{N,1995} & \dots & t_{N,2021} \end{matrix} \quad (3.3)$$

440,856,000 × 28

Climatology is computed as an average across the entire period from the space-time data matrix. Anomalies are calculated by subtracting the temperatures $T_{i,y}$ from the climatology (Eq. (3.4)).

$$A = a_{i,y} = T_{i,y} - \frac{1}{M} \sum_{y=1994}^{2021} T_{i,y} \quad (3.4)$$

To adjust for a 3D computation, each anomaly value is weighted according to its spatial position, reflecting its geometric size. In other words, anomalies are multiplied by a spatial weight. The components of the spatial weight are twofold:

- (i) Horizontal weight: each latitude-longitude is centered at a grid box that will change in size depending on its latitude (ϕ). As a result, equatorial boxes will be much larger than polar boxes.

$$\sqrt{\cos \phi_i} \quad (3.5)$$

- (ii) Vertical weight: each depth layer has a unique thickness (Δd) affecting the volume of the grid box. The depth thickness equals the difference between the current depth value (bottom of the grid box) and the previous depth value (top of the grid box).

$$\sqrt{\Delta d_i} \quad (3.6)$$

All calculations that follow will require using the volume-weighted anomaly matrix. For simplicity, we will denote the volume-weighted anomalies matrix A and refer to it as the anomaly matrix.

3.2.2.1 Two Conventional Methods to Compute EOFs

EOFs in climate science were traditionally computed from the eigenvectors of a spatial covariance matrix of climate anomaly data, such as in the original EOF paper by [25], and the milestone papers by Gerald R. North and his colleagues [26,27]. This eigenvalue problem is defined as follows:

$$C_S u_k = \lambda_k u_k, \quad (3.7)$$

where u_k is the k th EOF, λ_k is the variance corresponding to this EOF, and C_S is the spatial covariance matrix equal to:

$$C_S = AA^T. \quad (3.8)$$

Here A^T is the transpose of the anomaly matrix A . Usually, N is much larger than M , because the number of entries is much larger than the number of time steps. However, exceptions can occur, e.g., analyzing the observed space-time data for a small number of stations.

In formal statistics terms, the unbiased estimate of the spatial covariance matrix should be divided by the length of time minus one (i.e., $M - 1$) [28]. The difference in eigenvalues, or variances, between the rigorous statistics definition of covariance and that presented here is a matter of a constant multiplier. Since many signal-oriented studies care only about the percentage variance explained by each mode, the difference in the constant multiplier is gone. Thus, in this calculation, we choose not to divide the covariance matrix by the length of time minus one.

A modern way of computing EOFs is via singular value decomposition (SVD) [28, 29]. The SVD approach is a space-time decomposition of a space-time anomaly data matrix, expressed as follows:

$$A = UDV^T, \quad (3.9)$$

where U is the $N \times M$ spatial matrix whose columns are the EOFs, V is the $M \times M$ temporal matrix whose columns are the principal components (PCs), and D is the diagonal matrix with diagonal elements:

$$d_k = \sqrt{\lambda_k}, \quad k = 1, 2, \dots, M. \quad (3.10)$$

When the size of the space-time data matrix A is more than 50% of the current computer memory, neither of the above methods will work, and the EOFs cannot be computed using either of these two conventional methods. This is due in part to both calculations being based on the dimensions of A . The spatial covariance method requires us to store an $N \times N$ matrix, while SVD requires storage of all $U_{N \times M}$, $D_{M \times M}$, and $V_{M \times M}$ matrices in addition to A . We can partially solve this issue by using covariance in time, which would only require storing an $M \times M$ covariance matrix:

$$C_T = A^T A. \quad (3.11)$$

When $M \ll N$, as is our case, solving the eigenvalue problem for C_T is computationally more feasible than for C_S [42]. Directly computing C_T , however, still presents a memory challenge, as the matrix multiplication would require RAM for both A and its transpose, effectively doubling the memory requirement. Instead, we can partition A , perform multiplication operations on the smaller sub-matrices, and compute C_T as a superposition of the resulting multiplications.

3.2.2.2 Partition Input Method on a Space-Time Data Matrix

The PIM allows you to partition a big space-time data matrix into many smaller partitioned matrices, called block matrices or sub-matrices. All sub-matrices are constrained to use only a predetermined percentage of the computer's memory. In our application, PIM partitions the space-time anomaly matrix $A_{N \times M}$, where $M \ll N$ (rows greatly exceed columns), which has a total storage requirement of 49 GB (Eq. (3.2)). The first sub-matrix A_1 from A would look like:

$$A_{N \times M} = \begin{pmatrix} \begin{matrix} a_{1,1994} & a_{1,1995} & \dots & a_{1,2021} \\ \dots & \dots & \dots & \dots \\ a_{p,1994} & a_{p,1995} & \dots & a_{p,2021} \end{matrix} \\ a_{p+1,1994} & a_{p+1,1995} & \dots & a_{p+1,2021} \\ \dots & \dots & \dots & \dots \\ a_{2p,1994} & a_{2p,1995} & \dots & a_{2p,2021} \\ \dots & \dots & \dots & \dots \\ a_{N,1994} & a_{N,1995} & \dots & a_{N,2021} \end{pmatrix}, \quad (3.12)$$

with

$$\Rightarrow A_1 = \begin{pmatrix} a_{1,1994} & a_{1,1995} & \dots & a_{1,2021} \\ a_{2,1994} & a_{2,1995} & \dots & a_{2,2021} \\ \dots & \dots & \dots & \dots \\ a_{p,1994} & a_{p,1995} & \dots & a_{p,2021} \end{pmatrix}_{p \times M}, \quad (3.13)$$

where p is the chunk size, or the total number of grid points/entries read in for each year. The sub-matrices follow a sequential pattern: the first partition (A_1) will start at row 1 and end at row p , the second partition (A_2) will start at row $p+1$ and end at row $2p$ (Eq. (3.14)), the third (A_3) covers rows $2p+1$ to $3p$, and so on until all N rows are accounted for.

$$A_{N \times M} = \begin{pmatrix} a_{1,1994} & a_{1,1995} & \dots & a_{1,2021} \\ \dots & \dots & \dots & \dots \\ a_{p,1994} & a_{p,1995} & \dots & a_{p,2021} \\ a_{p+1,1994} & a_{p+1,1995} & \dots & a_{p+1,2021} \\ \dots & \dots & \dots & \dots \\ a_{2p,1994} & a_{2p,1995} & \dots & a_{2p,2021} \\ \dots & \dots & \dots & \dots \\ a_{N,1994} & a_{N,1995} & \dots & a_{N,2021} \end{pmatrix} \quad (3.14)$$

In total, there will be n sub-matrices. The complete matrix A can be represented as a vertical stack of these sub-matrices:

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{n-1} \\ A_n \end{bmatrix}_{N \times M} \quad (3.15)$$

As mentioned previously, each sub-matrix A_j ($j = 1, 2, \dots, n$) is sized to consume only a portion of available computer memory, ensuring sufficient resources remain for matrix multiplication operations during covariance and 3D EOF calculations. While various libraries exist to handle partitioning in large-scale array operations, such as *PyTorch* for matrix multiplications, we develop a custom algorithm tailored to our specific needs. For climate data in NetCDF format, we leverage the *NetCDF4* library, which provides native support for chunked reading and writing through specified indices. This approach effectively partitions for the covariance and EOF calculations while maintaining compatibility with standard climate data formats. The following subsections detail the PIM implementation for both calculations and their optimal parameters.

3.2.2.3 Temporal Covariance and the PIM

Using the PIM, we can compute temporal covariance by partitioning A , performing multiplication operations on the smaller sub-matrices, and computing C_T as a superposition of the resulting multiplications. The PIM implementation of this multiplication is as follows:

$$C_T = A_1^T A_1 + A_2^T A_2 + \cdots + A_{n-1}^T A_{n-1} + A_n^T A_n. \quad (3.16)$$

Here, $A_1, A_2, \dots, A_{n-1}, A_n$ are the sub-matrices of the anomaly matrix A in Eq. (3.15). We can prove that Eqs. (3.11) and (3.16) are equivalent using Eq. (3.15):

$$\begin{aligned} C_T &= A^T A \\ &= [A_1^T A_2^T \cdots A_n^T]_{M \times N} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{n-1} \\ A_n \end{bmatrix}_{N \times M} \\ &= C_T = A_1^T A_1 + A_2^T A_2 + \cdots + A_{n-1}^T A_{n-1} + A_n^T A_n. \end{aligned} \quad (3.17)$$

When computing the chunk size p for the covariance equation, we can exploit consistencies within the data to save memory. Let's first redefine the sub-matrices so they are taken from

A^T . The first sub-matrix would be:

$$A_{M \times N}^T = \begin{pmatrix} a_{1994,1} & a_{1994,2} & \dots & \dots & a_{1994,N} \\ a_{1995,1} & a_{1995,2} & \dots & \dots & a_{1995,N} \\ \dots & \dots & \dots & \dots & \dots \\ a_{p,1} & a_{p,2} & \dots & \dots & a_{p,N} \\ \dots & \dots & \dots & \dots & \dots \\ a_{2021,1} & a_{2021,2} & \dots & \dots & a_{2021,N} \end{pmatrix}. \quad (3.18)$$

Defining sub-matrices in this way allows us to think of them as a series of anomaly years read in at once. For example, if our chunk size is $p = 3$, then we would be reading 3 anomaly years at a time to construct a sub-matrix. We exploit the consistent positions of NaN values for each year by indiscriminately removing them while anomalies are read in; these non-values represent land and are not needed for this calculation. After the land removal, the total amount of entries for a single year, $N = 440,856,000$, decreases to $N' = 257,398,868$. This reduces memory storage to less than 1 GB a year, allowing us to read in more anomaly years at a time.

The chunk size is computed by finding the maximum integer divisor of half the available memory divided by the memory needed for each year (Eq. (3.19)). The available RAM is halved because 2 sub-matrices need to be stored at once to compute temporal covariance.

$$\begin{aligned} p &= \frac{\text{Available RAM}}{2 \times \text{Total bytes per year}} \times \frac{1}{\text{smallest prime factor}} \\ &= \frac{\text{Available RAM}}{2,059,190,944 \text{ bytes}} \times \frac{1}{\text{smallest prime factor}} \end{aligned} \quad (3.19)$$

The covariance PIM was run for different p values to test the run time outlined in Table 3.2. The total GB required must increase by a factor of ≈ 2 GB because the calculation requires two sub-matrices be stored at once (i.e., $2 \times N' = 514,797,736$ entries ≈ 2 GB).

The maximum chunk size using 16 GB of RAM is 7 years, as 8 would require just above 16 GB.

There is a large jump in run time, around 30 minutes, between 1 and 2 years, as seen in Fig. 3.1. This is likely due to fewer calls to the reading, writing, and multiplication functions. After this jump, the run time decreases by an average of 7 minutes until it plateaus at $p = 5$ years with 33.33 minutes, and reaches a minimum of 33.03 minutes at $p = 6$ years (12.36 GB of RAM). The gradual decline between 4.12 and 10.3 GB is due to the constrained 2 GB memory increase for each additional year in p , which in turn limits the reduction in runtime. The maximum chunk size of 7 years, or 14.41 GB of RAM, required a longer run time (34.1 minutes) than that of the previous chunk size step. This slight increase could be a result of the computer starting to reach its memory limit at 14.41 GB.

Table 3.2: Run time for various chunk sizes in years, total GB for the PIM computation of the covariance matrix.

Chunk size (yrs)	Total GB required	% RAM	Run time (min)
1	2.06	12.5%	87.2
2	4.12	25.75%	55.8
3	6.18	38.6/%	46.8
4	8.24	51.5%	41.01
5	10.3	64.38%	33.33
6	12.36	77.25%	33.03
7	14.41	90.06%	34.1

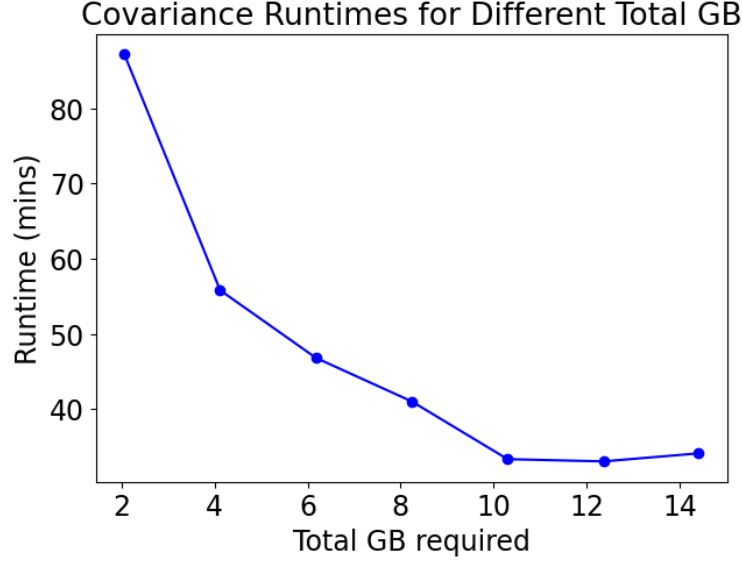


Figure 3.1: The run time for the PIM covariance matrix calculation for different total RAM gigabytes.

3.2.2.4 EOFs and the PIM

After n times of input of the partitioned matrices and their multiplications for the n sub-covariance matrices $A_j^T A_j (j = 1, 2, \dots, n)$, the temporal covariance matrix C_T is in the memory. We compute the eigenvalue problem of this small matrix C_T , a 28×28 square matrix, for the GLORYS data as:

$$C_T v_k = \lambda_k v_k, \quad k = 1, 2, \dots, M. \quad (3.20)$$

Here, the unit eigenvectors v_k are the PCs, and their spatial counterparts are the EOFs [28, 29]. Each EOF u_k can be computed by the projection of A onto v_k and normalizing:

$$u_k = \frac{A v_k}{|A v_k|}, \quad (3.21)$$

where $|A v_k|$ is the Euclidean norm of the vector $A v_k$.

Due to the memory limit, the matrix multiplication for the projection Av_k cannot be computed directly. We use the PIM again in the following way:

$$A\vec{v}_k = \begin{bmatrix} A_1 v_k \\ A_2 v_k \\ \vdots \\ A_{n-1} v_k \\ A_n v_k \end{bmatrix}_{N \times 1}, \quad (3.22)$$

where

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_{n-1} \\ A_n \end{bmatrix}_{N \times M} \quad \text{and} \quad \vec{v}_k = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_M \end{bmatrix}_{M \times 1}. \quad (3.23)$$

These sub-matrices are the same as in Eq. (3.15). To avoid any redundancies, our algorithm identifies the location of land/NaN values and only reads and writes to points with information. Therefore, in practice, our A matrix is $N' \times M$, reducing run time since there are fewer points to read and write to. The multiplication of $A\vec{v}_k$ results in an N -dimensional vector that will be reshaped as a depth-latitude-longitude array and saved in separate files according to their ‘mode’ k . Later, these files are individually read in, normalized, and overwritten as EOFs defined by Eq. (3.21). Though reading and writing are time-consuming, saving EOFs at each step will free computer memory for the calculation.

The resulting SVD matrices of A can be constructed as follows:

$$\begin{aligned} U &= [u_1 u_2 \cdots u_n]_{N \times M}, \\ V &= [v_1 v_2 \cdots v_n]_{M \times M}, \\ D &= \text{diag}(\sqrt{\lambda_k}, k = 1, 2, \cdots, M)_{M \times M}. \end{aligned}$$

Our EOF PIM algorithm requires us to load an EOF file (1.75 GB) to read and write to, along with one sub-matrix during the calculation. Therefore, to calculate our chunk size, we will need to subtract the storage of one EOF file from the total memory we want to use (e.g., $16 - 1.75 \text{ GB} = 14.25 \text{ GB}$). We define p as the total number of entries read in for each year:

$$p = (\text{Available RAM} - (1.75 \times 10^9 \text{ bytes})) \times \frac{1}{4 \text{ bytes}} \times \frac{1}{28 \text{ years}} \quad (3.24)$$

The EOF PIM was utilized for six different p values, outlined in Table 3.3, ranging from 2 to 15 GB of computer memory. As expected, our run time decreased as the chunk size increased (Fig. 3.2). There is a considerable reduction in run time of 31.4 hours from 2 to 4 GB. The number of entries we can read in is flexible and reflects the exact amount of computer memory used. Unlike the covariance PIM, the EOF PIM chunk size is not limited by how much it can increase. Our increase from 2 to 4 GB translates to a drastic increase in entry sizes of 2,232,142 to 20,089,285, explaining the jump in run time. After 4 GB, the run time decreases slowly, leveling out at 8 GB with 4.6 hours. The minimum run time was at 15 GB and totaled 2.5 hours, only 12 minutes faster than the previous size of 12 GB. The exact cause of this minimal decrease in run time could be due to the computer reaching its memory limit, causing a slowdown due to various computer environment factors (e.g., use of virtual memory, background processes, maintenance, etc.).

Table 3.3: Run time for the PIM computation of EOFs versus their total GB.

Chunk size (no. entries)	Total GB required	% RAM	Run time (hrs)
2,232,142	2	12.5%	55
20,089,285	4	25%	17.2
37,946,428	6	37.5%	7.9
55,803,571	8	50%	4.6
91,517,857	12	75%	2.7
118,303,571	15	93.75%	2.5

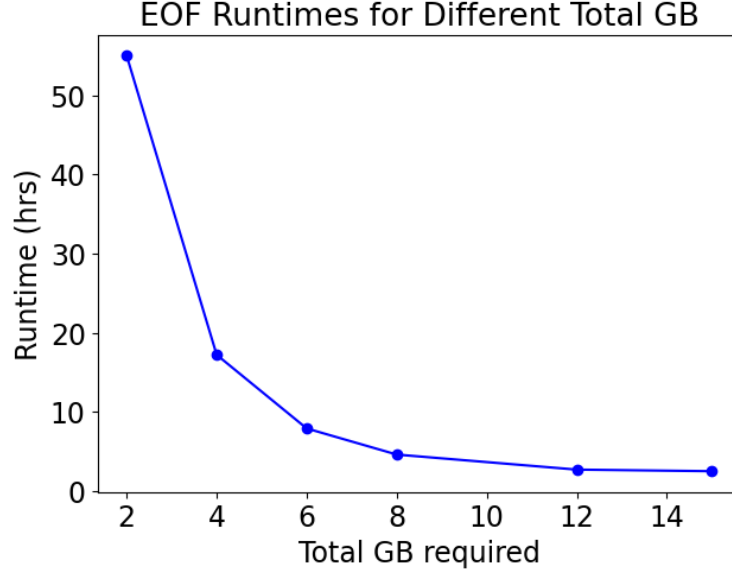


Figure 3.2: The run time for the PIM EOF calculation for different total RAM gigabytes.

3.3 Results

3.3.1 Creating 3D Visualization for EOFs

We visualize our results to contain latitude-longitude-depth elements for a meaningful spatial domain to provide a complete view of across-layer variability. 3D visualization is ideal for displaying the complex spatial structures of different ocean dynamics in the 3D EOFs. However, creating these figures requires careful consideration of memory constraints.

Plotting a single EOF requires loading a 1.7 GB EOF array into memory. Each 3D plot also requires latitude, longitude, and depth arrays with dimensions matching the plotted data.

These dimension variables will vary in size depending on the plotted spatial domain. For example, if we are plotting the entire globe, the dimension variables are the same size as the EOF array (1.7 GB), requiring an additional 5.1 GB. In total, we would need 6.8 GB of memory for a 3D global plot, our most memory-intensive case.

Most of the ocean’s variability is found in the upper 600 meters. In the GLORYS data, this is the first 32 layers (see Table 3.1). Therefore, in practice, our dimension variables are reduced to at most 1.13 GB each, totaling 5.1 GB including the EOF. Shrinking our spatial domain will only decrease the memory needed for our dimension variables further. In general, we can compute how much memory is needed for data and dimension variables as:

$$1.7 \text{ GB} + 3 \text{ variables} \times (\text{latitude} \times \text{longitude} \times \text{depth}) \times 4 \text{ Bytes.} \quad (3.25)$$

The remaining data manipulation involves indexing parts of the EOF array already in memory and will not consume additional RAM. The rest of the available RAM can be used for creating each figure, which will also vary. We use *matplotlib*, a Python library for creating visualizations, and its module called *mplot3d* for all 3D figures. Plotting a single depth layer for the global domain would require 740 MB; that is, one layer of Fig. 3.3. All three layers require around 2 GB and 28 seconds to run, tripling in both memory and execution time. The global multi-layer plot is the most memory-intensive figure in this section; caution should be used when handling large data and systems with limited RAM.

For our 3D figures, we use EOF 1, which captures the traditional El Niño-Southern Oscillation (ENSO). We reference the well-known zonal dipole structure of El Niño previously researched in 2D [7, 14], to define El Niño’s 3D structure. Our 3D methods capture the same characteristic warm eastern and cold western tropical Pacific subsurface signature of El Niño’s zonal dipole. Each visualization displays the dipole’s 3D structure and is helpful in quantifying the volume of both halves while highlighting various features of El Niño.

EOF 1

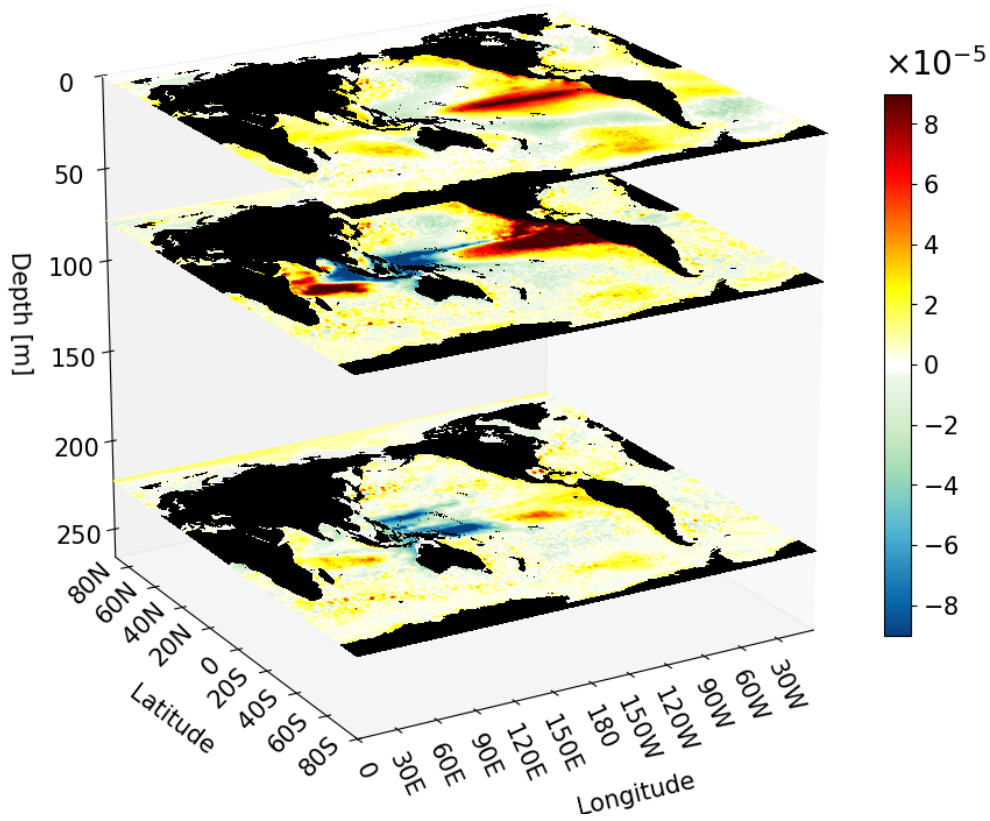


Figure 3.3: EOF 1 plotted for the entire globe at three depth layers (surface, 77 meters, and 222 meters). These depths were chosen to show the dynamics of El Niño’s zonal dipole, and in accordance with GLORYS depths (see Table 3.1).

Using a global domain will show the effect of El Niño on other ocean basins. The layer-by-layer figure (Fig. 3.3) displays the eastern Pacific surface warming, the dipole structure at 77 meter depth, and the signal’s absence beyond 200 meters. The Indian Ocean displays a positive dipole evolving alongside El Niño, with clear signatures at 77 meters. This global perspective emphasizes basin-wide dynamics and supports theorized relationships between different oceanic phenomena. For the case of EOF 1, our global visualization demonstrates the relationship between El Niño and a positive Indian Ocean Dipole (pIOD) [5, 80].

Taking zonal and meridional cross-sections of EOF 1 emphasizes the subsurface characteristics of El Niño (Fig. 3.4). The zonal portion starts at 160°E to capture the westernmost

El Niño index, Niño 4 ($5^{\circ}\text{S} - 5^{\circ}\text{N}$, $160^{\circ}\text{E} - 150^{\circ}\text{W}$), and ends at 60°W just west of Niño 1+2 ($10^{\circ}\text{S} - 0$, $90^{\circ} - 80^{\circ}\text{W}$). The meridional cross-section starts at the equator, where some of the strongest warming of El Niño lies, and ends at 60°N to encompass some of its northern teleconnections. These cuts reveal notable subsurface cooling lying beneath the eastern Pacific warming, clearly displaying the characteristic thermocline displacement of El Niño described in the next section. The regional surface plot provides an intuitive spatial context for the cross-section’s locations while revealing El Niño’s possible effects along the North American coast. Such effects include enhanced warming near Papagayo, the Isthmus of Tehuantepec, and Baja California to the U.S. West Coast [5, 15, 81, 82]. This visualization requires only 124 MB per dimension variable and 100 MB for figure creation. Its light memory demand is due to the downsized spatial domain.

A zonal cross-section is particularly effective in displaying the vertical extent of a feature. Extending the longitude region (160°E to the eastern boundary) with multiple zonal cuts ($10^{\circ}\text{S} - 10^{\circ}\text{N}$) quantifies EOF 1’s precise three-dimensional structure (Fig. 3.5). The same longitudes were used in this figure as in the previous one to capture all El Niño indices. Latitudes were selected to include the southernmost Niño index (Niño 1+2) at 10°S , with additional latitudes at 5° increments to provide comprehensive coverage.

The equatorial cut in Fig. 3.5 has strong eastern warming across multiple depths, while other slices show weaker surface warming reaching maximum intensities at deeper layers. Additionally, the zonal cuts show that north of 10°S , the western tropical Pacific experiences its strongest negative signal below 50 meters and reaches deeper than the eastern Pacific warming. The depth location of these signals coincides with the location of the thermocline, providing vital information on the vertical motion of the thermocline during El Niño. The cross-section figure requires the least amount of memory to create, less than 1 MB, with dimension variables requiring only 48 MB each, making it the most computationally efficient approach while providing maximum structural detail.

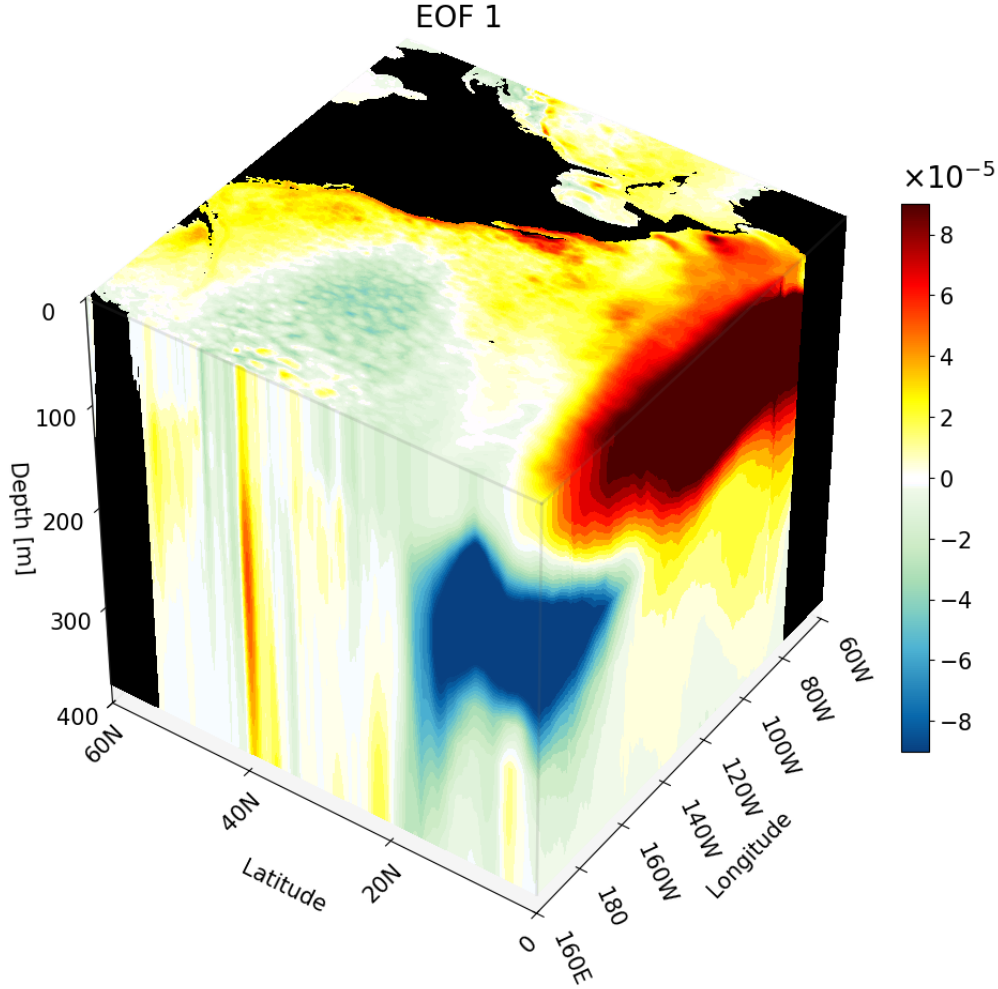


Figure 3.4: Three-dimensional visualization of EOF 1 with cross sections taken at the equator to 60°N and 160°E-60°W. The depth domain is limited to the upper 400 meters as it provides a clear display of the ENSO dynamical pattern. The right-side view cross-section shows the strong ENSO zonal dipole. The front meridional view displays the Northern Hemisphere extent of the western subsurface cold pole in the 3D ENSO signature. The top view shows ENSO’s surface pattern, commonly used in the climate research community.

Though the exact depth, longitudes, and latitudes of EOF 1’s pattern were not thoroughly described in this section, they can be defined using the 3D visualizations. Each figure emphasizes a different aspect of 3D ocean variability. A layer-by-layer figure shows global variability across multiple layers. The 3D cube with a surface, meridional cut, and zonal cut gives a general outline for 3D structure along with regional expressions. Multiple zonal cuts provide a detailed description of 3D structure.

Our high-resolution results could lead to issues in creating 3D figures if the user isn't mindful of the domain they are visualizing. Using tools that will visually scan and crop multiple cross-sections, such as those presented in Chapter 5, help identify appropriate domains for 3D plotting. Having an understanding of where most of the ocean's variability will be located can significantly cut the memory needed for dimension variables, leaving the rest of the RAM available for visualization.

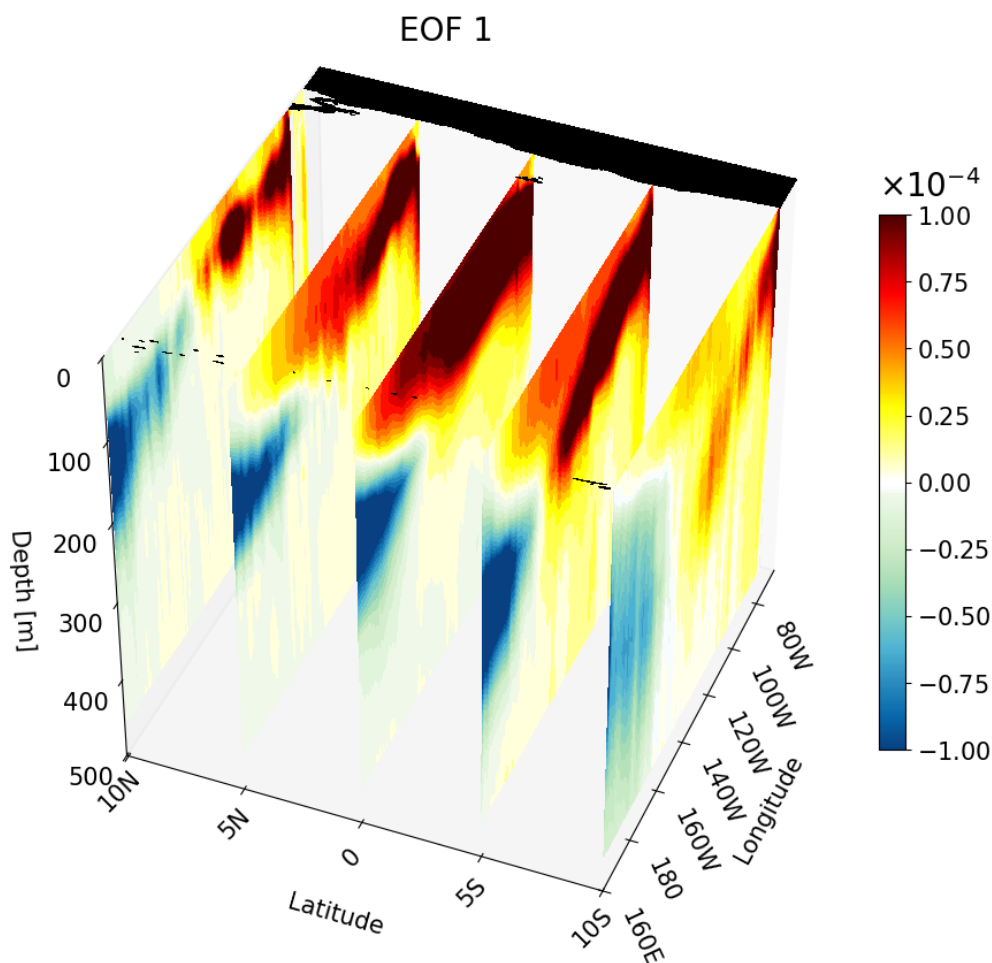


Figure 3.5: Vertical cross-sectional view of EOF 1 at 10°S, 5°S, 0°, 5°N, and 10°N in the Pacific from 160°E - 80°W. Depths go from the surface to 454 meters. These cross-sections also show the signal of ENSO's zonal dipole pattern.

3.3.2 Global 3D EOFs Interpretation

The first EOF with 18% variability is the canonical El Niño-Southern Oscillation (ENSO). The maxima in PC 1 describe the strong El Niño events in 1998 and 2016 (Fig. 3.6(a)), while its minima capture the La Niña years in 1999, 2000, 2008, and 2011. The spatial pattern shows the El Niño dipole signature with a warm eastern and cool western tropical Pacific throughout the first 222 meters (left column of Fig. 3.6). This can also be seen in the two cross-sections of Figs. 3.7(a) and (d), as the heightened cold/warm variability in the first 250 meters starting at 120°E - 60°W.

In Fig. 3.7(d), most of the variability is found within the thermocline, indicated by the close isotherms. The 20°C isotherm (blue line) delineates the average position of this transition zone. The maximum warming lies between 25 - 120 meters across the longitude band ~140° - 80°W. The strongest cooling occurs at 50 - 250 meters between 120°E and just east of the date line. This asymmetric pattern, with deeper cooling anomalies in the west compared to the east, reflects the basin's characteristic upward west-to-east thermocline slope. The pronounced variability results from the vertical displacement of the thermocline; shoaling in the western and deepening in the eastern tropical Pacific generate large-scale anomalies.

The first EOF additionally captures the pIOD, characterized by a warm western and cool eastern tropical Indian Ocean. At the surface, this mode exhibits widespread warming with cooling from the Australian coast to ~80°E (Fig. 3.6(a)). However, the subsurface contains a clear dipole pattern: strong eastern tropical cooling at 50 - 150 meters and a maximum western warming in the same layers (Fig. 3.7(d)). The absence of surface dipole expression reflects the pIOD decay during winter months, preserving only the subsurface signature [18].

The second EOF (11% variance), displays a global warming signal dominating the spatial pattern (Fig. 3.6 (middle column)). The associated time series demonstrates notable linearity, evidenced by the strong positive linear trend in PC 2 ($r = 0.90$). The warming carries

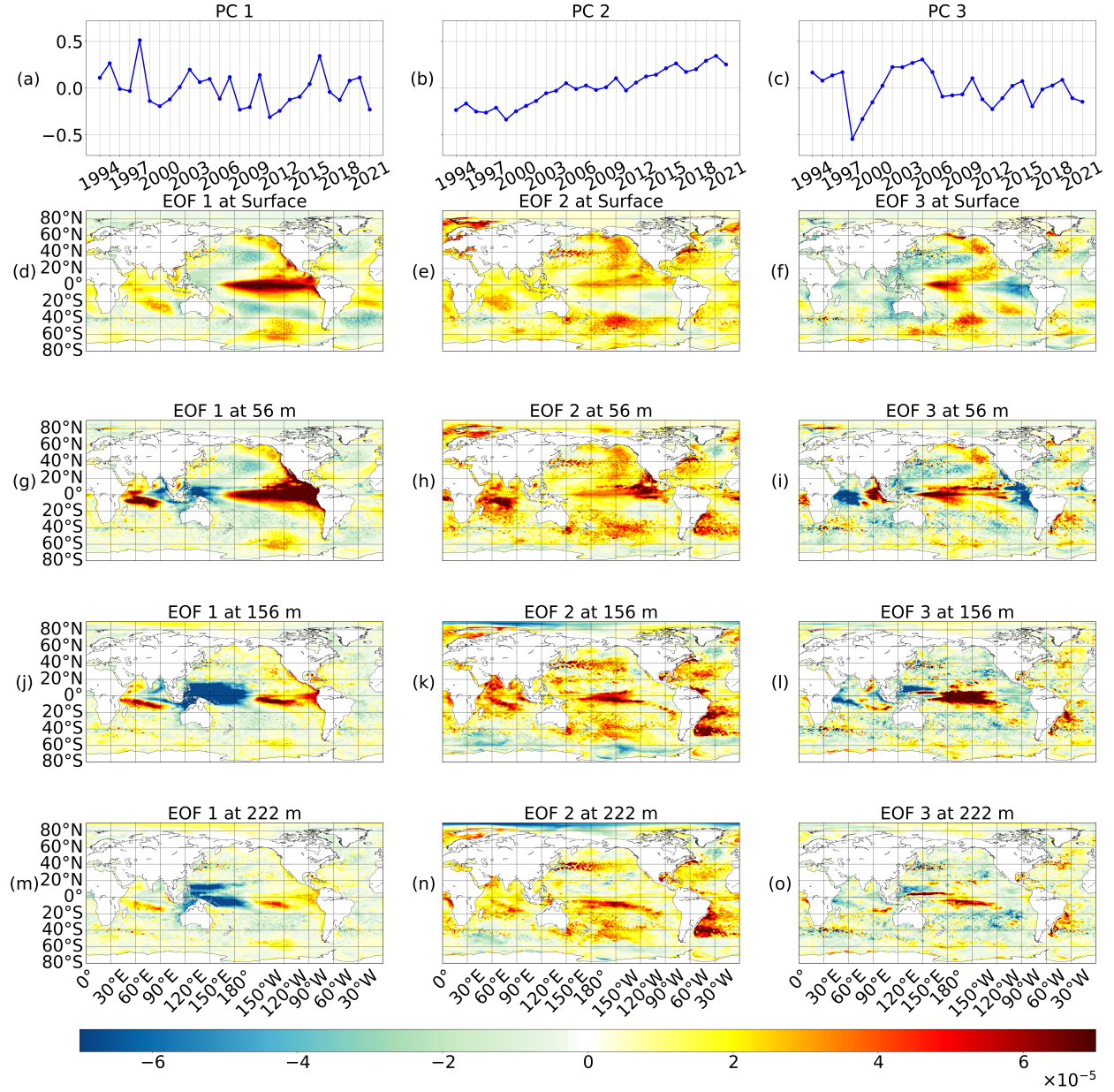


Figure 3.6: Visualization of the first three EOFs and their PCs: 3D EOF 1 and PC 1 (left column), 3D EOF 2 and PC 2 (middle column), and 3D EOF 3 and PC 3 (right column) for the global ocean. EOFs were chosen to be displayed at four different depth layers (surface, 56 meters, 156 meters, 222 meters). These depths were chosen for EOF 1 to show the traditional El Niño, EOF 2 to show warming and central Pacific El Niño, and EOF 3 to show the meridional modes.

into considerable depths, maintaining strength until 750 meters before weakening and ending around 1,000 meters (Fig. 3.7(b)).

Despite the overwhelming global warming signature, EOF 2 contains Central Pacific (CP) El Niño variability. The presence of CP El Niño is not immediately obvious from the time series. A correlation analysis between PC 2 and standard CP El Niño indices shows no significant relationship. For this analysis, we used two CP El Niño indices: the El Niño Modoki index (EMI) computed according to [6] and the Trans-Niño index (TNI) based on the definition in [83]. Both are computed using GLORYS sea surface temperature anomalies (SSTA), with the TNI using normalized SSTA. A complete description of the calculation can be found in Appendix A.

Nevertheless, the spatial patterns clearly reveal the CP El Niño influence in EOF 2. The CP El Niño signal appears as surface warming around the date line in the equatorial Pacific that extends north towards Baja California, coupled with cooling at $20^{\circ}\text{S} - 0^{\circ}$ from 130°W to the South American coast (Fig. 3.6(a)). The equatorial cross-section further reveals intensified variability in the upper 200 meters at $180^{\circ} - 70^{\circ}\text{W}$ (Fig. 3.7(b)). This indicates thermocline displacement characteristic of CP El Niño, which shoals in the east and deepens in the Central Equatorial Pacific [84, 85].

EOF 2 exhibits distinct regional warming at the surface, most prominently in the North Pacific ($\sim 20^{\circ} - 60^{\circ}\text{N}$, $160^{\circ} - 130^{\circ}\text{W}$), Southern Ocean ($60^{\circ} - \sim 35^{\circ}\text{S}$, $180^{\circ} - \sim 100^{\circ}\text{W}$), and southwestern Indian Ocean ($30^{\circ} - 10^{\circ}\text{S}$, Madagascar coast to 80°E). At a depth of 56 meters, these regions strengthen considerably along with a new warming that emerges in the Atlantic along South America (Fig. 3.6). These localized warming signals correspond to marine heatwave (MHW) signatures, whose inclusion in this EOF suggests coupling between MHW, global warming, and CP El Niño variability. Both the warming trend and CP El Niño influence MHW frequency and intensity in these specific regions [11, 77], explaining the presence of these regional patterns embedded within the broader signal.

The third EOF (5% variance) captures the Meridional Mode (MM) across multiple ocean basins. In the Pacific, both northern and southern components are evident: the North

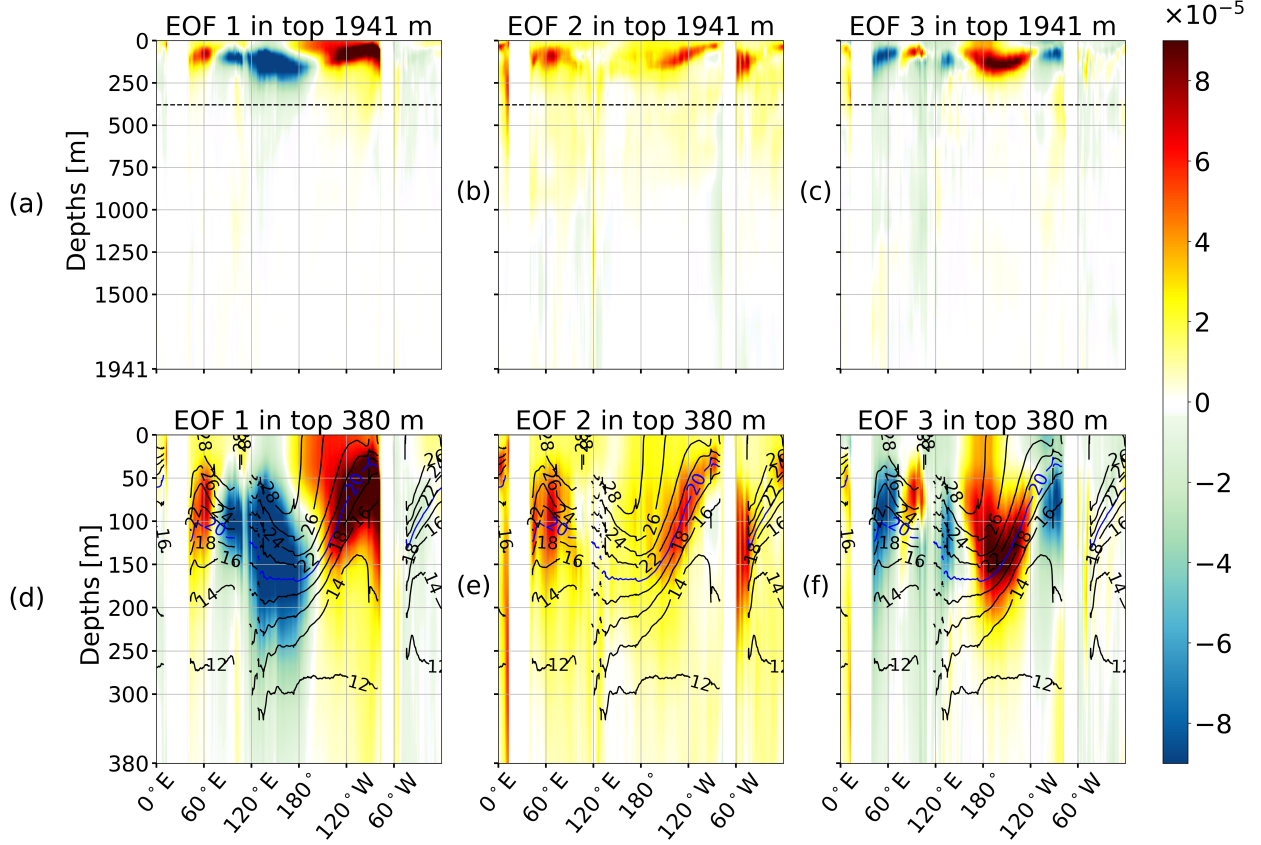


Figure 3.7: Depth-zonal view of the first three EOFs over the tropical region. The values displayed in color are taken as the average of 3D EOFs from 5°S to 5°N. The three panels in the top row, i.e., (a), (b), and (c), are from the surface to 1,941 meters, showing that strong temperature variability occurs in the top 380 meters. The three panels in the bottom row, i.e., (d), (e), and (f), are zoomed in to the layers in the top 380 meters, indicated by the black dotted line in the top row. Black contours are isotherms from boreal winter (DJF) climatology. The blue contour is the 20°C isotherm to represent the thermocline. Variability in the depths between 2,000 - 5,500 meters is very small and not displayed in this figure.

Pacific Meridional Mode (NPM) manifests as surface warming extending from the western equatorial Pacific toward the coast of North America, with large variability around the date line. The cooling in the eastern tropical Pacific (20°S - 15°N, 150° - 80°W) completes the NPM characteristic meridional gradient across the mean intertropical convergence zone (ITCZ) latitude at 8°N (mean ITCZ defined by [9]).

The NPM in EOF 3 displays central tropical Pacific surface warming between 10°S - 10°N and 150°E - 150°W (Fig. 3.6(f)). As shown in the cross-section of Fig. 3.7(f), the central tropical Pacific warming remains consistent in the upper 50 meters. From 50 - 100

meters, this warming intensifies with depth and extends eastward to lie between 150°E - 120°W . Maximum warming occurs between 100 - 175 meters along the thermocline, where it narrows in longitude to lie between the date line and 120°W . Similar to EOFs 1 and 2, vertical thermocline displacement generates large anomalies through deepening in the central tropical Pacific. However, thermocline displacement alone does not fully explain the subsurface spatial structure. Previous studies [10, 16] demonstrate that this subsurface warming pattern is a result of the trade wind charging (TWC) of heat content driven by surface wind stress curl anomalies.

The South Pacific Meridional Mode (SPMM) manifests in EOF 3 as warming between 40° - 15°S , extending from the South American coast to 130°W . The warming reaches maximum intensity at the surface within 40° - 20°S and 120° - 90°W . This surface signal diminishes rapidly with depth, weakening substantially within the upper 50 meters (Figs. 3.6(f) and (i)). While portions of the warming penetrate into deeper layers (Fig. 3.6(o)), most of the signal begins weakening at 20 meters and dissipates between 40 - 50 meters (Fig. 3.8). The SPMM signal is confined to shallow waters due to the underlying subsurface processes that drive this mode. During boreal winter, shallow mixed layer depths in this region increase the SST sensitivity to latent heat flux anomalies, effectively generating a strong SPMM [86, 87]. These subsurface processes, in turn, dominate the variability observed in the shallow waters.

The Atlantic Meridional Mode (AMM), similar to the NPMM, has a surface gradient across the mean ITCZ at 5°N [88]. Northern tropical warming extends from the African coast to South America, while equatorial cooling centers primarily in the Gulf of Guinea. This is the weakest of all MMs, likely due to the spring peak timing rather than the winter period used in this analysis.

We validated PC 3 against established meridional mode indices using a modified approach due to our temperature-only dataset. Since the standard method, maximum covariance analysis (MCA), requires both SST and wind data [9], we computed all MM indices for

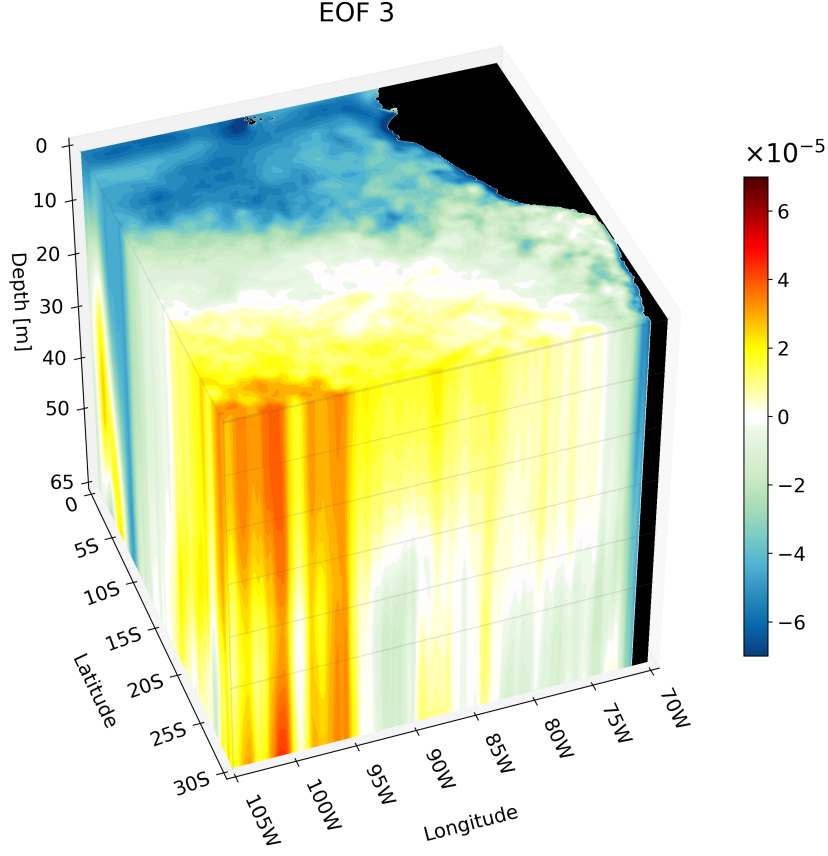


Figure 3.8: EOF 3 visualized for the SPMM at the surface and with a zonal-depth cut at 30°S and meridional cut at 105°W. Cross-sections were chosen to bisect the warming from the SPMM, and are from surface to 65 meters to interpret the subsurface structure of SPMM. Grey lines in the zonal cut are at every 10 meters to align with the depth axis.

GLORYS according to [4,10]: (i) remove the linear trend from anomalies, (ii) remove ENSO influence by linearly regressing out the first EOF of tropical Pacific SSTA (20°S - 20°N, 140°E - 90°W), and (iii) computing EOFs and PCs from regional SSTA fields defined by Table 3.4.

Table 3.4: Meridional mode region definitions to compute their indices. Regions are defined by [4].

Meridonal Mode (MM)	Latitude	Longitude
NPMM	20°S-30°N	175°E-95°W
SPMM	35°S-20°S	180°-70°W
AMM	0°-30°N	75°W-African Coastline

The Pacific meridional modes demonstrate statistically significant correlations with PC 3. The NPMM yields $r = 0.56$, while the SPMM shows stronger agreement at $r = 0.69$. This difference reflects the seasonal preferences of each PMM relative to our boreal winter analysis period: the NPMM peaks during spring, whereas the SPMM peaks during winter. The AMM has no significant correlation, consistent with its weaker expression in the spatial pattern. Additionally, PC 3 successfully captures the behaviors typical of PMM where both warm and cold ENSO events will project onto the negative meridional mode index [9]. El Niño years (1998 and 2016) and La Niña years (1999, 2000, 2008, and 2011) result in negative values for PC 3.

3.3.3 2D Versus 3D EOF Comparison

We computed EOFs layer-by-layer (2D) to compare with our 3D results. Our first 2D EOF displays the traditional ENSO pattern (Figs. 3.9 (left column)), consistent with 3D EOF 1. While 2D EOF 1 shows a stronger signal magnitude, the spatial pattern remains similar. This magnitude inflation likely stems from the calculation methodology, which considers variability for only a single layer at a time. Given these similarities in 2D and 3D EOF 1, we focus our analysis on the second and third EOFs, which display more notable differences.

The 2D EOF 2 preserves some of the warming from 3D EOF 2 while displaying additional cooling regions (Figs. 3.9 (middle column)). At the surface, eastern Pacific cooling spans the southern tropics from the eastern boundary to the date line between 25°S to the equator. The western tropical Pacific shows weak warming that extends toward Mexico and Baja California, resembling the NPMM. Meanwhile, South Pacific warming occurs from 120° - 90°W at 45° - 25°S, similar to the SPMM. Unlike 3D EOF 2, which lacks these PMM signatures, 2D EOF 2 captures the surface NPMM and SPMM features typically found in 3D EOF 3.

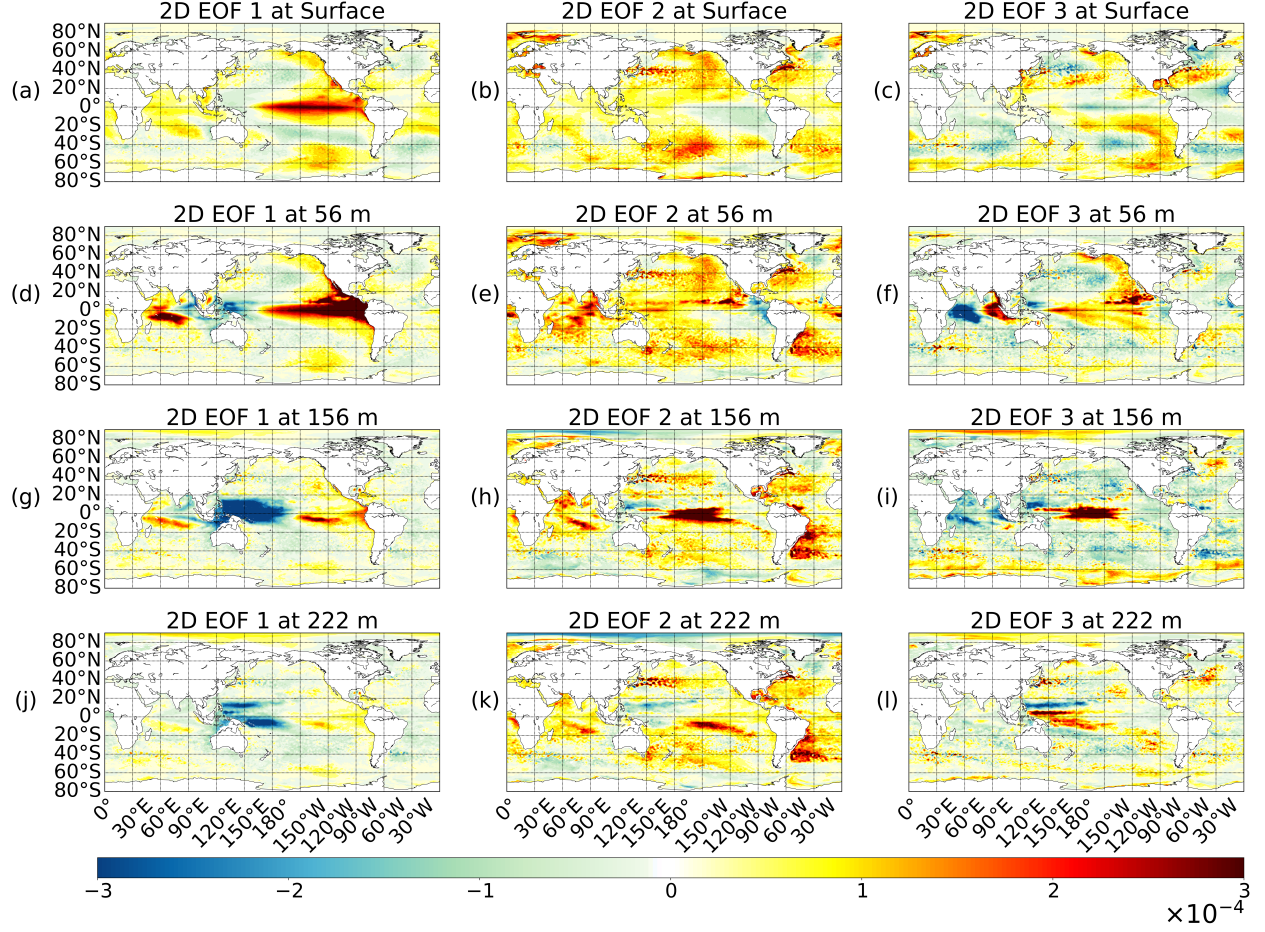


Figure 3.9: Same as Fig. 3.6 but for the 2D calculation. We exclude the time series as each layer outputs multiple PCs.

Also in alignment with 3D EOF 2, the surface of 2D EOF 2 contains warming associated with MHW in the North Pacific ($\sim 20^\circ - 60^\circ\text{N}$, $160^\circ - 130^\circ\text{W}$) and Southern Ocean ($60^\circ - \sim 35^\circ\text{S}$, $180^\circ - \sim 100^\circ\text{W}$). Although 3D EOF 2 captured a MHW at the surface of the southwestern Indian Ocean ($30^\circ - 10^\circ\text{S}$, Madagascar coast to 80°E), this feature is absent from the 2D EOF 2 surface. Instead, it appears as warming at 56 meters, alongside the South Atlantic MHW along the South American coast (Fig. 3.9(e)).

A subsurface analysis confirms the NPMM signature in 2D EOF 2. Between 100 and 175 meters, subsurface warming centered around 150°W (Figs. 3.10(b) and (e)) captures the TWC temperature variability associated with NPMM. Thus, the NPMM shifts to the second mode for our 2D results, where it is coupled with MHW and warming signals. Notably,

compared to the NPMM subsurface warming pattern in 3D EOF 3 (Fig. 3.7), the warming in 2D EOF 2 reaches further west (Fig. 3.10). Within the first 150 meters, NPMM warming in 3D EOF 3 reaches 150°E, while 2D EOF 2 extends well past this to 130°E. Compared to 2D EOF 2, the subsurface warming in 3D EOF 3 is more consistent with published NPMM subsurface variability [10, 16]. This discrepancy may result from the warming or MHW signals accompanying the NPMM in 2D EOF 2.

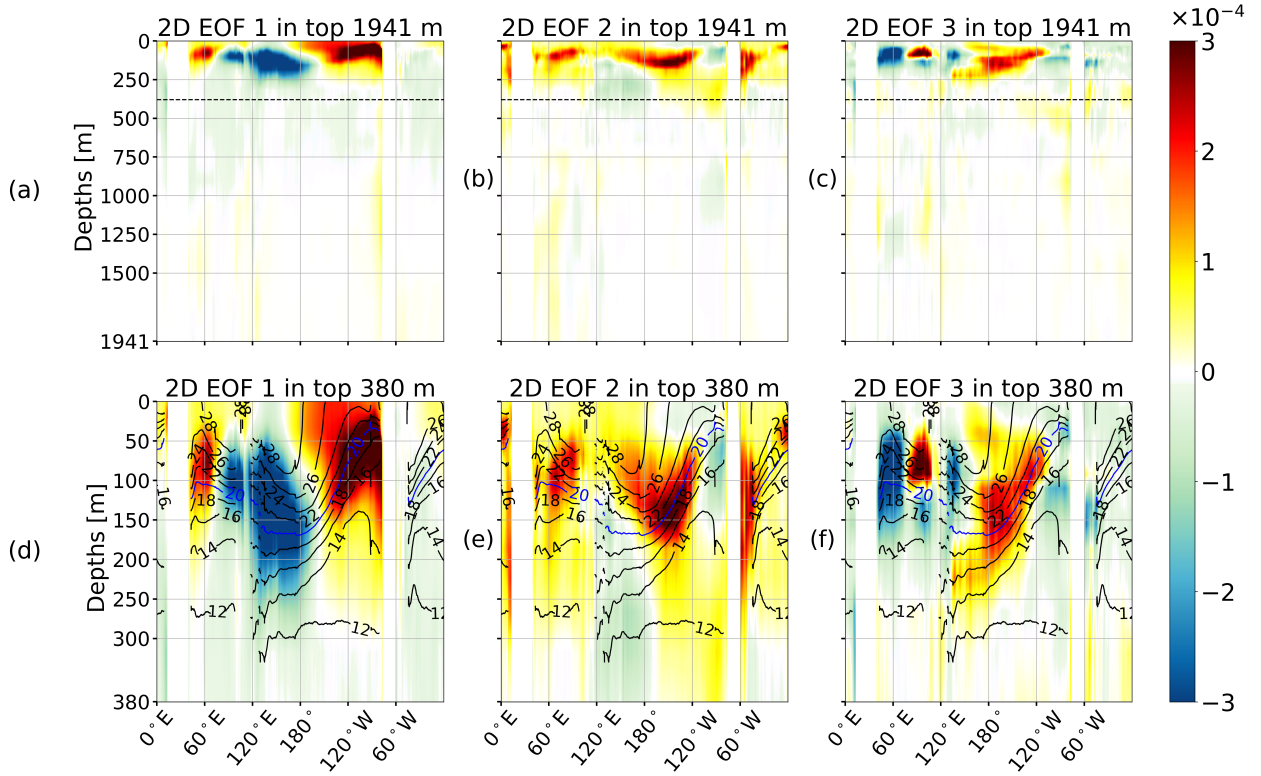


Figure 3.10: Same as Fig. 3.7 , but with the 2D calculation results.

To explore meridional mode representation in 2D PC 2, we computed correlations of the NPMM, SPMM, and AMM indices to 2D PC 2 at each layer. The NPMM index shows significant correlations ($r = 0.53 - 0.88$) to 2D PC 2 throughout the upper 266 meters (Fig. 3.11(a)). Correlations become negligible and statistically insignificant after 266 meters. The SPMM has weaker correlations ($r = 0.38 - 0.73$) over the same depth range, also declining below 266 meters.

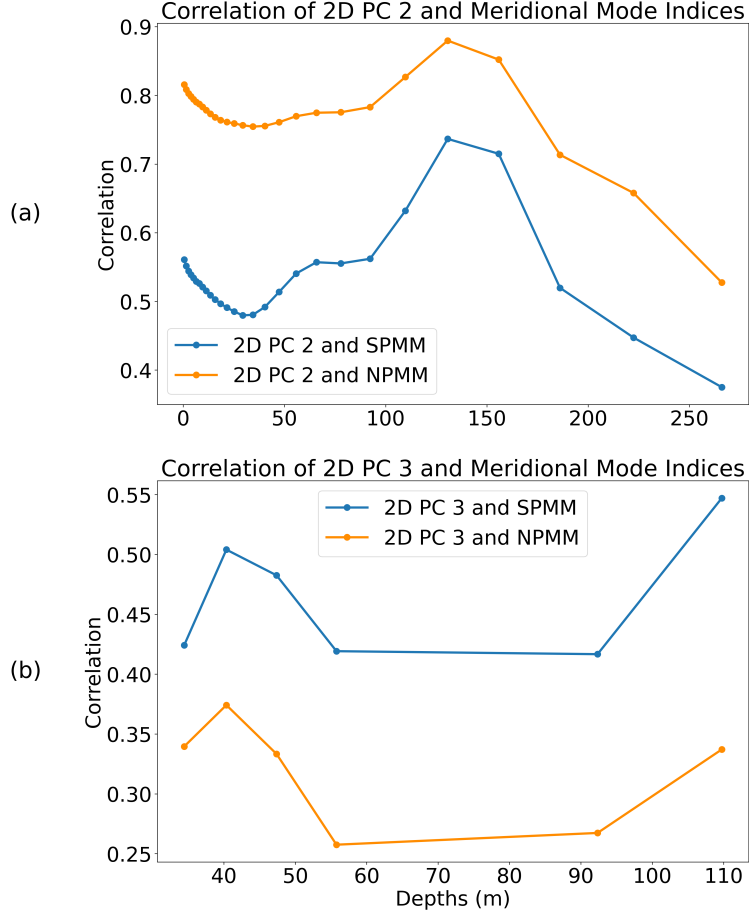


Figure 3.11: Correlations at every depth between meridional modes and (a) 2D PC 2 and (b) 2D PC 3. Figure (a) displays SPMM and NPMM correlations, omitting AMM due to statistical insignificance. Figure (b) shows only SPMM for the same reason. Depths without statistically significant correlations are also omitted.

Extending the analysis to CP El Niño indices demonstrates a notable relationship with the TNI (Fig. 3.12), consistent with the spatial pattern. The TNI, computed as the difference between SSTA in the eastern tropical Pacific (Niño 1+2) and the central tropical Pacific (Niño 4), yields a negative value when strong central tropical Pacific warming occurs with eastern tropical Pacific cooling or weaker warming. The negative correlation between 2D PC 2 and TNI ($r = -0.71$ - -0.41) at 56 - 186 meters coincides with depths of strong central tropical Pacific warming and weak eastern tropical Pacific cooling (i.e., NPMM) in 2D EOF 2 (Fig. 3.10(e)). The absence of CP El Niño and AMM signal in the spatial pattern results in

no significant correlation between 2D PC 2 and either the EMI or AMM. Thus, 2D PC/EOF 2 correspond to PMM variability (NPMM and SPMM) represented in 3D PC/EOF 3.

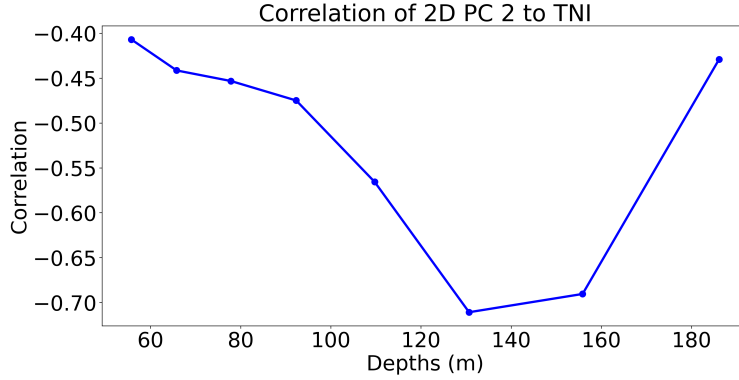


Figure 3.12: Correlation at every depth between 2D PC 2 and CP El Niño indices. EMI correlations were not statistically significant and are not shown. Depths without statistically significant correlations are also omitted.

Correlation analysis between 2D PC 2 and the 3D PCs (Fig. 3.13(a)) reveals that for the first 266 meters, the strongest correlation is with 3D PC 2 ($r = 0.78 - 0.97$). The maximum relationship of $r = 0.97$ occurs at 92 meters, and its minimum of $r = 0.78$ occurs at 266 meters. This strong relationship likely stems from the warming signal present in both time series. This warming signal is also observable in 2D EOF 2 as an overall positive signal across the globe for each layer (Fig. 3.14). The correlation magnitude is also consistent with the spatial pattern, where 92 meters experiences stronger warming than 266 meters.

At 266 meters, 2D PC 2 presents overlapping correlations with 3D PC 2 ($r = 0.78$) and 3D PC 1 ($r = 0.57$), suggesting a possible mixture of warming and ENSO signals at this depth. The 2D EOF 2 spatial pattern at 266 meters supports this interpretation, displaying both a prominent warming signal globally (Fig. 3.14) and a weak ENSO-like dipole structure in the tropical Pacific (Fig. 3.10(b)). As determined by the equatorial cross-section of 3D EOF 1 (Fig. 3.7(a)), most of El Niño's dipole signal occurs within the first 250 meters, leaving only weak cold/warm equatorial Pacific waters from 250 - 500 meters. The equatorial cross-sections of both 3D EOF 1 and 2D EOF 2 (Fig. 3.10(b)) show comparable negative loadings in the western tropical Pacific and positive loadings in the eastern tropical Pacific at 250

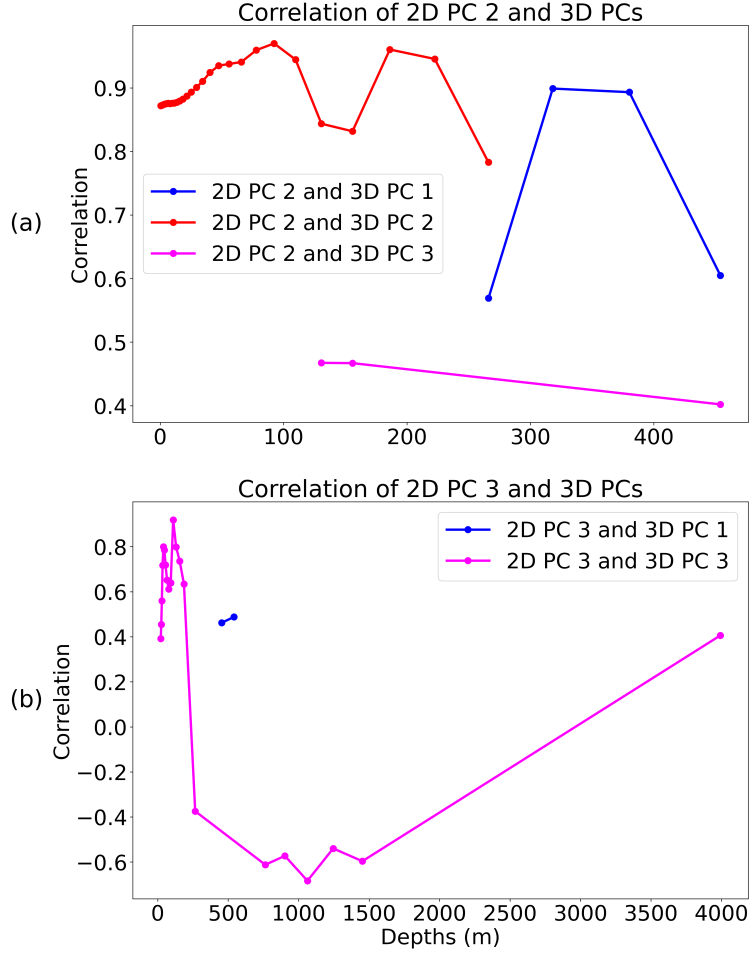


Figure 3.13: Correlations between 3D and 2D PCs. Panel (a) displays 2D PC 2 and correlations with all 3D PCs. Panel (b) same as (a) but for 2D PC 3. Results between 2D PC 3 and 3D PC 2 are not shown due to statistical insignificance. Depths without statistically significant correlations are also omitted.

- 500 meters. This similarity explains the overlapping correlations at 266 meters, and the continued correlation from 318 - 454 meters. However, at 454 meters, the global ocean signal in 2D EOF 2 is weak, suggesting the correlation likely results from noise (Fig. 3.14). Apart from this similarity, 2D EOF 2 shows no other features associated with ENSO, like the pIOD. The 2D PC/EOF 2 relationship to 3D PC/EOF 1 does not indicate ENSO presence in 2D PC/EOF 2, but rather reflects similarities between the time series and spatial patterns at these specific depths.

Of the 3D PCs examined, 3D PC 3 has the weakest correlation with 2D PC 2 overall. Nevertheless, three depths show statistically significant correlations: 131 ($r = 0.47$), 156 (r

2D EOF 2

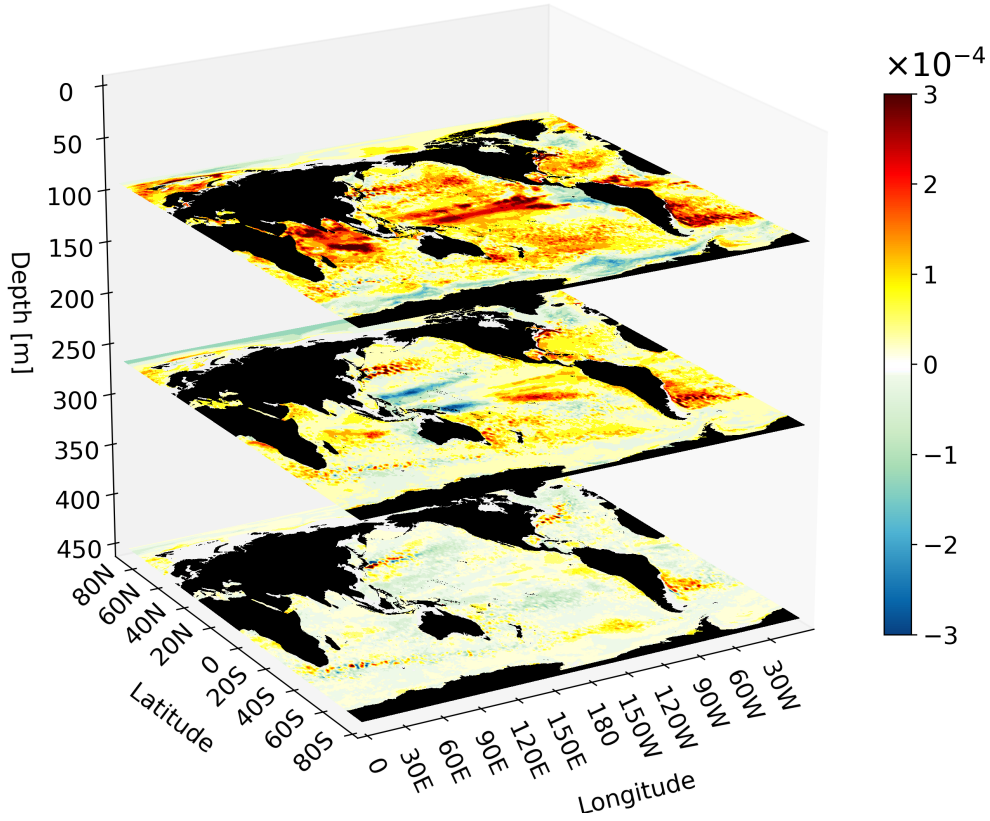


Figure 3.14: 2D EOF 2 visualized at three depths: 92, 266, and 454 meters. Depths 92 and 266 are displayed to interpret the global warming signal throughout the layers, with the 266 meter layer also displaying a weak dipole signal in the tropical Pacific. The 454 meter depth was chosen to show a globally weak signal with no identifiable climate modes, providing context for the final correlation depth between 2D PC 2 and 3D PCs 1 and 2.

$= 0.47$), and 454 meters ($r = 0.40$). Previous analysis of 3D EOF 3 showed most of the NPM and SPM variability occurs within the upper 200 meters, with maximum warming at 100 - 175 meters along the thermocline (Fig. 3.7(f)). Consistent with this, the 2D EOF 2 cross-section (Fig. 3.10(b)) exhibits similar behavior in the thermocline at 131 and 156 meters, supporting these correlations. In contrast, equatorial cross-sections for 3D EOF 3 (Fig. 3.7(c)) show weak to negligible signals past 380 meters. Since the 454 meter layer of 2D EOF 2 also displays weak signals throughout (Fig. 3.14), the correlation at this depth likely does not result from matching climate modes.

The surface of 2D EOF 3 is distinct from any patterns captured by the 3D EOFs (Fig. 3.9(c)). The surface contains a cold-warm-cold tripole pattern in the North Atlantic (10° - 70° N, North American coast to African coast) associated with the North Atlantic Oscillation (NAO) [89]. The Southern Pacific features warming extending from the South American coast to the date line between 60° S to $\sim 5^\circ$ S, creating a ‘C’-shaped warming pattern around cool southern central Pacific waters. This signal does not match the SPMM seen previously in 3D EOF 3 or 2D EOF 2 but resembles SPMM signals with the cold tongue index removed (Fig. 2(d) in [87]). For clarity, we refer to the 2D EOF 2 SPMM as SPMM-CTI to distinguish it from the 3D EOF 3 modes.

While the surface expression in the Pacific exhibits the SPMM-CTI pattern, the equatorial subsurface contains a composite signal incorporating both CP El Niño and PMM components. The surface of 2D EOF 3 in Fig. 3.9(c) shows little equatorial warming from 120° - 150° E, 10° S- 5° N. At 56 meters, this warming strengthens and grows in area, reaching further east (120° E - $\sim 100^\circ$ W) and spanning the tropics; the ‘C’-shaped SPMM-CTI warming blends into the central tropical Pacific warming, creating a pattern resembling the PMM in 3D EOF 3 and CP El Niño in 3D EOF 2. We determine the arrangement of the signals through a zonal cross-section analysis of 2D EOF 3. In Fig. 3.10(f), we observe the maximum warming in the thermocline at 50 - 200 meters roughly between 180° - 120° W. The tall half at 100 - 200 meters between 180° - 140° W resembles the NPMM TWC thermocline warming, and the narrow half at approximately 50 - 130 meters from 140° - 110° W matches the narrow thermocline warming from CP El Niño’s vertical thermocline displacement.

Beyond the central Pacific, the remaining subsurface of 2D EOF 3 most closely resembles 3D EOF 3. The Indian Ocean contains the same positive and negative loadings in the eastern and western equatorial regions at 50 - 100 meters (Figs. Fig. 3.7(f) and 3.10(f)). The Pacific also displays similar negative loadings starting at 56 meters (Fig. 3.9(f)) in the western (0° - 10° N, 120° - 150° E) and eastern tropical Pacific (10° S - 10° N, 100° - 80° W). The Atlantic

subsurface of 2D EOF 3 is unique, containing multiple areas with strong positive variability. This contrasts the South Atlantic warming in 3D EOF 3, most notably at 156 meters off the South American coast in Figs. 3.6(l) and 3.9(i).

To identify the combination of signals in the surface and subsurface, we correlated 2D PC 3 with meridional mode indices, EMI, TNI, and 3D PCs. The two CP El Niño indices, EMI and TNI, show nearly opposite correlations (Fig. 3.15). This is expected given their definitions: EMI subtracts eastern and western tropical Pacific SSTA from central tropical Pacific SSTA, while TNI subtracts the central from eastern tropical Pacific SSTA (see Appendix A for definitions). The surface of 2D EOF 3 contains cooling along the equator from 170°E to the coast of South America (Fig. 3.9(c)). With depth, central tropical Pacific warming strengthens, most notably at 29 and 34 meters (Appendix Fig. A.1). These depths correspond to where correlations with TNI (29 meters, $r = -0.378$) and EMI (34 meters, $r = 0.42$) begin, and explain their relatively small magnitudes compared to their following correlations.

The TNI correlations span a slightly larger depth range (29 - 3,221 meters) compared to EMI (34 - 3,221 meters). However, given the weak signals at deeper layers, exemplified by the absence of identifiable climate mode patterns at 1,452 meters (Fig. 3.18(c)), we focus our analysis on the upper 500 meters (Fig. 3.15(b)). Within this layer, significant correlations occur from 29 - 186 meters for the TNI ($r = -0.68$ - -0.38) and 34 - 266 meters for the EMI ($r = -0.41$ - 0.57). The TNI correlations remain consistently negative throughout these depths, while EMI shows a positive relationship from 34 - 131 meters ($r = 0.40$ - 0.57) before transitioning to a negative value at 266 meters ($r = -0.41$). This negative correlation corresponds to central tropical Pacific cooling instead of warming in 2D EOF 3 at 266 meters (Fig. 3.16). Examining the time series at this depth (Fig. 3.19(a)) confirms 2D PC 3 fails to capture CP El Niño events, which should appear as positive peaks in 2005, 2010, 2015, 2019, and 2020.

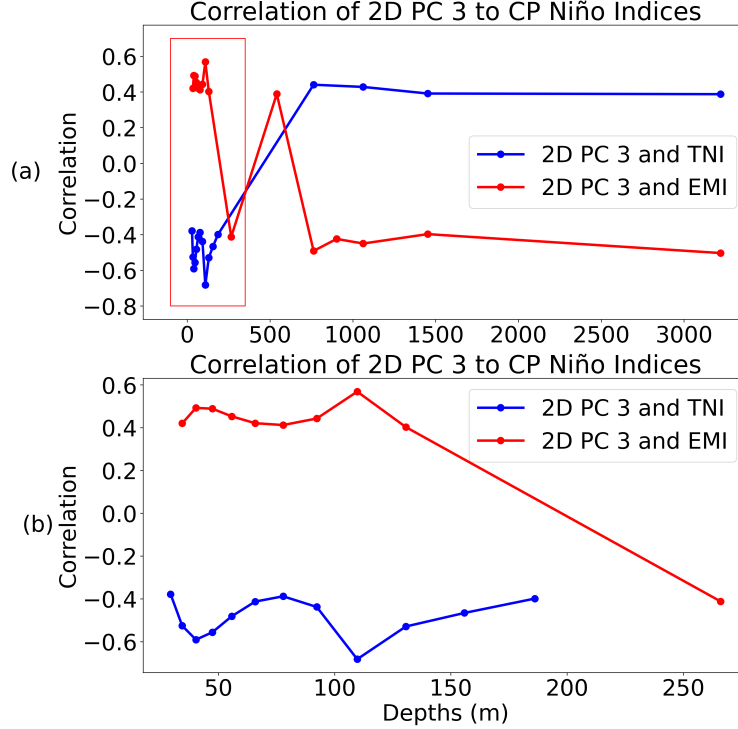


Figure 3.15: Correlation between two CP El Niño indices (TNI and EMI) and 2D PC 3. Panel (a) displays all statistically significant correlations. Panel (b) shows results from surface to 266 meters (red box in (a)). The first 266 meters are chosen to interpret correlations for depths where CP El Niño and the PMM experience the most variability.

The positive EMI and negative TNI correlations at 29 - 186 meters suggest CP El Niño features are present in these layers. 2D EOF 3 displays central tropical Pacific warming that emerges (Fig. A.1), intensifies, weakens, and dissipates (Fig. 3.10(f)) within these depths. Notably, the strongest correlations at 40 and 110 meters align reasonably well with the characteristic CP El Niño thermocline warming depth range of 50 - 130 meters observed in 2D EOF 3.

That said, not all correlations are easily explained. While TNI and EMI definitions are designed to capture CP El Niño's unique structure, both indices are inherently biased toward central tropical Pacific warming. Consequently, correlations with either index could arise from central tropical Pacific warming associated with PMM or CP El Niño. We address this ambiguity by supplementing our correlation analysis with spatial patterns and known CP El Niño mechanisms such as thermocline variability. Nevertheless, the mixture of CP

2D EOF 3

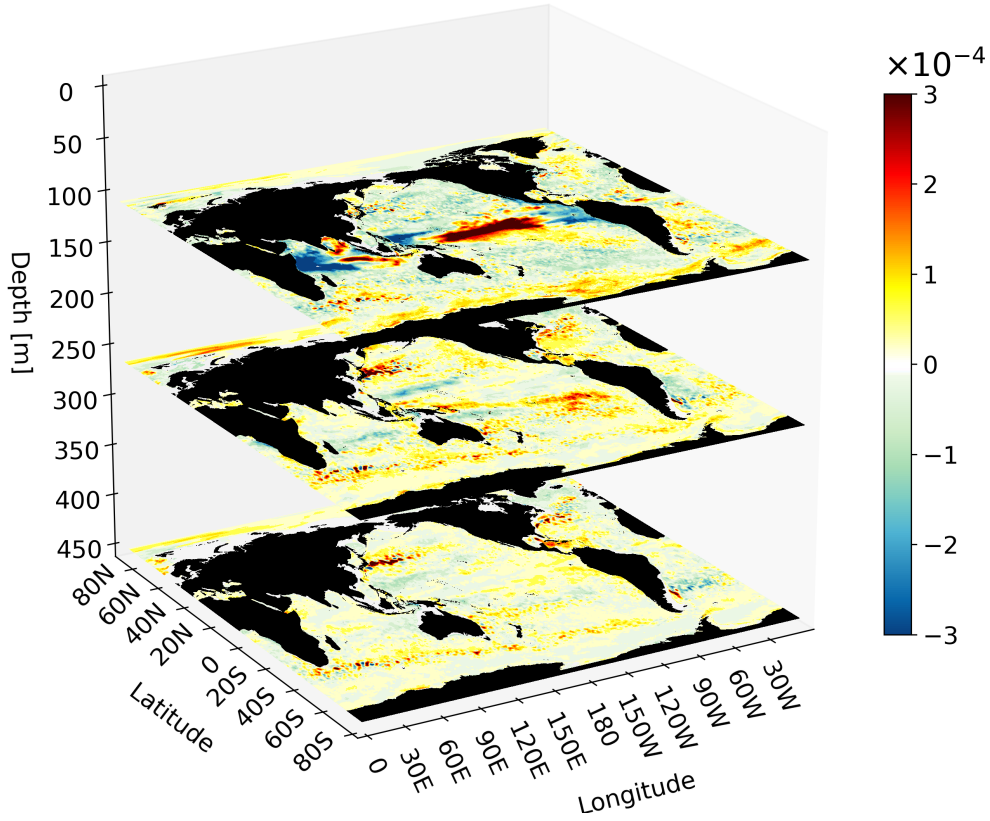


Figure 3.16: 2D EOF 3 displayed at three depths: 110, 266, and 454 meters. The 110 meter layer is displayed to interpret the PMM and CP El Niño signal. The depth 266 meters is displayed to explain the negative EMI and 3D PC 3 correlation, and 454 meters is displayed to examine the spatial pattern associated with a positive correlation between 3D PC 1 and 2D PC 3.

El Niño and PMM signals in 2D PC/EOF 3 makes it difficult to conclusively determine which mode drives the EMI/TNI correlations in the upper 200 meters. In general, 2D PC 3 correlations align with spatial patterns that contain subsurface central tropical Pacific warming that exhibits characteristics of both CP El Niño and the NPMM. Due to this signal mixing, the remaining correlation analysis shows that 2D PC 3 produces similar relationships with the EMI, TNI, SPMM, NPMM, and 3D PC 3, where all correlations are significant at overlapping depths and reach their maxima at the same layers (Fig. 3.17).

Among the meridional mode indices, NPMM ($r = 0.26 - 0.37$) and SPMM ($r = 0.42 - 0.56$) showed statistically significant correlations with 2D PC 3 between 34 - 110 meters (Fig.

3.11(b)). As discussed previously, correlations begin at 34 meters where central tropical Pacific warming becomes apparent (Fig. A.1). Both SPMM and NPMM correlations experience two maxima: one at 40 meters (NPMM: $r = 0.37$, SPMM: $r = 0.50$) and another at 110 meters (NPMM: $r = 0.34$, SPMM: $r = 0.56$). Spatial patterns are consistent with these maxima, displaying an NPMM signal at 40 meters (Fig. A.1) and central tropical Pacific warming at 110 meters that exhibits characteristics of both CP El Niño and the NPMM (Fig. 3.16). These significant correlations extend only to 110 meters, shallower than the characteristic NPMM thermocline warming previously identified in the zonal cross-section of 2D EOF 3 (100 - 200 meters, 180° - 140° W; Fig. 3.10(f)). This depth mismatch indicates our correlation analysis captures only shallow NPMM variability present in 2D EOF 3.

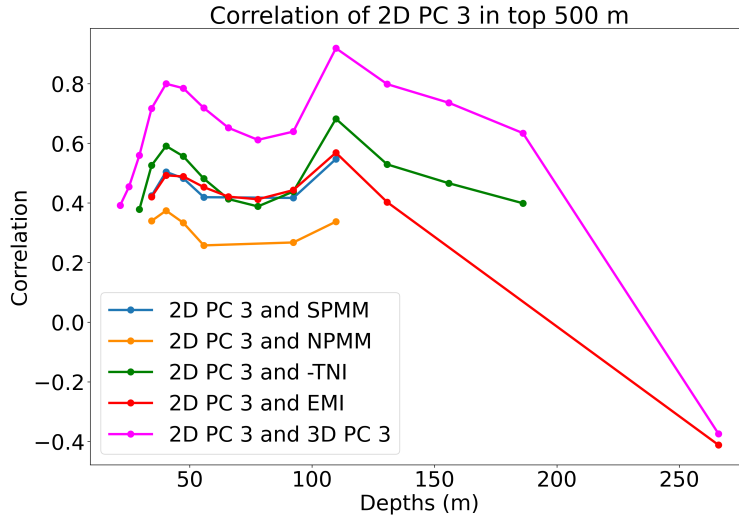


Figure 3.17: All correlation results for 2D PC 3 in the top 500 meters. AMM, 3D PC1, and 3D PC 3 show no significant correlations above 500 meters and are not displayed in this figure.

Correlation analysis of 2D PC 3 with various climate indices showed relationships primarily with SPMM, NPMM, and CP El Niño. As expected, this results in no relationship with 3D PC 2 (warming). There are moderate correlations between 2D PC 3 and 3D PC 1 at 454 meters ($r = 0.46$) and 541 meters ($r = 0.49$). The spatial patterns at these depths, however, do not show any signals relevant to the traditional ENSO pattern, as demonstrated by the 454 meter layer in Fig. 3.16. Large correlations in deep layers (763 - 3,992 meters) are also unsupported by the spatial pattern. For example, Fig. 3.18 shows that at 1,452

meters, the depth with minimum correlation, both 2D EOF 3 and 3D EOF 3 contain no identifiable climate mode signals. As such, correlations from 763 - 3,992 meters do not result from matching climate dynamics between 2D EOF 3 and 3D EOF 3.

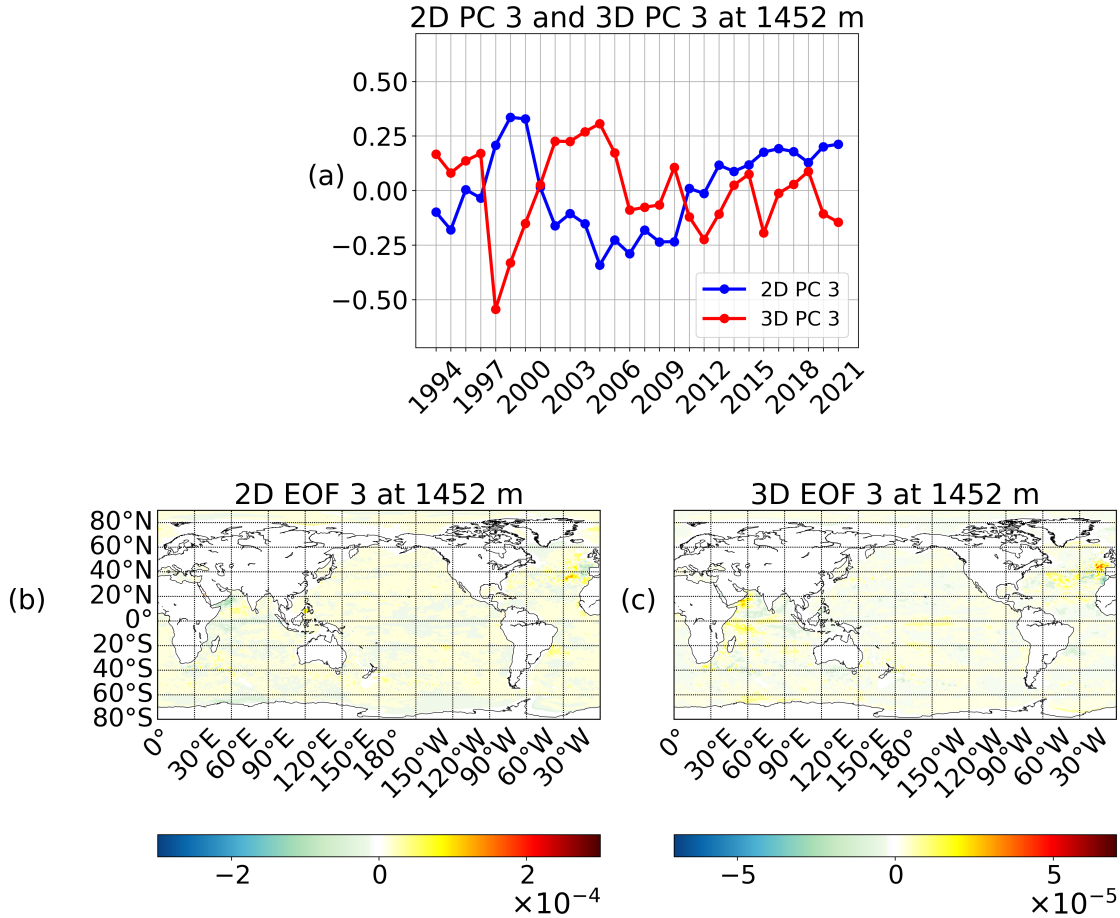


Figure 3.18: Visualization of 2D PC/EOF 3 and 3D PC/EOF 3 at 1,452 meters. PCs and EOFs are displayed to compare spatial patterns and time series of the depth layer with minimum correlation between 2D PC 3 and 3D PC 3 as defined by Fig. 3.13(b). The top row shows (a) 2D PC 3 and 3D PC 3. The bottom row shows the corresponding EOFs: (c) 2D EOF 3 and (d) 3D EOF 3. EOF plots use different colorbar ranges to effectively visualize cold/warm signals for comparison.

In the first 500 meters, the correlation between 2D PC 3 and 3D PC 3 displays similar behavior as the previous climate mode correlations. Figure 3.17 shows all correlations to 2D PC 3 for the top 500 meters. For easier comparison, the TNI sign is flipped to appear as positive. Correlations with 3D PC 3 are positive ($r = 0.39 - 0.91$) from 22 - 186 meters and, like the EMI, is negative at 266 meters ($r = -0.37$).

The negative correlation with 3D PC 3 at 266 meters reflects differences in both time series and spatial patterns. Unlike 3D PC 3, 2D PC 3 lacks PMM characteristics. Figure 3.19(a) shows that 2D PC 3 fails to produce negative values during warm and cold ENSO events. Comparing the spatial patterns (Figs. 3.19(b) and (c)) reveals striking differences: 2D EOF 3 displays signals opposite to 3D EOF 3 throughout much of the globe. In the Atlantic off the South American coast, for instance, 3D EOF 3 displays warming, where 2D EOF 3 displays cooling.

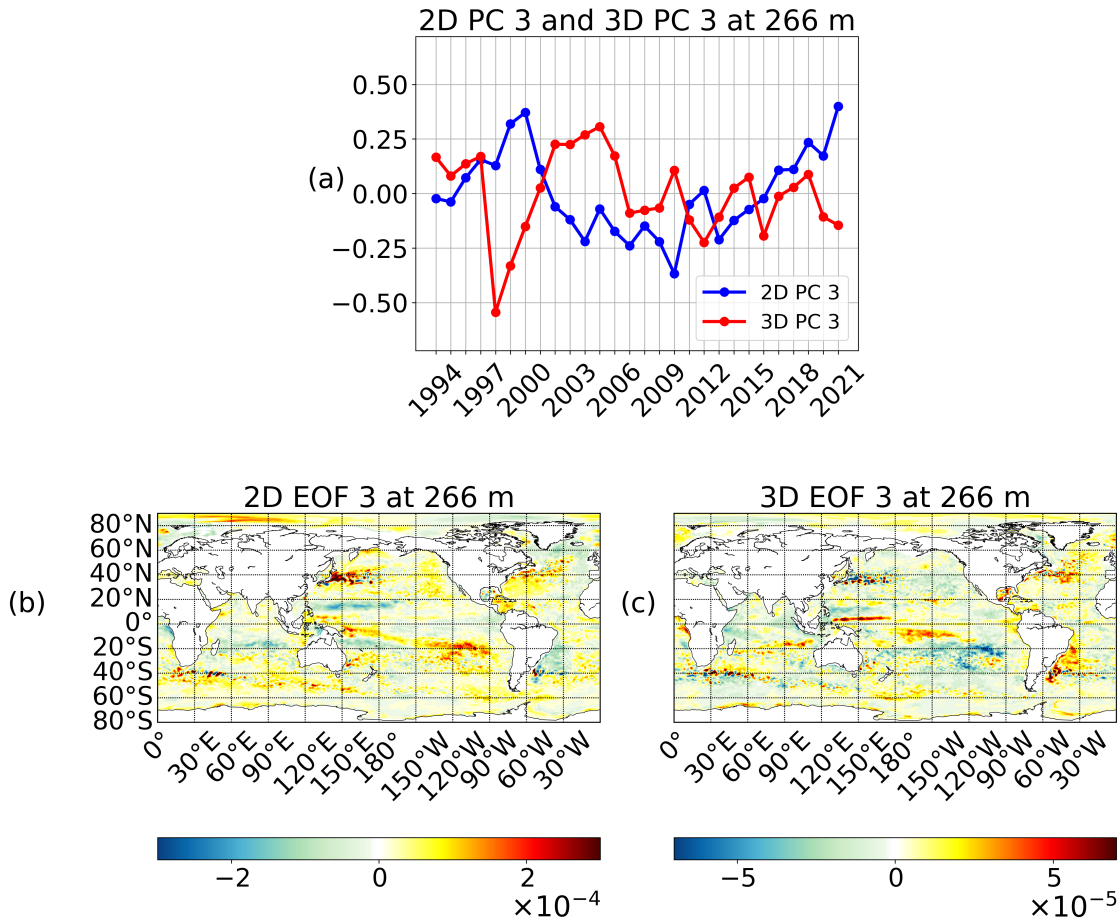


Figure 3.19: Same as Fig. 3.18 but for 266 meters.

Like the previous climate indices, 3D PC 3 exhibits two correlation maxima with 2D PC 3 at 40 meters ($r = 0.80$) and at 110 meters ($r = 0.92$) (Fig. 3.17). We attribute the maxima to the same central tropical Pacific warming associated with CP El Niño/PMM variability described for other indices. The remaining positive 2D and 3D PC 3 correlations increase

and decrease along with central tropical Pacific warming (Fig. A.1), again matching the correlation behavior of previous indices. The overlapping correlation depths and consistent behavior confirm mixed subsurface CP El Niño and PMM signals in 2D PC/EOF 3. Considering the well-documented relationship between PMM and CP El Niño [10,90–92], it can be difficult to separate the two signals. Though our 3D EOFs successfully separate the signals, the 2D EOFs do not.

In the 2D calculation, PMM variability does not appear as a single EOF as it does in the 3D calculation. While PC correlations with meridional mode indices provide valuable insight, they are unable to capture the complete NPMM seen in 2D EOF 3 (e.g., thermocline variability) due to its complex climate mode composition. For additional clarity, we use the NPMM subsurface signature in 3D EOF 3 to identify the NPMM signal that is distributed between 2D EOF 2 and 2D EOF 3. We define these features to align with known NPMM subsurface mechanisms [16]. The characteristic NPMM subsurface pattern consists of:

- (i) Central Pacific warming at 5°S - 5°N between 150°E - 120°W from surface - 200 meters.
- (ii) Maximum warming between 100 - 160 meters along the thermocline between the date line and 120°W .
- (iii) Subsurface warming centered on 5°N with varying longitudes based on Fig. 3.6 (right column): 56 (150°E - 120°W), 156 (130°E - 120°W), and 222 meters (130°E - 170°W).
- (iv) Warming between 10°S and the equator at 156 (150°E - 110°W), and 222 meters (170°E - 120°W).

Analysis of equatorial cross-sections (Figs. 3.5 and 3.10) shows how features (i) and (ii) are distributed between the 2D EOFs. Although the warming pattern in 2D EOF 2 resembles that of 3D EOF 3, its western extent reaches too far past the dateline, and its maximum warming terminates prematurely at approximately 130 meters. The central tropical Pacific

warming in 2D EOF 3 successfully captures the spatial domain of the warming found in 3D EOF 3, but fails to reproduce the shape of the maximum thermocline warming as effectively as 2D EOF 2. Combining the shape of the thermocline warming in 2D EOF 2 and the spatial domain on 2D EOF 3 reproduces features (i) and (ii). At 56 meters (Figs. 3.9(e) and (f)), feature (iii) appears primarily in 2D EOF 2, though the rest of the global 3D EOF 3 pattern is evident in 2D EOF 3. Features (iii) and (iv) appear identically in both 2D EOFs at 156 meters. At 222 meters, feature (iii) corresponds to the northern band in 2D EOF 3, while feature (iv) represents the southern band in 2D EOF 2.

3.4 Conclusion and Discussion

We have developed a computational approach for 3D EOF analysis of large oceanographic reanalysis datasets by using the PIM. The PIM automatically partitions an input matrix for temporal covariance or 3D EOF calculations, given any specified amount of memory. Our testing showed that larger memory allocations significantly reduce computation time. However, users should avoid utilizing their entire RAM capacity to prevent performance degradation at the upper memory limits.

Our 3D visualizations of EOFs can help intuitively reveal the 3D structure of dominant variability, eliminating the need to mentally reconstruct 3D structures from multiple 2D views. Detailed memory requirements to create each figure are given to demonstrate our visualization methods' capabilities in handling large datasets. Although our 3D visualization computer is efficient, it is limited in the total layers it can visualize from an EOF file.

The 3D EOF analysis revealed distinct climate signals, with the second EOF capturing the interconnected dynamics of warming, CP El Niño, and MHW. This finding supports existing theories regarding the symbiotic relationships between these phenomena. The in-

creased frequency and intensity of CP El Niño events over the past three decades coincide with increased warming [6, 93], suggesting that warming may contribute to growing CP El Niño occurrence [84]. Conversely, because CP El Niño events predominantly occur as warm phases, they may contribute to the overall warming trend [8]. Similarly, MHW intensity and frequency have increased since the 1990s [77], demonstrating a comparable relationship with warming as observed for CP El Niño events. The connection between CP El Niño and MHWs depends on the region. Our results identified four regions with substantial surface and subsurface warming: the North Pacific ($\sim 20^\circ - 60^\circ\text{N}$, $160^\circ - 130^\circ\text{W}$), Southern Ocean ($60^\circ - \sim 35^\circ\text{S}$, $180^\circ - \sim 100^\circ\text{W}$), southwestern Indian Ocean ($30^\circ - 10^\circ\text{S}$, Madagascar coast to 80°E), and a signal starting at 56 meters in the Atlantic along South America. These patterns align with documented areas where CP El Niño influences MHW characteristics through western/central Pacific variability [91, 94]. The relationship between the mentioned dynamics raises the question of the effect of anthropogenic warming. Although the answer is beyond the scope of this paper, it has been a part of other ongoing research [31, 77, 84, 95].

We computed 2D EOFs for the same spatial regions to compare with our 3D results. Our 2D EOFs indicated a shift of NPMM to the second EOF, where it coupled with warming and MHW signals, and a mixture of PMM and CP El Niño variability in the third EOF. Using the resulting time series at every depth, we applied correlation analysis to further identify the mode composition at each depth. 2D PC 2 correlations confirmed the presence of NPMM, SPMM, and warming. Due to 2D EOF 3's spatial structure, 2D PC 3 produced similar correlations with all climate indices (excluding AMM) and 3D PC 3 (PMM), confirming the presence of both CP El Niño and PMM signals. Through visual comparison of the resulting spatial patterns, we identified each signal in 2D EOF 3 and explored how NPMM features were divided between 2D EOF 2 and 3. We find the 2D results do not contain clear representations of CP El Niño or NPMM as the 3D results do. Because the 2D results are limited to individual layers, this potentially leads to a distorted view of variability. 3D EOFs

address this issue by computing EOFs across all layers simultaneously, capturing cross-layer covariances and providing a more realistic representation of ocean dynamics.

The PIM can accommodate the big data computational scenarios for most personal computers. It removes computational constraints on 3D EOF calculations, enabling the use of large high-resolution climate datasets like GLORYS. While our temperature-focused analysis demonstrates the technique’s effectiveness in revealing cross-layer climate dynamics, extending this approach to additional oceanographic variables would enhance understanding of the underlying physical mechanisms driving these climate phenomena.

Chapter 4

Four-dimensional North Pacific Variability and Regional Impacts from a High-Resolution Ocean Reanalysis

4.1 Introduction

The Pacific contains several large-scale phenomena that contribute to ocean temperature variability. El Niño-Southern Oscillation (ENSO) [2], Eastern Pacific (EP) and Central Pacific (CP) El Niño [6, 8, 96], and the North Pacific meridional mode (NPMM) [9] are widely studied leading modes of climate variability in this region. While sea surface temperature anomalies (SSTA) are commonly used for these analyses, subsurface ocean temperature anomalies (SOTA) provide additional insight into distinguishing characteristics of these modes.

Multi-layer ocean analyses have advanced our understanding of phenomena through their exploration of subsurface structure. Previous studies have used SOTA to identify distin-

guishing features and mechanisms for the two El Niño types [6, 14, 96], define a subsurface index to help classify and separate the two types of El Niño [13], and link the trade wind charging mechanism related to the NPMM as a precursor to mature El Niño [16]. The studies [6, 13, 96] and [16] analyzed subsurface patterns using principal component analysis (PCA) of SSTA and lagged correlations to SOTA. In contrast, [14] computed subsurface variability using PCA on a zonal average between 5°S - 5°N. These studies emphasized the importance of multi-layer dynamics, but limited ocean variability to either the surface or a zonal cross-section. Therefore, an analysis that considers all layers simultaneously to capture 3D variability would provide a comprehensive understanding of important ocean dynamics.

Our study aims to quantitatively describe the 3D structure of ocean variability in the North Pacific by computing 3D empirical orthogonal functions (EOFs) on the high-resolution Global Ocean Physics Reanalysis (GLORYS) ocean temperature data. We explore the subsurface patterns for ENSO, the NPMM, EP El Niño, and CP El Niño throughout the region to the depths of 541 meters. Our use of high-resolution ($1/12^\circ$ latitude-longitude) data allows for detailed analysis of regional effects from both El Niño types, specifically along the U.S. West Coast (USWC) and within the California Current System (CCS).

The remainder of this paper is organized as follows: Section 4.2 describes the dataset, data processing methods, and the approach used to isolate different ENSO types. Section 4.3 presents the results from the 3D EOF calculation and the rotated indices. Section 4.4 offers the conclusion and discussion from the analysis.

4.2 Data and Methods

The ocean temperature in this study is from the Global Ocean Physics Reanalysis (GLORYS) [79]. The data is a high-resolution ($1/12^\circ$ by $1/12^\circ$) eddy-resolving global ocean simulation

available at Copernicus Marine Environment Monitoring Service. We use latitude (20°S - 70°N) and longitude (120°E - 60°W) ranges similar to [2] for the North Pacific to capture ENSO and the NPMM variability. Although these modes are primarily tropical phenomena, their influence extends into the North Pacific [30, 97]; therefore, we extend our latitude coverage to 70°N . The GLORYS data contains 50 depth layers from the surface to 5727.92 meters. We use only the first 32 depths (surface - 541.09 meters), which are sufficient to represent the variability of the aforementioned modes. Following the timing of El Niño in boreal winter, we use a December-January-February (DJF) mean of GLORYS ocean temperature data from 1993/1994 boreal winter to 2020/2021.

The 3D EOFs are calculated from temporal covariance following methods in [42]. Using temporal covariance is preferable because it requires much less computer memory than spatial covariance methods, which can be prohibitively expensive for high-resolution 3D data. The covariance is calculated from a volume-weighted anomaly matrix constructed following [23] for a 3D analysis. To further consider computational memory cost, the matrix multiplication required for calculating temporal covariance and 3D EOFs are performed using the partitioned approach described in Chapter 3. For temporal covariance, partitioned segments of the transposed anomaly matrix are sequentially read in and multiplied by corresponding partitioned segments of the anomaly matrix until the full multiplication is complete. Similarly, 3D EOFs are calculated by multiplying partitioned segments of the anomaly matrix with the eigenvectors of the temporal covariance matrix.

Our first two principal components (PCs) and EOFs will contain ENSO and the NPMM signals. For clarity, we interpret ENSO as warm anomalies in the eastern tropical Pacific represented by the leading mode in our EOF calculation. Following methods in [32], we apply a 45° rotation to our first two EOFs and PCs and create two new indices that separate

El Niño events into Eastern Pacific (EP) and Central Pacific (CP) El Niño (see Eq. 4.1).

$$\begin{aligned} C &= (PC1 + PC2)/\sqrt{2} \\ E &= (PC1 - PC2)/\sqrt{2} \end{aligned} \tag{4.1}$$

The C index represents the CP El Niño and demonstrates its maximum variability near the dateline. The E index reflects EP El Niño and is similar to the ENSO signal but its maximum variability favors the eastern Pacific. For the interpretation of each feature, we focus on the positive spatial pattern expressions of the modes (i.e., ENSO, NPMM, CP, and EP El Niño) in our results.

4.3 Results

4.3.1 EOF 1 and 2

The first two EOFs, representing a 46% cumulative variance, characterize ENSO and the NPMM. We focus on the first two EOFs/PCs because they represent a large variance in the data, and are the only components necessary to create the C and E indices. Our resulting 3D spatial patterns reproduce the expected surface expressions from previous 2D studies that define the rotated indices [32, 98], while including their subsurface signatures.

The first EOF (37% variance) has a surface ENSO signal in its positive phase with equatorial warming extending from the South American coast to the date line (Fig. 4.1(c)). The surface warming covers three ENSO index regions: Niño 1+2 (10°S - 0, 90° - 80°W), Niño 3 (5°S - 5°N, 150° - 90°W), and Niño 4 (5°S - 5°N, 160°E - 150°W). Consistent with this spatial pattern, the corresponding time series (PC 1) shows strong correlations with these regional indices (0.813, 0.924, and 0.873, respectively). The PC 1 maxima at 1998 and 2016 reflect

strong El Niño events, while its minima correspond to La Niña years in 1999, 2000, 2008, and 2011.

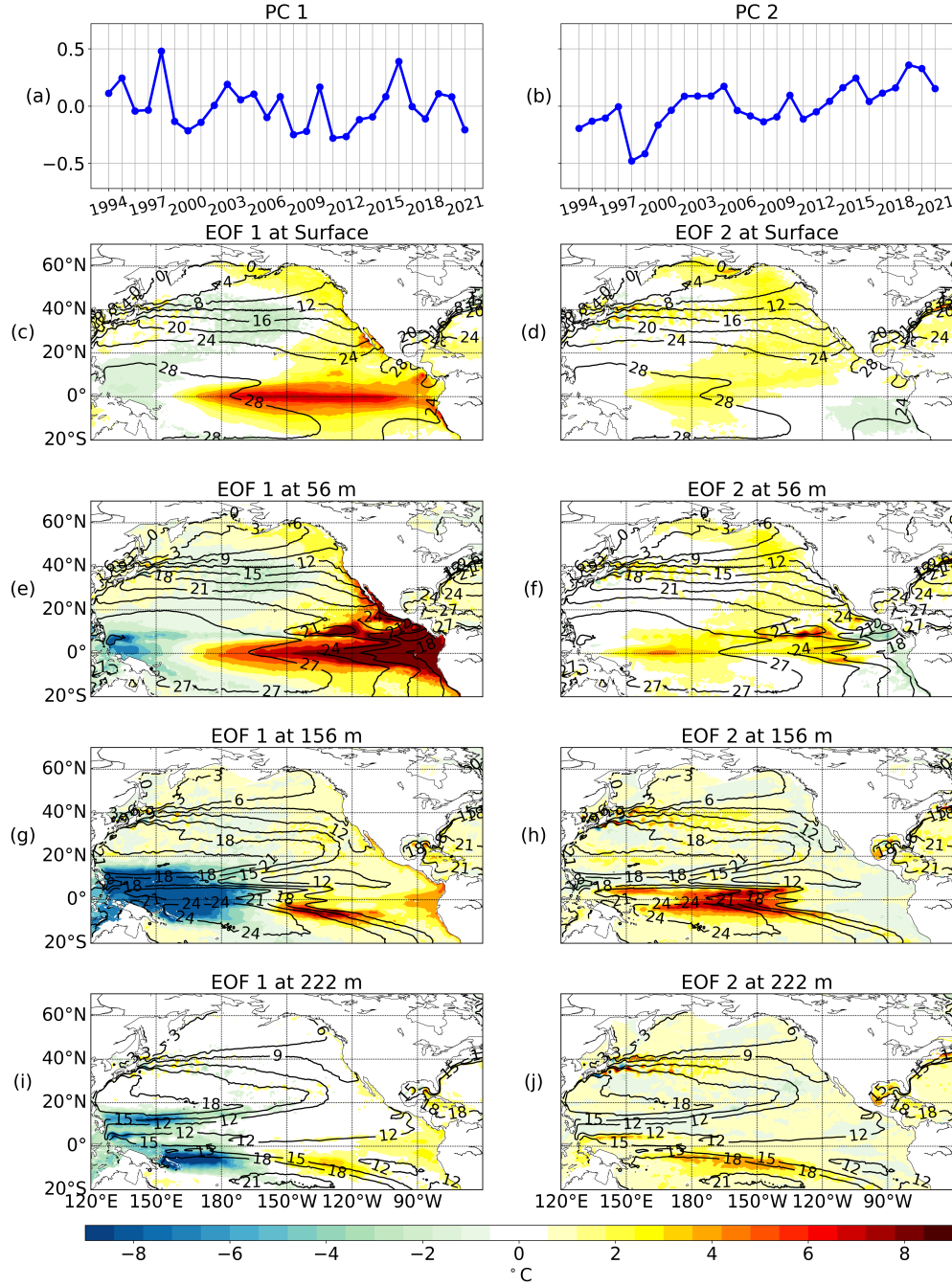


Figure 4.1: Visualization of the first two EOFs and their PCs: 3D EOF 1 and PC 1 (left column) and 3D EOF 2 and PC 2 (right column) for the North Pacific Ocean. EOFs are displayed at four depth layers (surface, 56 meters, 156 meters, 222 meters). These depths were chosen for EOF 1 to show ENSO, and EOF 2 to show the PMM. The depth values are in accordance with depth layers defined by GLORYS.

The subsurface temperatures of EOF 1 contain the canonical dipole pattern of ENSO, with a cool western tropical Pacific and warm eastern tropical Pacific. Using a zonal cross-section averaged from 5°S to 5°N (Fig. 4.2(a)), we determine the vertical extent of the dipole pattern and the locations of its maximum variability. In the upper 70 meters, the eastern Pacific warming consistently extends past the date line with maximum intensity between 145° - 80°W along the thermocline between 25 - 120 meters. Below 150 meters, the dipole's warm half is a weak warm signal ($\sim 1.25^\circ\text{C}$), with a small portion containing stronger warming ($\sim 2.5^\circ\text{C}$) centered at 90°W and extending beyond ~ 320 meters. The dipole's western cooling is weak at the surface and also reaches its extreme in the thermocline (50 - 175 meters). This strong cooling expands longitudinally from an initial range of 120° - 155°E at 50 meters to eventually span from 120° - 170°W at 150 meters.

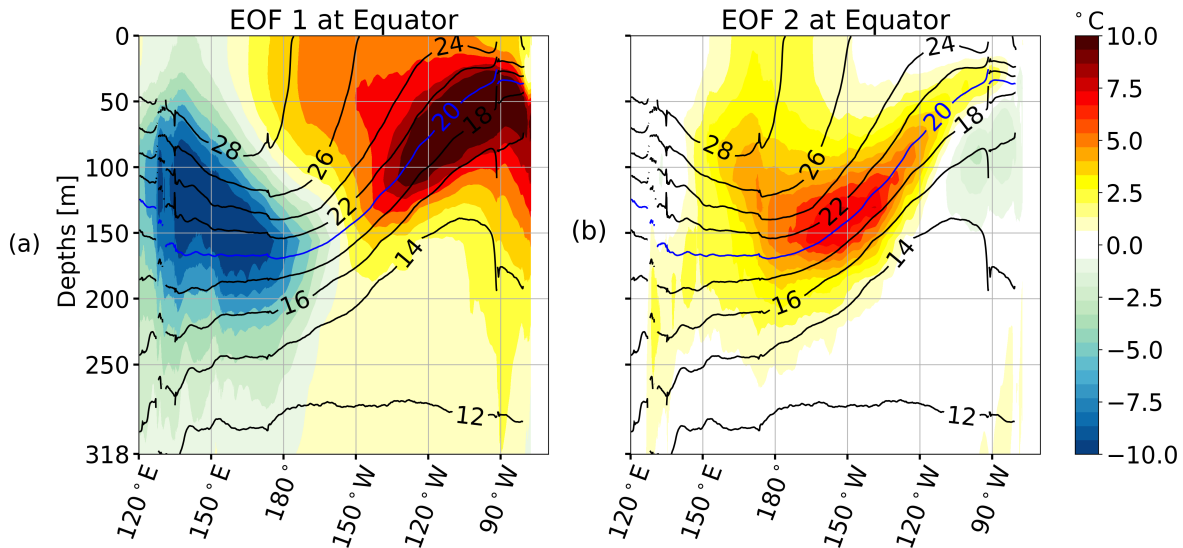


Figure 4.2: Depth-zonal view of the first two EOFs over the tropical region. The values displayed in color are taken as the average of 3D EOFs from 5°S to 5°N. Both panels are from surface to 318 meters; the weak variability after 318 is not shown. Black contours are isotherms from boreal winter (DJF) climatology. The blue contour is the 20°C isotherm to represent the thermocline.

The second EOF, capturing 9% of the 3D temperature anomaly variance, represents the NPM. In its positive phase, EOF 2 shows warm temperature anomalies at the equator around the dateline, with a branch extending northeast toward North America (Fig. 4.1(d)). This warming, though weak, extends into the North Pacific, covering a larger area off the

North American coast than EOF 1's North Pacific surface warming. The equatorial subsurface EOF 2 does not contain any other signatures besides the warming in the central Pacific in the first 50 meters, warming in the thermocline, and weak cooling at 120° - 80° W from 50 - 150 meters. Elevated warming signals are visible in the center of the basin between 50 - 200 meters, with the maximum warming occurring from 100 - 160 meters at 170° - 120° W. This maximum coincides with warm subsurface anomalies of the NPMM related to trade wind charging [10, 16, 99].

Our PC 2 incorporates variability associated with the NPMM in that both El Niño and La Niña events project onto negative values in the NPMM time series [9]. This applies to all previously mentioned years except for the recent event in 2016. This year is a local minima, but it is possible that an upward linear trend ($r = 0.48$) in PC 2 influenced the year to be nonnegative. We also compare PC 2 against the Pacific meridional mode index using a modified approach that only requires temperature data [4, 10]. To calculate the NPMM index for the GLORYS data we: (i) remove the linear trend from anomalies, (ii) remove ENSO influence by linearly regressing out the first EOF of tropical Pacific SSTA (20° S - 20° N, 140° E - 90° W), and (iii) compute EOFs and PCs from regional SSTA fields (20° S - 30° N, 175° E - 95° W). Even though the NPMM index is computed for just the surface, our 3D PC 2 has a high correlation of 0.91.

In both EOFs, the strongest signals occur in the thermocline, indicated by closely-spaced isotherms (Fig. 4.2). In EOF 1, the maximum western signal is deeper than the eastern signal, consistent with the thermocline sloping upward from west to east. Similarly, the maximum signal in EOF 2 occurs at 100 - 160 meters, in the thermocline near the middle of the basin. Due to the strong vertical temperature gradients in these regions, vertical displacement of the thermocline can induce large temperature anomalies.

4.3.2 E and C Indices

Our first two EOFs and PCs align with the findings of [32] and [98]; the leading modes are ENSO and the NPMM with similar surface signatures to their EOF results. Additionally, our EOF's subsurface patterns reflect known ENSO and NPMM dynamics, and our PC 1 and PC 2 correlate highly with El Niño and NPMM indices, respectively. Therefore, we extend the 2D methods from [32] to 3D by combining our first two PCs (Eq. 4.1) and EOFs to define two new orthogonal indices, C and E, that separate CP and EP El Niño events. The two El Niño types differ in the locations of their maximum SSTA [6,8,96]. Our C and E patterns depict temperature anomaly variance for CP and EP El Niño throughout the water column.

Similar to the dipole pattern EOF 1 ENSO (Fig. 4.1), the E index experiences strong warming in the eastern tropical Pacific with negative loadings in the western tropical Pacific (Fig. 4.3). However, differences emerge in the locations and behavior of maximum warming starting at the surface. While EOF 1 exhibits a strong warming across the eastern equatorial Pacific, the E pattern is strongest in the east (120° - 90° W) and weakens westward. We can also observe this trait in our time series by comparing the E index to the same Niño indices as PC 1. The westernmost index (Niño 4) has no significant correlation, while the more eastern indices Niño 3 ($r = 0.66$) and Niño 1+2 ($r = 0.75$) have higher correlations. Moreover, our E index time series satisfies the definition from [32] of capturing only strong El Niño events: all La Niña years are nearly 0 (Fig. 4.3(a)), and its maxima at 1998 and 2016 correspond to strong El Niño years.

The subsurface E index pattern (Fig. 4.4(a)) has a different vertical structure than the EOF 1 ENSO pattern. In the first 50 meters, the western cooling has a similar weak signal to EOF 1 but extends further east (120° - 170° E). Between ~ 75 - 200 meters, we observe the bulk of the eastern cooling; in these depth layers, cooling reaches as far east as $\sim 160^{\circ}$ W,

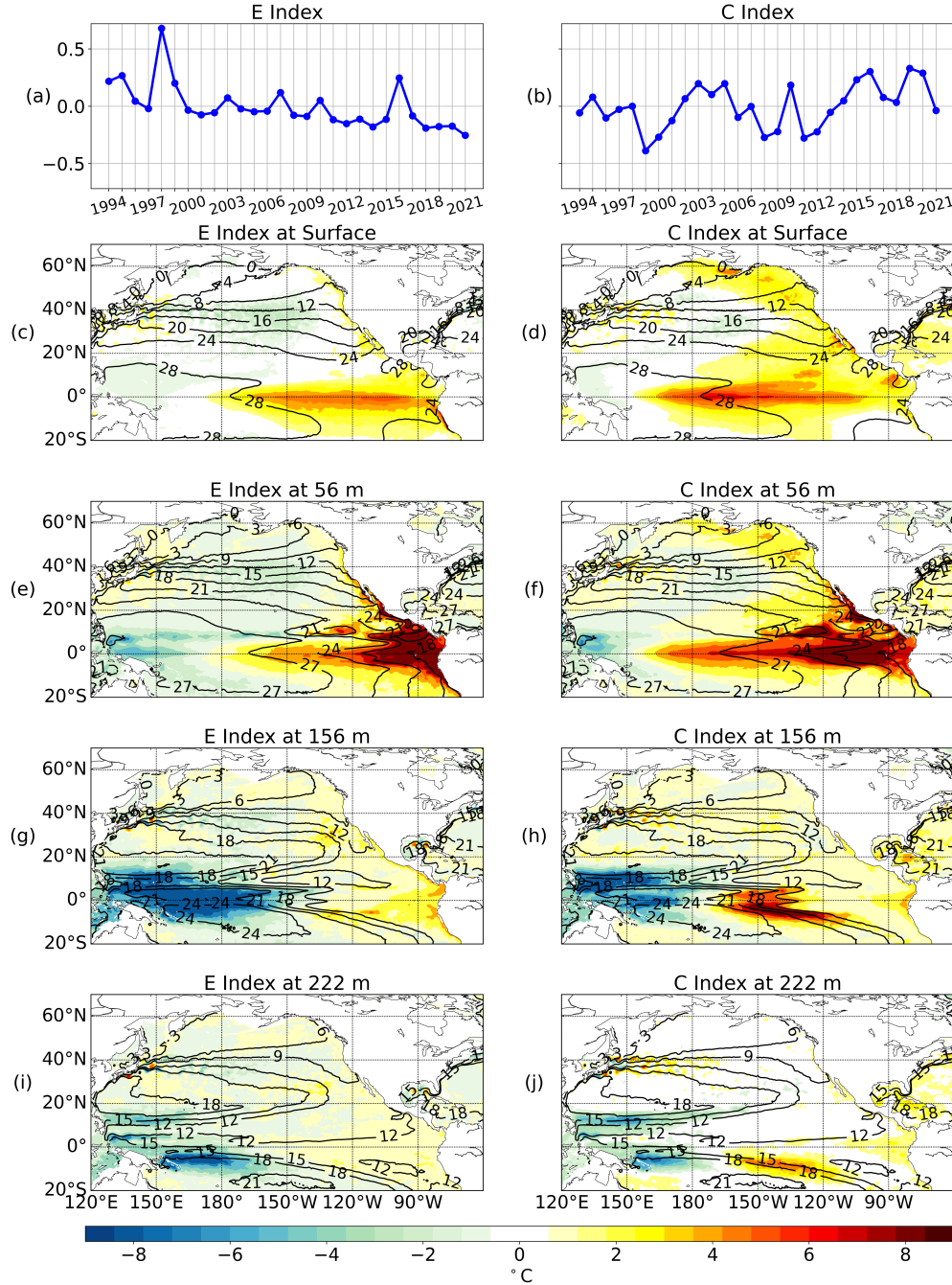


Figure 4.3: Same as Fig. 4.1 but for the E and C index.

covering a larger longitude but smaller depth range than EOF 1. The warm half of the dipole has a much weaker signal than EOF 1 and ends in shallower layers (200 meters). Below 200 meters, the eastern Pacific contains only weak warming. The E pattern's eastward bias is also reflected in the maximum warming ($\sim 40 - 100$ meters, $\sim 155^\circ - 80^\circ\text{W}$), which never

extends past 120°W as EOF 1 does. However, like EOF 1, maximum variability for both halves of the dipole are in the thermocline, with the western maximum lying deeper than the eastern maximum, appropriately characterizing the thermocline slope described in the previous section.

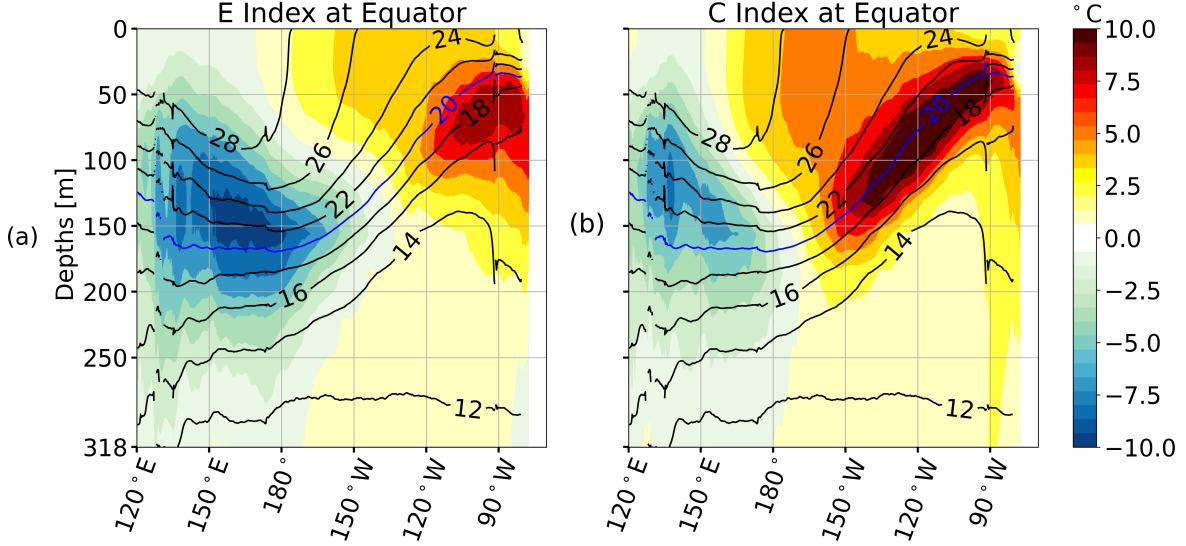


Figure 4.4: Same as Fig. 4.2, but for the E and C index.

As mentioned previously, the CP El Niño is characterized by maximum anomalies near the dateline. The surface of the C pattern (Fig. 4.3(d)) has a strong signal from 160°E - 110°W in the equatorial Pacific spanning various Niño indices. Consequently, the C time series correlates most with the westernmost index Niño 4 ($r = 0.90$), followed by Niño 3 ($r = 0.64$), while Niño 1 + 2 has no statistically significant relation. As reflected in the correlations and surface expression, the C index can capture the distinct central Pacific variability expected from CP El Niño. We also compare C time series to the known El Niño Modoki index described by [6] as an area-average of multiple Pacific regions defined to capture the cold-warm-cold loadings in CP El Niño. The three areas are: A (10°S - 10°N, 165°E - 140°W), B (15°S - 5°N, 110° - 70°W), and D (10°S - 20°N, 125° - 145°E). Both regions B and D are expected to have negative loadings at the surface, but our C pattern's surface expression shows weak to no signal in these areas. However, our time series still indicates a considerable relation to the EMI ($r = 0.83$). Our C index time series also captures the CP

El Niño events in its maxima (2005, 2010, 2015, 2019, and 2020), and satisfies the definition in reference [32], capturing La Niña events in its minima (1999, 2008, and 2011).

The C spatial pattern subsurface (Fig. 4.4(b)) contains a strong warming signal from the dateline to 140°W in the upper 30 meters. From 30 - 50 meters, this warming extends eastward to span the entire eastern Pacific ($\sim 180^{\circ} - 80^{\circ}$). After 50 meters, this warming begins to bottleneck, coming to a point centered at 150°W and ending at 175 meters. Though the overall shape of the warming's vertical structure is consistent with other 2D subsurface analysis [13, 14, 96, 100], our results differ in the location of the maximum warming. The 2D studies present a maximum signal in the first 100 meters starting as west as 160°E and ending as far east as 150°W . Our 3D results for the C pattern indicate strong positive loadings in the central Pacific for the first 30 meters, but the strongest signal occurs in the thermocline. Though the central Pacific warming is not a maximum, it captures the tendency for CP El Niño to favor the central Pacific. Furthermore, the maximum warming at the thermocline is narrow, suggesting small vertical thermocline displacement in the central to eastern Pacific.

By comparing the subsurface of the C and E patterns, we can leverage their characteristic signatures to define regions that may help differentiate the two El Niño types. The cross-section in Fig. 4.4(a) indicates the maximum signal of EP El Niño is found in the upper 100 meters at the thermocline from $\sim 110^{\circ} - 80^{\circ}\text{W}$. The same zonal cut for the C pattern has a maximum warming along the thermocline (40 - 140 meters) further west ($140^{\circ} - 90^{\circ}\text{W}$). The maximum E and C pattern warming signals experience the least overlap at 50 - 140 meters between $140^{\circ} - 110^{\circ}\text{W}$ and at 50 - 100 meters from $90^{\circ} - 80^{\circ}\text{W}$. Subsurface analyses have described the $90^{\circ} - 80^{\circ}\text{W}$ region as an area of interest for a subsurface EP index [13], but their definition uses a larger depth range (surface - 100 meters). Temperature anomalies in the region (50 - 140 meters, $5^{\circ}\text{S} - 5^{\circ}\text{N}$, $140^{\circ} - 110^{\circ}\text{W}$) could be used to supplement/distinguish the subsurface CP index.

Though our focus has been primarily on the tropical Pacific, the North Pacific contains substantial variability. Both E and C patterns display warming along the coast of North America in their surface pattern expressions (Figs. 4.3(c) and (d)). The characteristics of the northeast Pacific (NEP) warming are vastly different: the E pattern is confined to the coast, while the C pattern extends further west. These surface expressions, consistent with [11], exemplify the distinct effects of El Niño diversity on NEP temperature variability. To explain the 3D structure of the NEP variability, we visualize the U.S West Coast (USWC) from California to Washington in the next section (Fig. 4.5).

4.3.3 Regional Expressions in the California Current System

The California Current System (CCS) is home to a highly productive marine ecosystem that supports a wide range of economically and culturally important human activities. Its sensitivity to ENSO has been widely documented (e.g., [15, 82]), though the 3D structure of ENSO’s influence, as well as the difference between CP and EP El Niño, have received relatively little attention. The CCS experiences warming during the warm phases of both CP and EP El Niño. Their differences in the spatial distribution of their SSTA result in differing dynamics and influence on the CCS [11, 97]. The mechanisms that drive each dynamic are reflected in the 3D temperature variability structure of the E and C index patterns.

The coastal warming of the E index shows differences in behavior depending on latitude. The E index (Fig. 4.5(a)) indicates warming ($\sim 2^\circ\text{C}$) north of 36°N that does not extend past 100 km from the coast. South of 36°N , there is weaker warming ($\leq 1.6^\circ\text{C}$), and the rest of the region contains little to no warming. The EP El Niño intensifies at 56 and 78 meters (Figs. 4.5(c) and (e)). Just like the surface, these subsurface waters (56 and 78 meters) experience the strongest warming near the coast, with different rates of increase deviating at 36°N . To exemplify this, we compute the coastal temperature increase from surface to

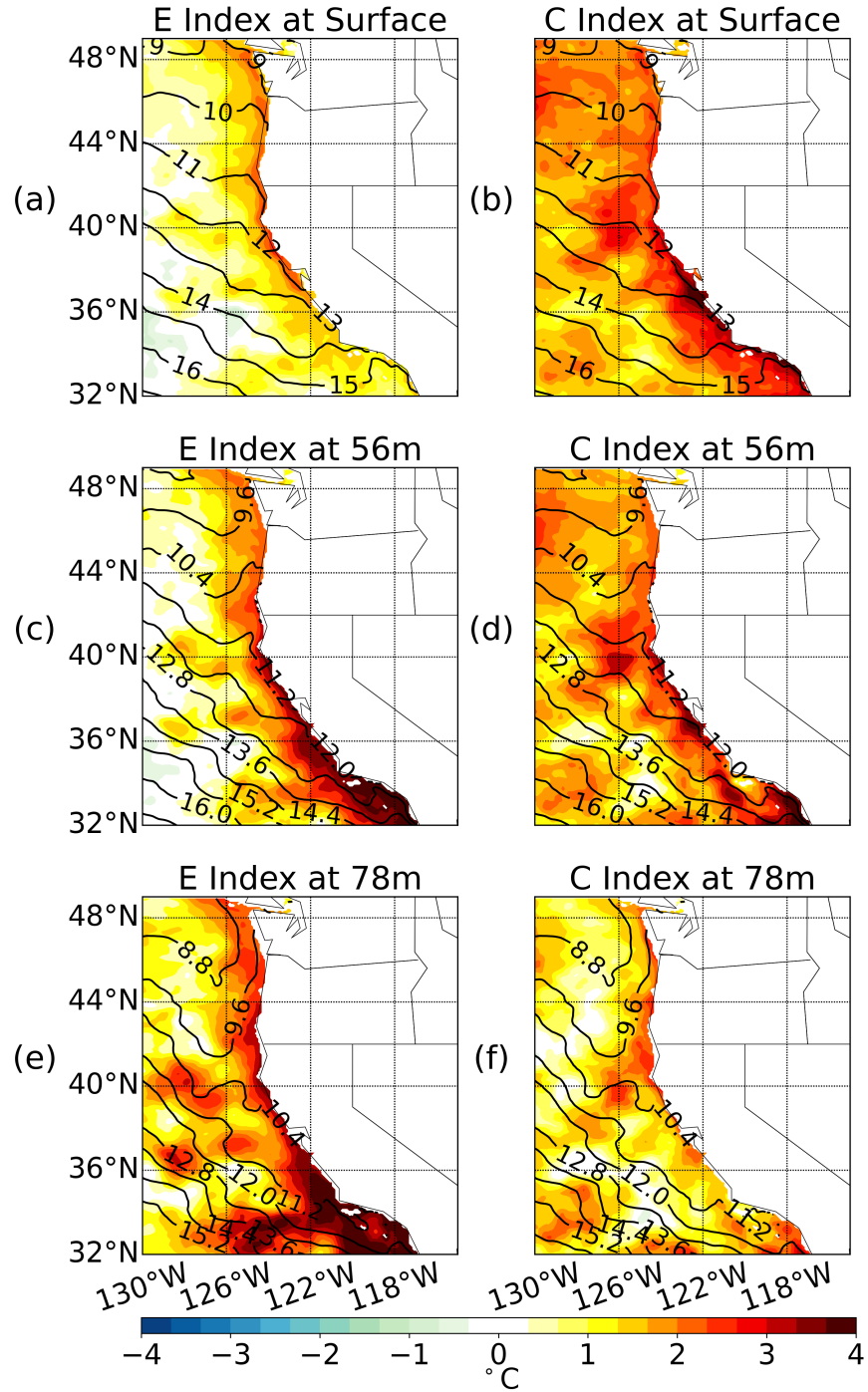


Figure 4.5: The (left column) E and (right column) C index for the U.S West Coast from surface to 222 meters. Layers are specified according to GLORYS depth layers. These depths, longitudes, and latitudes were chosen for the interpretation of EP (E index) and CP (C index) El Niño and their effect on the CCS and USWC.

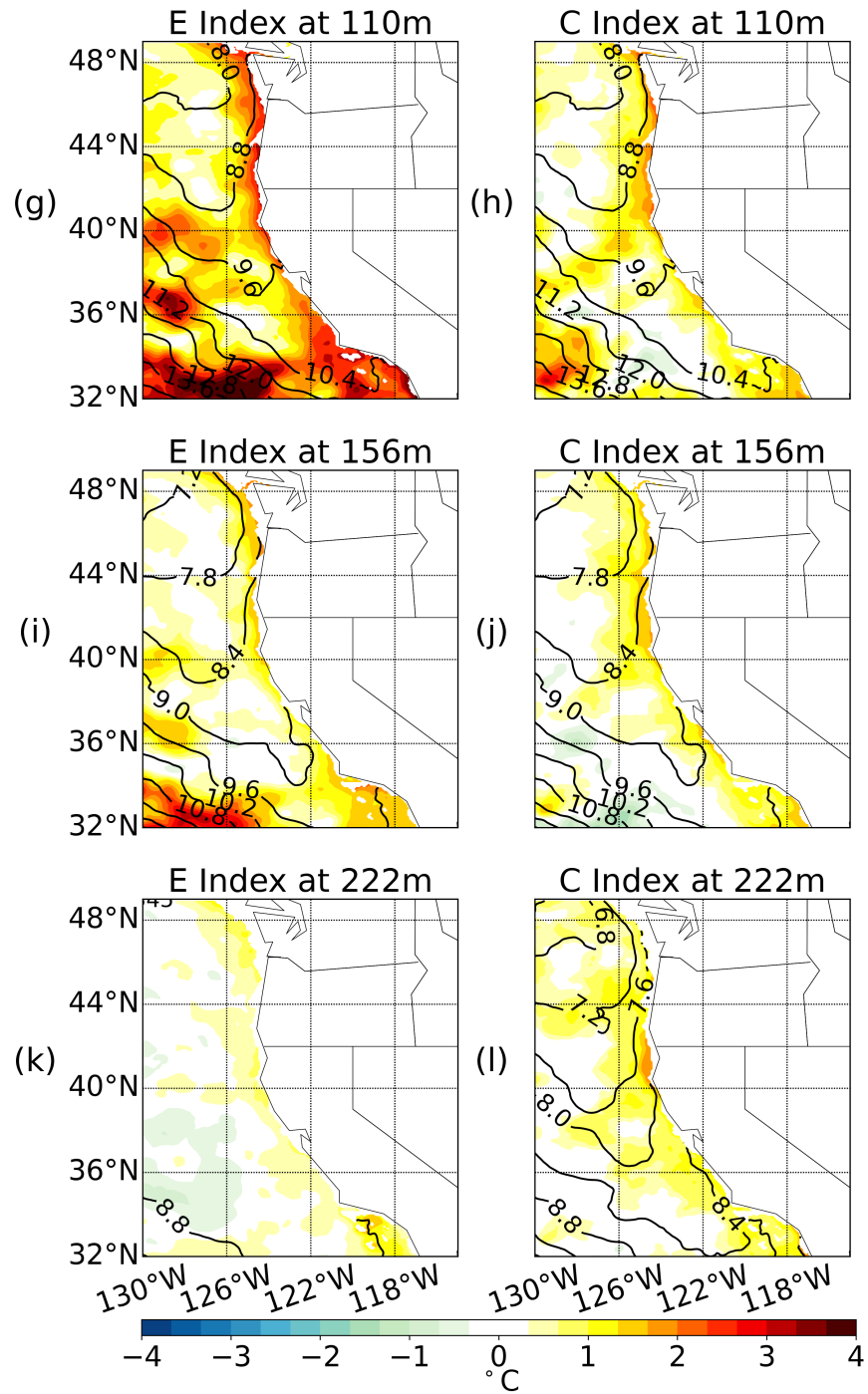


Figure 4.5: The (left column) E and (right column) C index for the U.S West Coast from surface to 222 meters (continued).

the beginning of maximum warming at 50 meters from two cross-sections in Fig. 4.7. We choose one cut below 36°N at 34°N and one above 36°N at 37°N . The southern cut shows an increase of 2.7°C from 1.3°C to 4°C , while the northern cut has a smaller increase of 1.6°C from 3.6°C to 2°C . Meaning, the vertical coastal temperature gradient is smaller north of 36°N .

The deviation in coastal warming behavior at 36°N also includes the area size of maximum warming. The zonal cross-sections along the USWC (Figs. 4.6 and 4.7) show how far west the coastal warming will reach in the northern (north of 36°N) and southern (south of 36°N) halves of the USWC. Each zonal cut of the E index shows maximum warming between 50 - 100 meters. Maximum warming in the cut at 34°N is between 122° - 119°W , while the maximum at 37°N is between 123° - 122°W . The cut at 34°N has a small interruption at 199°W , but a cut at 33°N (not shown) has the same longitude range for the maximum warming. With higher latitudes, the warming in the CCS narrows to the coast, and the vertical temperature gradient weakens.

The CCS warming signature is consistent with coastally trapped wave behavior. EP El Niño produces oceanic kelvin waves that propagate eastward along the equator; once they reach the eastern boundary, they continue poleward as coastally trapped waves. These waves will deepen the thermocline, causing large anomalies at 50 - 100 meters [15]. We identify the thermocline in the zonal cuts of Fig. 4.6 as the closely spaced climatology isotherms in the same depths (50 - 100 meters). Maximum warming occurs at the thermocline, consistent with vertical thermocline displacement from EP El Niño's coastally trapped waves [15]. As mentioned previously, northern cuts have a weaker and smaller coastal warming signal compared to the one at 34°N , decreasing with higher latitudes. This coincides with coastally trapped waves attribute to weaken with higher latitudes and decay with distance from the shore [101].

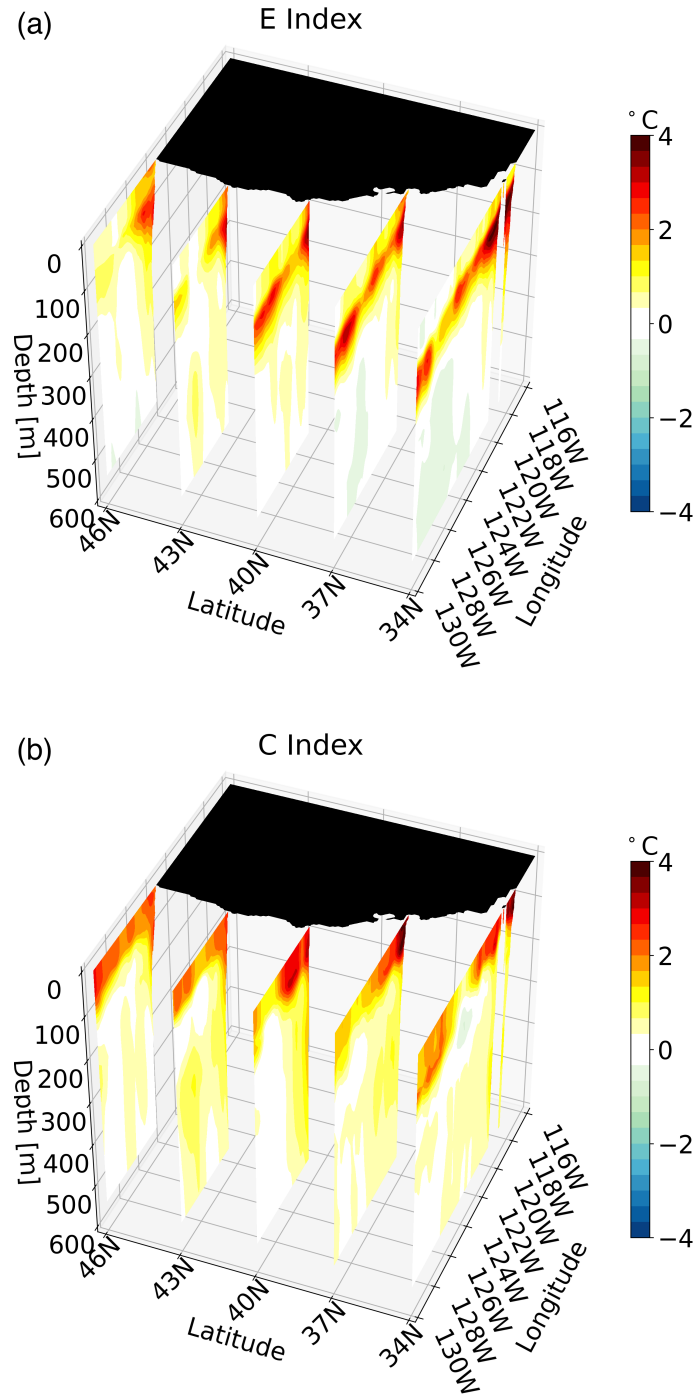


Figure 4.6: Zonal cross-section cuts from the USWC of the (a) E and (b) C index. Cuts were taken from 34°N to 46°N in increments of 3° to capture the effects of both El Niño types up the coast. Longitudes were chosen to match Fig. 4.5.

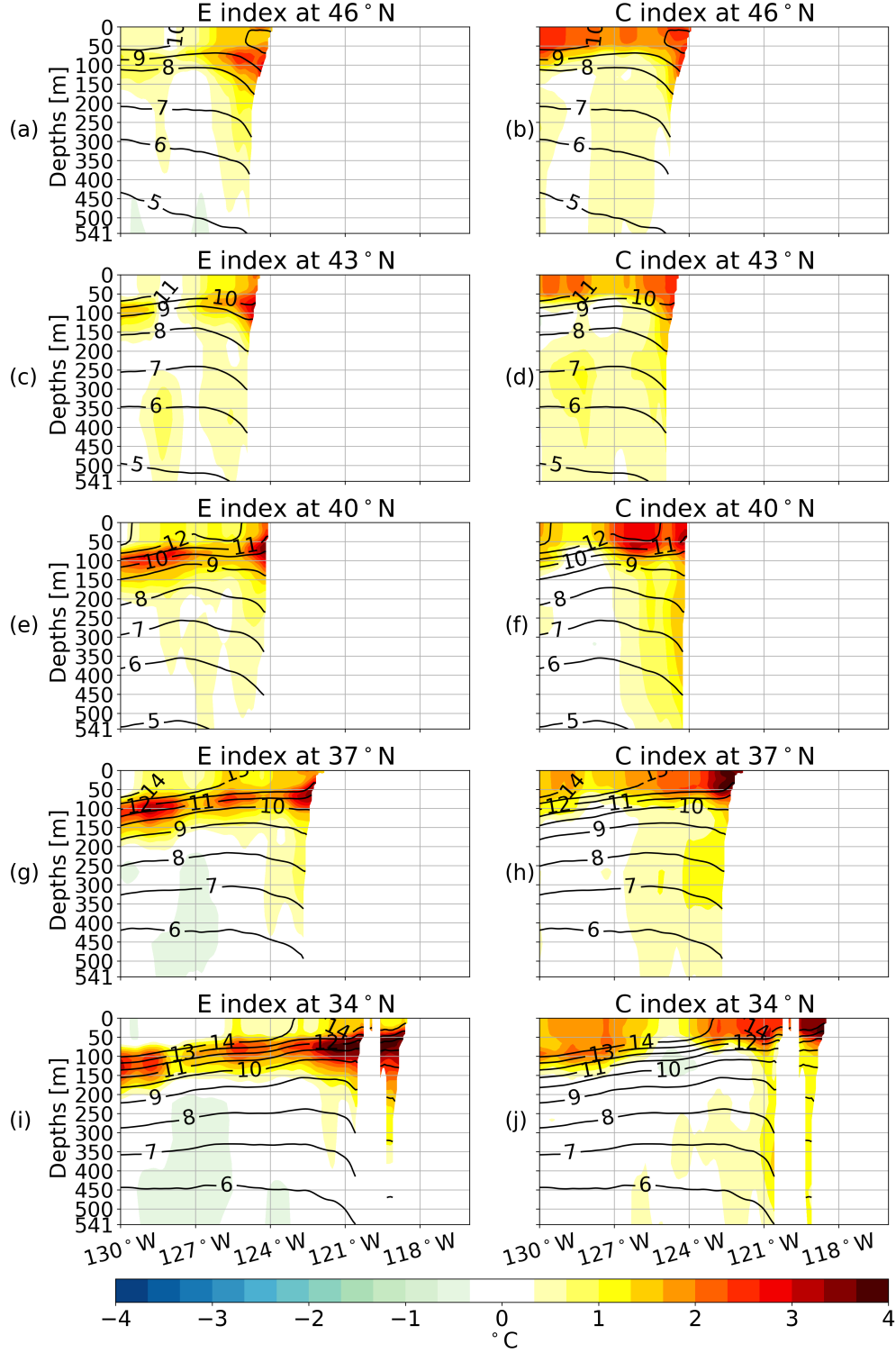


Figure 4.7: The same Zonal cross-section cuts from the USWC of the (left column) E and (right column) C index as in Fig. 4.6, visualized in 2D for clarity. The left column is the E index, and the right column is the C index. Cuts were taken from 34°N to 46°N in increments of 3° to capture the effects of both El Niño up the coast. Panels were ordered top to bottom from north to south. Longitudes were chosen to match Fig. 4.5. Black contours in are the GLORYS boreal winter climatology

EP El Niño's influence extends beyond coastal warming. In Fig. 4.5(e) there are two climatology isotherms that bend into 'S' shapes. The most obvious is the 10.4°C isotherm that bends at 40°N from 127° - 125°W. The other is the 11.2°C isotherm that bends at 126°W from 36°-38°N. These curved isotherms result from meanders in the CCS that cause anticyclonic and cyclonic eddies [3,102]. In the E pattern at 78 meters, we observe two warm ovals in areas of known meanders [3]. The largest warm oval is centered 40°N and 128°W, and a smaller oval is centered at 37°N and 125°W. We identify the depth location of the ovals from the zonal cuts at 37°N and 40°N (Figs. 4.7(e) and (g)). The southern oval (37°N) has warming of $\sim 1.3^\circ\text{C}$ in the first 30 meters at $\sim 126^\circ\text{-}125^\circ\text{W}$ and reaches its maximum at the thermocline at 60 - 80 meters between $\sim 126^\circ\text{-}125.5^\circ\text{W}$. The northern oval (40°N) also has its maximum at the thermocline, but its warming blends into the rest of the thermocline warming in Fig. 4.7(e). Using a smaller depth range and coloring to amplify the signal (Appendix Fig. B.1), we locate the warming. The north oval has consistent warming in the first 50 meters between $\sim 128.5^\circ\text{-}127.5^\circ\text{W}$ and reaches its maximum at 75 - 100 meters between $\sim 129^\circ\text{-}127^\circ\text{W}$. During strong El Niño events, the meanders tend to weaken [21]. This disruption could explain the temperature variability seen in the E index at the exact location of known meanders.

Compared to the E index warming, the C index surface (Fig. 4.5(c)), contains stronger warming with smaller zonal and meridional temperature gradients. North of $\sim 41^\circ\text{N}$, all warming is primarily within $1.6^\circ\text{-}2.3^\circ\text{C}$, a noticeable difference from the E pattern range in the same region ($0.3^\circ\text{-}2^\circ\text{C}$). South of 41°N we start to observe lower offshore surface minimums of 1.3°C , and south of 37°N a maximum warming of 4°C on the coast. That is, with southern latitudes, the USWC will experience a larger zonal temperature range. However, even at its largest range (36°N , $1.3^\circ\text{-}4^\circ\text{C}$), the C pattern is smaller than the E pattern's temperature range in the same region ($-0.6^\circ\text{-}2.6^\circ\text{C}$).

Warming in the C pattern decreases after initial surface warming throughout most of the USWC region except the coast (Fig. 4.5(d)). Most coastal areas at 56 meters have remained the same or increased. After 56 meters, coastal warming gradually decreases while the rest of the region decays rapidly to show little variability. The zonal cuts of the C index (Fig. 4.7) reveal that coastal warming experiences its strongest intensity from surface to 50 meters and remains within 1° longitude from the coast. The C index maximum warming is shallow compared to the maximum warming in EP El Niño. The differing dynamics attribute to this: CP El Niño is not associated with strong coastally tapped waves like EP El Niño [103]. Therefore, it does not experience the same thermocline deepening as EP El Niño that causes large variability at 50 - 100 meters and instead has its strongest warming near the surface.

The limitation of the dominant coastal warming to the upper 50 meters does not translate to weaker coastal warming for the C index. Excluding the northernmost cross-section in Fig. 4.7 due to the large amount of non-values in deep coastal waters, all other cuts show the C index has warming of at least 1°C in deep layers ranging from 350 - 541 meters (our analysis depth limit). All zonal cuts whose 1°C warming ends before 541 meters (37°N and 40°N) have a weaker warming of $0.3^\circ - 0.6^\circ\text{C}$ until 541 meters. The E index does not have this same coastal warming feature. The end of the E pattern coastal warming ranges from 400 - 500 meters, ending in shallower layers than the C index warming. Overall, coastal warming along the USWC driven by CP El Niño is most intense at the surface and extends significantly deeper than that associated with EP El Niño.

4.4 Conclusion and Discussion

In this study, we quantitatively described the 3D structure of the dominant modes of ocean variability in the North Pacific from surface to 541 meters. We calculated 3D EOF on the eddy-resolving high-resolution GLORYS ocean temperatures for boreal winter to capture

ENSO and the NPMM in the first two EOFs. Surface patterns align with previous 2D results from [32] and [98], while subsurface results showed consistencies with the known dipole structure of ENSO and the trade wind charging mechanism of the NPMM. The 3D PCs are significantly correlated with surface-based Niño and NPMM indices, which support using 3D results to extend 2D methods.

We combine our first two 3D PCs and EOFs to define two new orthogonal indices that separate CP and EP El Niño. The resulting EP El Niño index (E index) showed a spatial pattern favoring the eastern Pacific, while the CP El Niño index (C index) exhibited characteristic central Pacific variability. Both E and C index time series satisfy their defined behaviors from the original [32] study: the E index explains strong El Niño events, and the C index captures the warm and cold CP events. Using the multi-layer E and C indices, we identified regions with maximum variability where the two El Niño differ most. We propose that these regions could help identify the two El Niño types and supplement other subsurface indices such as [13].

Our use of high-resolution ($1/12^\circ$ latitude-longitude) data allowed for detailed analysis of regional effects from both El Niño types, specifically along the U.S. West Coast (USWC) and within the California Current System (CCS). Our regional analysis reveals significant differences in how EP and CP El Niño affect the CCS. During positive EP events, warming is confined to the coast, reaches its maximum in the thermocline, and presents latitude-dependent behaviors. This response aligns with EP El Niño’s oceanic forcing through coastally trapped waves on the USWC. In contrast, CP events have a broader warming that reaches further west, and reaches its maximum in the first 50 meters. Although the coastal warming is weakened after 50 meters, it persists into layers deeper than the broader region, influencing depths greater than the EP coastal signal.

While our analysis provides valuable insights into regional expressions of El Niño diversity, we have not fully exhausted all potential coastal analyses enabled by the high-resolution

GLORYS. Future work could include a detailed 3D analysis of other coastal areas like Central America or Baja California, which have known connections to EP and CP El Niño [81, 97, 100, 104]. Expanding our focus to other seasons would also contribute to our interpretation of the USWC variability and potentially reinforce the theorized relationships of USWC warming/marine heat waves and El Niño diversity [11]. Therefore, further applications of our methods would advance understanding of important climate phenomena.

Chapter 5

Development of 3D Visualization Tools for Ocean Data

5.1 Introduction

Considerable effort has been dedicated to developing visualization tools for climate data that are accessible to the general public. These tools feature simple and intuitive interfaces designed for users without coding expertise, making them excellent educational resources that encourage public engagement with climate science. However, a user-friendly interface typically comes at the cost of increased coding complexity, making these applications difficult to adapt for other datasets unless they were originally designed with such flexibility in mind. This limitation became evident during our research. We needed a convenient tool that could instantly and accurately visualize our results through various cross-sections while remaining flexible enough to handle ocean temperature data with different resolutions and domains.

Many web-based applications exist that are exceptional in visualizing climate data. Notable examples include NASA's Geospatial Interactive Online Visualization ANd aNalysis

Infrastructure (Giovanni) [105], NASA’s Worldview [106], NOAA’s system of Web-based Re-analyses Intercomparison Tools (WRIT) [107], and the 4D visual delivery (4DVD) [108,109]. While each serves slightly different purposes, outlined in Table 5.1, their goals can be simplified as: visualize downloadable data to facilitate data access and analysis. Although their objectives differed from ours, these software applications served as valuable references for our tool development. They inspired us to develop a simple dashboard that would feature a combined interactive and visualization interface, buttons that instantly update visualizations, options to compare data, and an appropriate color palette for clear visual storytelling.

Web applications provide the convenience of ready-to-visualize datasets without requiring local data storage, though users have the option to download the complete datasets or subsets if needed. This convenience comes with a limitation: users are restricted to the provided datasets and cannot utilize these tools on their own data. As an alternative, we turned to downloadable tools such as Panoply [110] and Cross-Sectional Layers with Interactive Cropping and Exportation (XSLICE) [1]. Both applications can scan through depth, zonal, or meridional cross-sections and produce highly customizable figures through cropping and colorbar customization. For a comprehensive comparison of these tools’ purposes and capabilities, see Table 5.1.

Panoply’s ability to visualize any climate data of NetCDF or HDF file types aligned perfectly with our needs. In practice, Panoply has significant drawbacks. The system takes substantial time to generate separate interface and visualization windows, creating a time-consuming workflow that is problematic when several datasets require visualization. Additionally, the interface window frequently crashes, leaving only the visualization with no means of interaction. Beyond workflow issues, the interface is complex, requiring direct text inputs for most interactions, including cropping and cross-section scanning. XSLICE addresses these problems through its user-friendly dashboard that contains both an interactive and visualization

interface. However, in its current version, XSLICE is tailored to GODAS data and lacks the capacity to incorporate other datasets easily.

Building upon these existing applications, we developed a 3D visualization tool called XSLICE for GLORYS. This is a simplified version of XSLICE designed specifically for visualizing various datasets for analysis. Its usage is demonstrated in this chapter using the high-resolution Global Ocean Physics Reanalysis (GLORYS) ocean temperatures for the global ocean, hence the name. The combination of three key features distinguishes XSLICE for GLORYS from the aforementioned tools:

- A custom colorbar that effectively represents ocean dynamics and can be manipulated to enhance climate patterns.
- The ability to visually scan through zonal, meridional, and depth cross-sections and display multiple cross-section figures simultaneously.
- A user-friendly interface built on simple and adaptable Python code, facilitating analysis of various datasets.

The resulting tool streamlined the process of exploring and quantifying climate variability for our research. It also produced initial figures that we subsequently refined and compiled into 2D and 3D figures featured in our published papers [23,24], and throughout this dissertation.

The purpose of this chapter is to describe the motivation, development process, functions, and applications of XSLICE for GLORYS. The tool was applied to the EOF results from Chapter 3 to demonstrate the tool’s usage. The descriptions and examples illustrate the value each function provided for our analysis. All code examples, as well as XSLICE for GLORYS, are written in Python, a recently popular computer language used in climate research. In this way, the following tools are accessible to a broader coding audience.

The remainder of this chapter is organized as follows: Section 5.2 outlines the creation process for the colorbars used in this study, Section 5.3 describes the development of XSLICE for GLORYS and its four interfaces, and Section 5.4 provides a summary and discussion.

5.2 Customize Colorbars for Optimal Display of Ocean Dynamic Features

The colorbar is a critical component for visualizing climate data, as inappropriate color representation can lead to misinterpretation of results. There are two key aspects of the colorbar that we manipulate to influence this interpretation. The first is the colormap that systematically arranges colors later used to represent data values. The second is the colormap normalization that linearly maps data values to colors in a colormap. We will see examples of the use and misuse of either in this chapter.

The creation, regulation, and application of colormaps and colorbars are accomplished using the Python library *matplotlib* [111]. This library offers multiple colormaps across four categories: sequential, diverging, cyclic, and qualitative. Sequential colormaps progress in lightness and saturation, often using a single hue. Diverging colormaps change in lightness and saturation for two hues that meet in the middle at an unsaturated color (i.e., two sequential colormaps combined). Cyclic colormaps begin and end with identical hues while varying in lightness throughout the intermediate range. Qualitative colormaps employ distinct, unrelated colors with no inherent order.

We tailor the colormaps to our EOF results computed from ocean temperature anomalies. Since our EOFs are centered at zero, we use a diverging colormap. Our analysis explains negative EOF values as cold signals and positive values as warm signals. Accordingly, we employ cool tones (e.g., blue and green) for negative values and warm tones (e.g., yellow and red)

Table 5.1: Summary of published visualization tools for climate data. Applications appear in the order they are referenced in this Chapter.

Name	App type	Purpose	Features & advantages
Geovani	Web-based	Facilitate data accessibility and analysis by visualizing large satellite data users can download	Handles a large amount of satellite data. Produces various map plots (time-averaged, recurring), comparisons (correlation, scatter), time series, vertical cross-sections, and provides unit conversions
Worldview	Web-based	Present visualizations of current satellite data that engage a general audience through storytelling. Also used to identify data to download for further analysis	Updates satellite data frequently. Tool can easily overlay multiple datasets in one map (e.g., SST and precipitation), and dynamically compare data. Contains story-based tours that present data through notable events.
WRIT	Web-based	To quickly test hypotheses or plot basic climate data	Includes multiple tools for re-analysis plots (maps, anomalies, cross-sections), time-series, correlation analyses, trajectories, distribution, profiles, and Hovmoller analyses
4DVD	Web-based	Convenient and fast visualization of large space-time data. Allows users to identify and download data subsets necessary for analysis	Visualizes large climate datasets, produces maps/time series, 2D and 3D visualization options, data download, and statistical computing
Panoply	Downloadable application	For data viewing. Used to explore and analyze any data file of type HDF or NetCDF	Display users' data, visually scans depth/zonal/meridional cross-sections, data computing, wind vector plots, time series plots, and difference maps
XSLICE	Downloadable source code to run application	To quickly identify significant climate dynamic patterns or significant anomalies for further analysis on 4D data.	Simple user interface that visually scans depth, zonal, and meridional cross-sections for GODAS data. Has tools to produce difference plots, time series, time averages, and Hovmoller analyses.

for positive values. Although *matplotlib* provides various diverging colormaps incorporating these colors, they lack sufficient contrast to adequately represent our results. Therefore, we developed a customized colorbar by combining sequential colormaps with additional light transition colors at their convergence point.

The first iteration of our custom colorbar was created by combining *winter* with a reversed *hot* colormap (Fig. 5.1). This coloring was used for all two-dimensional (2D) figures¹ in Chapter 2, such as Fig. 2.4. For each colormap, the ‘under’ and ‘over’ outputs specify colors for values extending beyond the colorbar limits, while the ‘bad’ output designates the color for non-number (NaN) values. Since land areas are represented as NaN values in the EOFs, the ‘bad’ color determines their appearance.

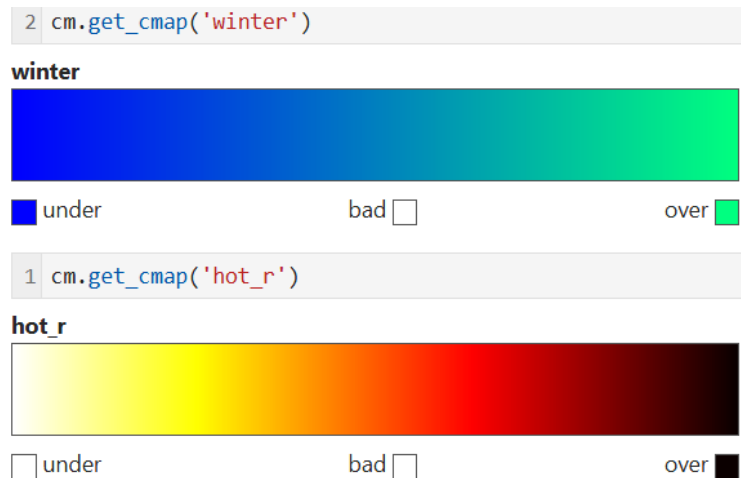


Figure 5.1: The two colormaps used for the custom colorbar in Chapter 2. The winter colormap is used for all negative EOF values, and the hot colormap is for all positive EOF values. The ‘under’ and ‘over’ boxes specify the colors of any values out of bounds in the colorbar, and ‘bad’ colors are for any NaN values.

The aggregation of both colormaps requires some processing. Each colormap stores its color values ranging from 0 to 1 for red, green, blue, and transparency in an $L \times 4$ array, where L is the length of the colormap. The array is accessed by calling the colormap with an array of values that match the desired length of the colormap. For the custom diverging colormap, each array will be half the length of the complete colormap. This step is crucial, as the

¹2D refers to the visual representation format, not the nature of the calculation.

colorbar will contain only a total of L colors. Once arrays are defined for each colormap, they are stacked vertically and converted back into a colormap object. The process of stacking the arrays is what merges the colormaps.

As an example, we create a diverging colormap that will contain 20 colors using the two colormaps in Fig. 5.1. The result is a simplified version of the colormap in Chapter 2. Once we save the *winter* and *hot* colormaps as objects matching their respective name, the color array can be accessed using:

```
colormap object(np.linspace(min_c_val, max_c_val, L))
```

where `min_c_val` is the minimum color value (typically 0), `max_c_val` is the maximum color value (typically 1), and L is the length of the colormap. For the *winter* and *hot* colormaps, this becomes:

```
winter(np.linspace(0, 1, 10))  
hot_r(np.linspace(0, .9, 10))
```

To retain all values in *winter*, the minimum color value is 0, and the maximum is 1. The *hot* colormap's end color is excessively dark, so the final color is removed by setting the maximum color value to 0.9. The final parameter specifies that *winter* and *hot* will produce 10 colors each. To combine the two, the color arrays are stacked vertically and then converted into a colormap object using matplotlib's *ListedColormap* function. Figure 5.2 displays the Python code and resulting 20-color colormap.

For improved gradients, the number of colors in the array can be increased. Chapter 2 figures demonstrate the custom colorbar from Fig. 5.2 using 256 colors, which results in a smooth, continuous color transition. Alternatively, figures in Chapter 4 utilize 30 colors, producing abrupt color transitions that make it easier to distinguish various climate modes and their effects.



Figure 5.2: The result of combining two colormaps to create a custom colormap. The Python code demonstrates conversion of colormap values into color arrays, vertical stacking of the arrays, and conversion back into a colormap object.

Implementing the colormap in a colorbar requires specifying the data values the colors will represent. All colormaps are scaled to reflect data from 0 to 1, and colorbars follow this convention accordingly. Creating a colorbar with the map from Fig. 5.2 would produce a scheme where 0 corresponds to dark blue, 0.5 represents the transition between green and white, and 1 appears as maroon. To modify this scale, we first employ matplotlib’s *Normalize* function, which creates an object that linearly maps any interval $[a, b]$ to $[0, 1]$. This function assures that the minimum value a aligns with the color value at 0, the maximum value b corresponds to the color value at 1, and all intermediate values receive their appropriate color assignments. The *ScalarMappable* function then maps the normalized data scale onto the colormap indices. This function serves as the essential component in connecting data values to colors and ensures that a colorbar presents the correct $[a, b]$ interval. In practice, this process is handled by the Python functions, but the parameters for the *Normalize* function require some thought.

Selecting an appropriate range requires careful consideration. The values a and b do not need to represent the true minimums and maximums of the data, as the colormap will display any values beyond these limits using the ‘under’ or ‘over’ colors. An optimal interval will accommodate most data values, and is visually apparent when it does not. A range that is too small causes all values to appear as ‘under’ or ‘over’ colors, and one that is too large will

only visualize outliers. Our results take great care in choosing an appropriate interval that will highlight large climate variability without exclusively amplifying the pattern’s signal.

As an example of the process for creating a colorbar, we use the same colormap as in Fig. 5.2. Assuming zero-centered data, as in our EOF analysis, the normalization interval must be $[-b, b]$. Setting the range so the minimum is the negative of the maximum ensures the colormap will diverge at zero with equal amounts of colors on either end. After generating random data, we select an arbitrary range $[-.3, .3]$ to define the normalization interval (line 6 of Fig. 5.3). To create a figure with the custom colorbar (Fig. 5.3), we pass the colormap object as an argument twice: once to create the contour plot of the data, and once to generate the colorbar.

The colormap requires creation only once for use across multiple figures. When results vary significantly between figures, as in Figs. 2.4(a) and (b), redefining individual colorbar limits and the corresponding normalization object is necessary. This assures appropriate color representation of climate dynamics for easier quantification of pattern structures.

The previous examples present a simplified version of the custom colorbar in Chapter 2. The colors were chosen to have distinct warm and cold tones to identify regions of strong variability with ease. However, their high contrast visually amplifies weak cold and warm signals. To better represent small negative values, we incorporated a transition color in the cold half of the colormap that mirrors the white in the warm half. This modification involved adjusting and refining color values through trial and error at the end of the cold color array. To resolve the highly saturated blue and green colors, the *winter* colormap was replaced with the *green blue* colormap, which contains lighter cool tones. Figure 5.4 compares both versions for 3D EOF 3 at a depth layer (186 meters) exhibiting substantial weak cooling. The reduced contrast in the revised colormap enhances the stronger signals throughout the map (e.g., warming in the tropical Pacific). The coloring minimizes visual clutter, allowing the user to focus on the structure of the dominant climate variability. This revised colorbar

```

3 data = np.random.randn(30, 30) # create random data
4 fig, ax = plt.subplots(figsize=(6, 5), layout='constrained') # creating an empty figure for the colorbar
5 ax.pcolormesh(data, cmap=newcmp2) # plot random data
6 norm = mpl.colors.Normalize(vmin=-.3, vmax=.3) # defining the maximums and minimums of the colorbar
7 cbar = plt.colorbar(mpl.cm.ScalarMappable(norm=norm, cmap=newcmp2),
8                     ax=ax, orientation='vertical') # creating colorbar with parameters and placing of figure

```

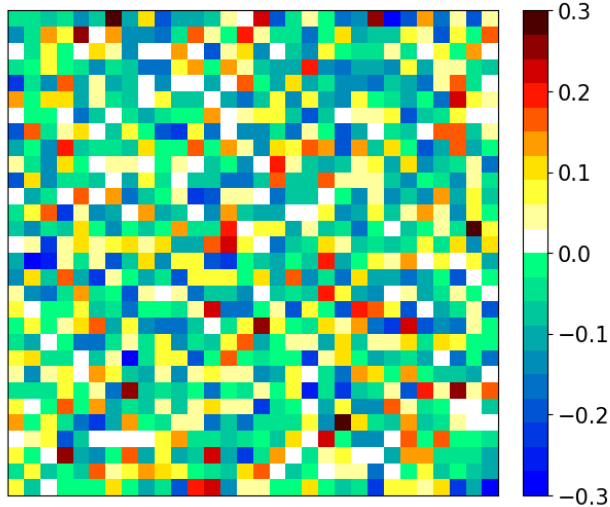


Figure 5.3: The colorbar implementing the colormap created in Fig. 5.2 in the range $[-.3, .3]$. The Python code generates random data, creates the figure for the visualization, applies the colormap to visualize data in a contour plot, and then applies the colormap to the colorbar.

is implemented in Chapter 3, with reduced colors in Chapter 4, and in the XSLICE for GLORYS tool described in the following section.

5.3 Customizing XSLICE for GLORYS: Creating a Visualization Interface for Three-Dimensional Data

Creating each figure in the previous Chapters required consideration for which depths, cross-sections, or regions best represent the variability of each climate mode. With this in mind, in tandem with the 3D EOF algorithm, we developed a computer tool to quickly visualize four-dimensional (4D) gridded space–time climate data called XSLICE [1]. Like a CT scan of a human body, XSLICE takes a cross-sectional scan of the ocean or atmosphere by fixing any two coordinate variables from 4D (x, y, z, t) climate data. Here, x indicates longitude,

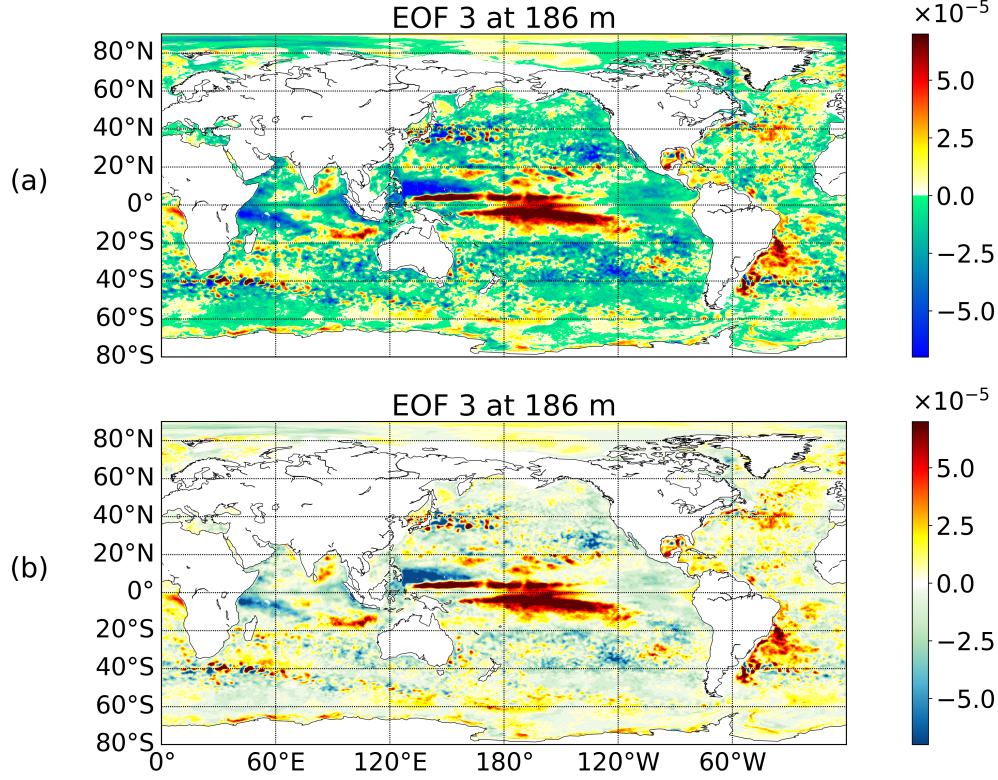


Figure 5.4: (a) Application of the colorbar from Fig. 5.3 with increased color values. (b) Updated colorbar featuring new cool tones and central transition colors. EOF 3 is displayed at 186 meters for both examples to compare the visual representation of weak cooling for both colorbars. The weak cooling appears amplified in (a) due to the saturated green tones, while the same cooling in (b) exhibits reduced saturation, better representing the true value.

y indicates latitude, z indicates depth, and t indicates time. Using sliding control buttons, XSLICE allows users to scroll through different altitude layers and vertical latitude–altitude or longitude–altitude cross-sections. It greatly simplifies diagnostic explorations of climate model data and aids in the efficient detection of climate patterns.

For example, when fixing coordinate variables y and t , as in Fig. 5.5, we visualize the vertical structure of El Niño defined on the depth and longitude coordinates, i.e., the zonal (xz) cross-section. This helps diagnose the 3D structure of the El Niño dynamics, such as the vertical thermocline variation [30]. Thus, XSLICE extends the conventional analyses based on the depth layers to six new dimensions: the cross-sections of xz , yz , xy , xt , yt , and zt . Once a cross-section is created, we can then crop the domain or vary the once-fixed variables. In the zonal cross-section example, we can vary y to demonstrate latitude variations of El

Niño's structure, vary t to explore the temporal variation of El Niño dynamics, or vary x and z to crop the domain to focus on El Niño variability for a specific region. Dramatic changes that appear during this process can detect the boundaries of climate patterns, identifying key areas to explore or incorporate into larger 2D or 3D visualizations.

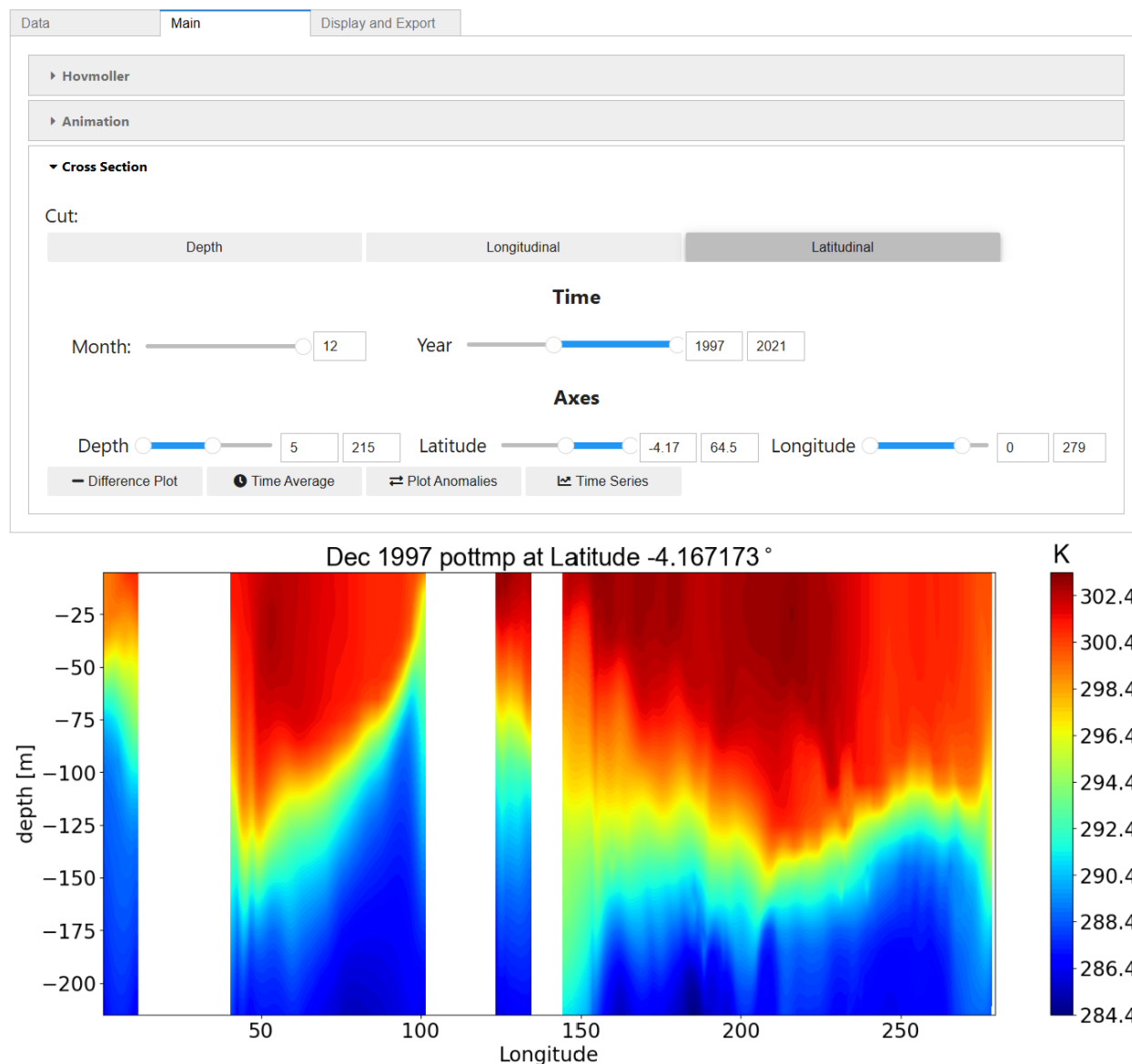


Figure 5.5: Zonal cross-section for the GODAS potential temperature at $\sim 4^\circ\text{S}$. Time was chosen as December 1997 to show temperatures for a strong El Niño.

The XSLICE interface (Fig. 5.6) was designed to be user-friendly for both classroom instruction and scientific research. It offers multiple services: (i) visualizing and cropping six

types of cross sections, (ii) highly customizable visualization with custom text fonts, text size, color maps, dark mode, and more (Fig. 5.7), and (iii) various image saving options including PNG, JPG, and PDF. In this way, XSLICE helps overcome the relative obstacle for people who are not proficient with computer coding but wish to visualize NetCDF datasets.

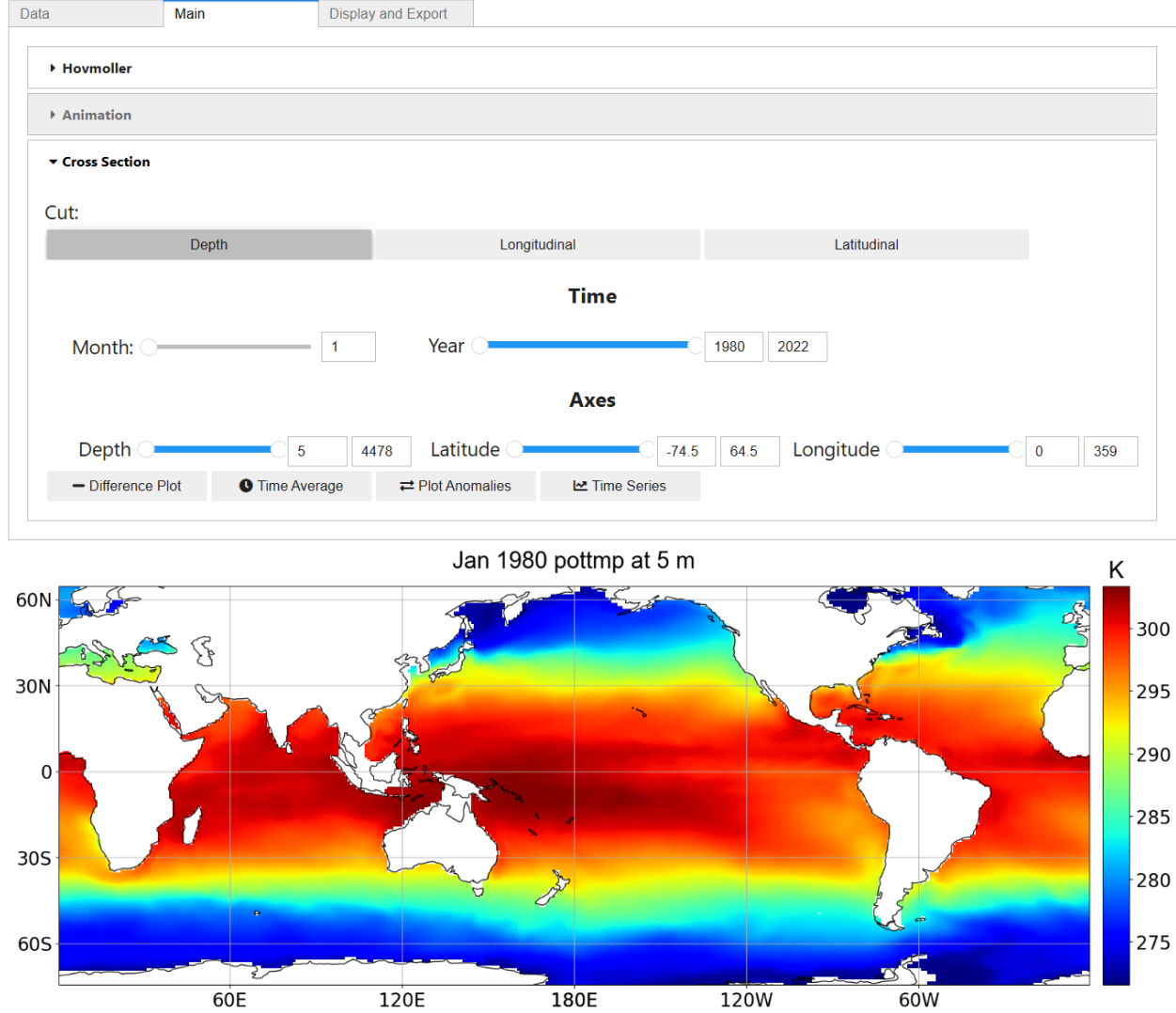


Figure 5.6: Main tab for XSLICE and the cross-section menu. Each cut tab is named after the variable held fixed for the cross-section. **Depth** is a longitude-latitude (xy) cut, **Longitude** is a meridional (yz) cut, and **Latitude** is a zonal (xz) cut. A complete tutorial on each button can be found in [1].

Unfortunately, the current version of XSLICE only uses the Global Ocean Data Assimilation System (GODAS) data [112], and the application of XSLICE on other datasets is nontrivial due to its many features. For this dissertation, we created a simplified version of XSLICE

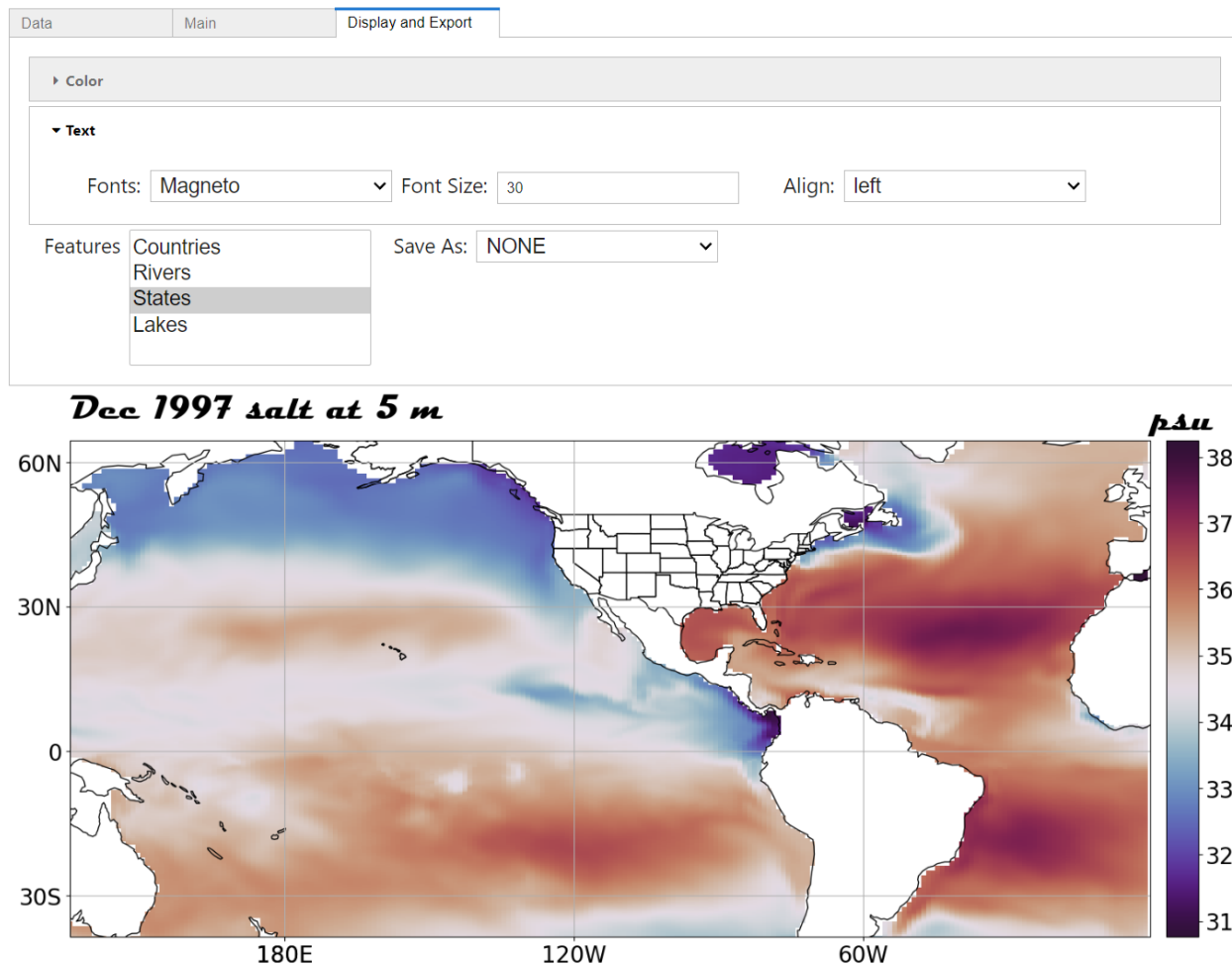


Figure 5.7: Additional color and display features in XSLICE that can change the figure’s aesthetic for export.

with fewer interactive widgets to streamline its application on various datasets. Specifically, our version reduces the cross-section tool (Fig. 5.6) to its primary functions of visualizing, cropping, and scanning cross-sections of 3D data.

The following summarizes the modifications made on the main XSLICE dashboard (Fig. 5.6). For cropping and scanning, we incorporated similar **Axes** sliding buttons as XSLICE. To accentuate important features in the data, a **Clip** text input was added that sets the interval limits of the colorbar. The original colorbar was also replaced with the custom colorbar made in Section 5.2 pictured in Fig. 5.4(b). Given this study is an analysis of various EOFs that are in depth-latitude-longitude format and do not contain time, the **Time** sliders, **Time Average**, **Plot Anomalies**, and **Time Series** buttons were removed. To compare various

results, instead of a **Difference Plot** button, we introduce a new interface called **Compare**, which will visualize the cross-section of multiple datasets simultaneously. To select the data for visualization, our **Cross-Section** interface uses a selection menu, and our **Compare** interface uses tags that the user can choose from.

We employ our simplified XSLICE to examine the first three 3D and 2D EOFs from Chapter 3. The EOFs are calculated using the Global Ocean Physics Reanalysis (GLORYS) [79] boreal winter ocean temperatures for the global domain. The data has a latitude-longitude resolution of $1/12^\circ$ by $1/12^\circ$ with 50 depth layers covering the entire global ocean (80°S - 90°N , 180°E - 180°W) from the ocean surface to $\sim 5,727$ meters (see Table 3.1 for details). We call this version developed for our analysis XSLICE for GLORYS. It contains a total of 4 interfaces: the **Depth**, **Latitude**, and **Longitude** interface, where each visualizes their respective cross-sections², and the **Compare** interface, which plots multiple EOFs at once. Each cross-section interface was developed individually and not as a part of a larger interface as in the Original XSLICE (Fig. 5.6). This approach ensures straightforward applications for other datasets while maintaining the option to merge all interfaces later.

To construct each interface, XSLICE for GLORYS uses multiple widgets from a Python package called *ipywidgets* [113]. The complete widget description and usage is as follows:

- Sliders: A single button on a slider is used to choose a single value. These are utilized to choose which depth, latitude, or longitude value to visualize in their respective cross-section.
- Range sliders: Two buttons on a slider that choose two values. The values represent which domain range the visualization is cropped to.
- Select menus: A menu with various options that the user can click on to select. For our EOF analysis, it is used to select which EOF will be visualized.

²In the original XSLICE this is equivalent to the xy , xz , and yz cross-sections.

- Text input: Appears as a text box that will take a number input. It defines the minimum and maximum values for the colorbar (i.e., $[a, b]$), which we call the clip value.
- Selection tags: Are tags with specific options that can be rearranged, added, or discarded. Used to select, discard, or arrange multiple EOFs for visualization.

XSLICE for GLORYS intentionally omits the customization and export features of the original version, as its purpose is not to create polished figures. Rather, it serves as a simple, data-adaptable visualization tool designed specifically for ocean dynamic detection and analysis. To demonstrate the successful fulfillment of this goal we: (i) apply XSLICE for GLORYS on multiple EOFs computed from GLORYS, (ii) identify regions of large climate variability that reflect key features or mechanisms of a climate phenomena, (iii) quantify and compare the 3D structure of various climate modes, and (iv) generate the initial research figures that are later modified for publication. In the following sections, we present the four interfaces of XSLICE for GLORYS and provide a comprehensive description of their features and successful implementation.

5.3.1 The Cross-Section Interfaces: Implementing XSLICE for Three Spatial Dimensions

5.3.1.1 Depth

The depth cross-section will scan through all 50 depth layers of the GLORYS data from surface to 5,727.92 meters. Figure 5.8 shows the initial output from the depth cross-section of EOF 1 at the surface for the entire latitude-longitude domain ($80^{\circ}\text{S} - 90^{\circ}\text{N}$, $180^{\circ}\text{E} - 180^{\circ}\text{W}$). Since we use the colorbar from Section 5.2, all land data is white, and EOF values are represented in warm or cool colors for positive or negative values, respectively. Mapping

functions, including drawing coastlines, grid lines, and producing labels, are implemented using the *Basemap* toolkit from matplotlib's library.

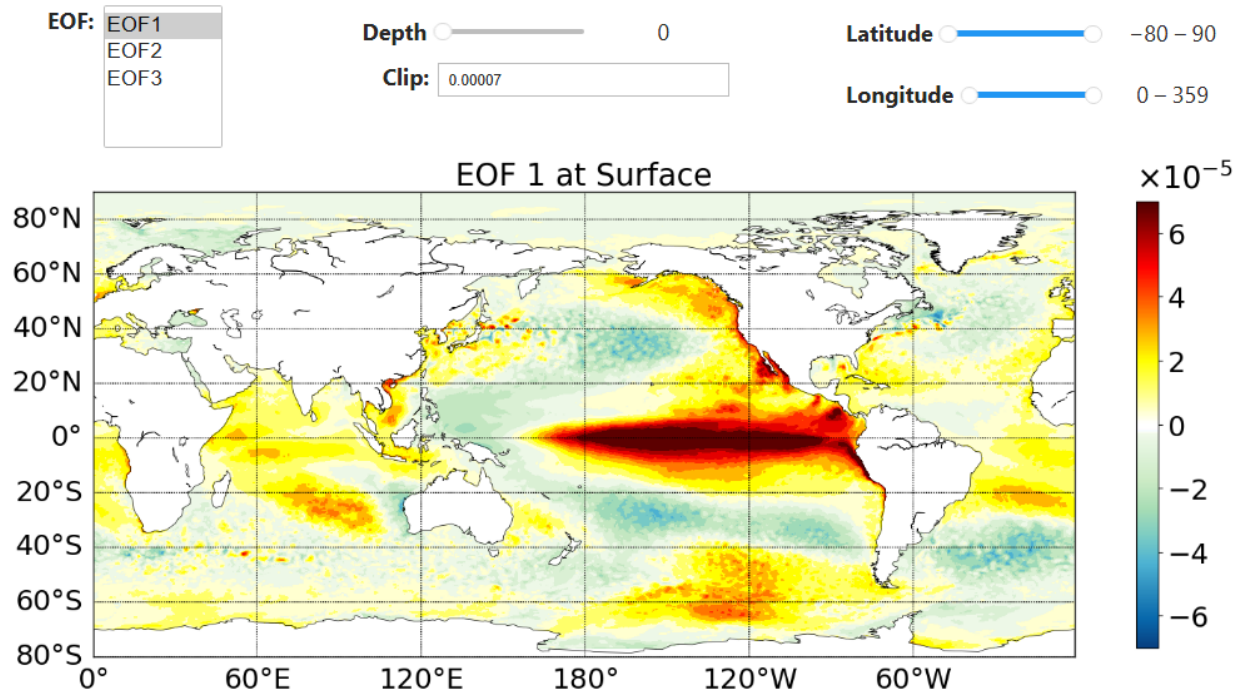


Figure 5.8: Initial interface for the depth cross-section of XSLICE for GLORYS. Above the visualization are the interactive widgets that will modify the figure. In its initial state, the interface shows the EOF 1 for the entire latitude-longitude domain at the surface. The signal captured in EOF 1 is the traditional El Niño pictured as the eastern Pacific warming.

The **Depth** interface consists of five interactive widgets: **EOF** selection menu, **Depth** slider, **Clip** text input, **Latitude** range slider, and **Longitude** range slider. The **EOF** selection allows users to choose which 3D EOF to visualize and remains consistent across all cross-section interfaces. The selection options include EOF 1, which contains the traditional El Niño (referred to as simply El Niño in this chapter); EOF 2, which contains warming coupled with marine heat waves and Central Pacific (CP) El Niño; and EOF 3, which represents the meridional modes. In this chapter, EOFs are visualized and described using the positive spatial pattern expression of their climate mode (e.g., El Niño and not La Niña). To exemplify the uses and benefits of the depth cross-section, we focus on EOF 1 and leverage the well-known zonal dipole structure of El Niño [7, 14].

The **Depth** slider scans through depth layers, with the corresponding readout indicating the current depth value. The title updates accordingly as users navigate through each layer. Depths are rounded to integer values for display purposes, but their actual values are outlined in Table 3.1. We define the first 0.49 meters as the surface (Fig. 5.8) to represent the layer spanning the surface to 0.49 meters. Subsequent depth layers are named after their ending layer. For example, the layer spanning 47 to 56 meters is labeled as 56 meters (Fig. 5.9).

Using the **Depth** slider, we scan through depths to describe the evolution of El Niño’s dipole throughout the water column. This is displayed as a warm eastern and cold western tropical Pacific. At the surface, EOF 1 mainly shows the eastern tropical Pacific warming (Fig. 5.8). At 56 meters (Fig. 5.9), a dramatic change occurs in the western tropical Pacific, where a strong cold signal appears. This marks the beginning of the cold half of El Niño’s dipole. Continued scanning through the depths of EOF 1 reveals additional dramatic changes to the dipole structure. Through this process, we identify depths that emphasize these dramatic changes, like the surface and 56 meters, and combine the resulting figures to produce multi-figure visualizations like Figs. 3.3 and 3.6. These figures are vital for communicating climate mode dynamics throughout the water column.

The **Clip** input text box manipulates the clip value, which defines both minimum and maximum bounds for the colorbar. The clip value corresponds to the parameter b in the interval $[a, b]$ described in Section 5.2. A zero-centered colorbar uses the symmetric interval $[-b, b]$, making the clip value equal to b . Every interface has a **Clip** input box with a clip value initialized at 0.00007. This value is initialized at 0.00007 because it produces reasonable figures for multiple EOFs across multiple depths (Fig. 3.6). The value can be changed by using the up and down arrows or through manual input.

Decreasing the clip value results in more values exceeding the colorbar range. Values outside the colorbar interval are visualized as the ‘over’ or ‘under’ color. In Fig. 5.9, any values less than -0.00007 appear dark blue, and any value above 0.00007 appears maroon. In the

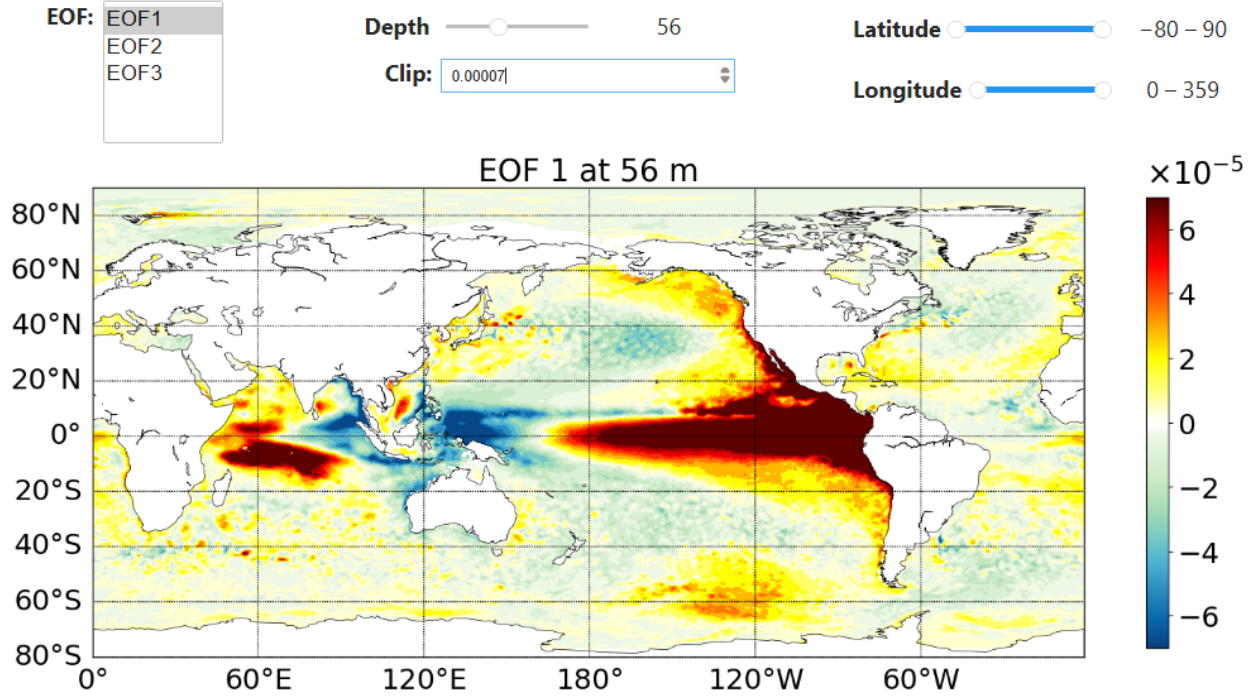


Figure 5.9: The depth interface after user interaction with the depth slider. This depth shows a distinct change compared to the surface in the western Pacific, due to the emergence of the subsurface zonal dipole component of El Niño.

eastern tropical Pacific of Fig. 5.9, many values surpass 0.00007 and are displayed as a single maroon color, accentuating the warming in that area. A similar process occurs in the western tropical Pacific, but for cooling. These cold and warm signals form the complete El Niño zonal dipole. The small clip value emphasizes signals throughout the figure to help identify the characteristics of a climate mode.

Increasing the clip value widens the colorbar range and reduces the number of values that exceed the colorbar limits. Values no longer at or exceeding the extremes lighten along with other intermediate values, resulting in a less saturated figure with fewer areas visualized as the ‘over’ or ‘under’ color. We observe an example of these changes in color through the cold western tropical Pacific signal of Figs. 5.9 and 5.10. The smaller clip value (Fig. 5.9) shows clear, strong cooling from the Maritime continent to 166°E. Increasing the clip value to 0.0001 lightens Fig. 5.9, producing Fig. 5.10. The resulting figure shows the western

tropical Pacific has a smaller area of strong cooling ending further west at 147°E. A larger clip value desaturates the figure and accents the climate pattern's details.

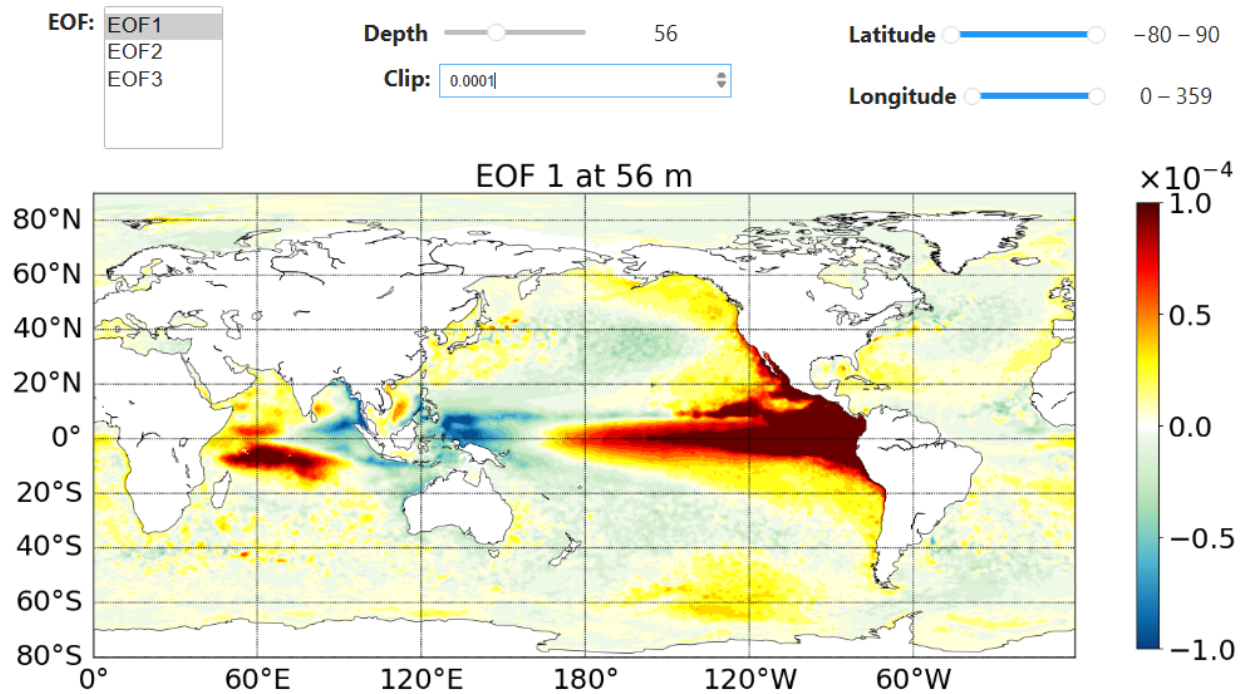


Figure 5.10: Same as Fig. 5.9 with a larger colorbar range, demonstrating how signals can be amplified or weakened depending on the clip values.

The **Latitude** and **Longitude** range sliders are dual-ended sliding controls that allow the user to select the latitude and longitude domain that is visualized. For latitude, southern values are negative and northern values are positive. Longitude values range from 0 - 360 and are converted to standard longitude values in the visualization. Although the EOFs have a $1/12^\circ$ by $1/12^\circ$ latitude-longitude resolution, the sliders are purposefully set to increment by 1° . This 1° increment facilitates quick comparisons to other results with coarse resolution, such as [23] and [33], and enables easier interpretation of basin-wide 3D structure.

Figures 5.11a and 5.11b demonstrate the cropping capabilities of the Latitude and Longitude range sliders applied to EOF 1. In each figure, the **Longitude** sliders are configured to increment by 1° (Fig. 5.11a) or $1/12^\circ$ (Fig. 5.11b), while all other widgets remain identical. Using the latitude slider, we select the same range as in [2] at $20^\circ\text{S} - 55^\circ\text{N}$, to capture

El Niño’s tropical warming along with its extra-tropical effects in the North Pacific. Both figures show that most of the eastern tropical Pacific warming occurs between 15°S - 15°N . For both increments, we set the easternmost longitude boundary at 70°W to capture El Niño reaching the South American coast. To identify the western boundary of El Niño’s warming, we adjust the left longitude slider to align with the westernmost point of the warming signal. For the increment of 1° the closest degree for the boundary was 162°E , while the increment of $1/12^{\circ}$ provides a more precise bound at 162.5°E . From this latitude-longitude cropping, we can identify the bounds of El Niño’s eastern tropical Pacific warming (15°S - 15°N , 162.5°E to the coast of South America). Although the $1/12^{\circ}$ increment offers more accuracy, locating 162.5°E in figures with large longitude ranges, such as Fig. 5.8, is more difficult than locating the whole degree 162°E . Additionally, integer degree bounds offer a more straightforward comparison to other studies, such as [2]. Therefore, larger increments are more practical for basin-wide analysis.

The latitude or longitude increment can be changed manually to any step size that matches the data. For GLORYS, this increment could be $1/12^{\circ}$, $1/6^{\circ}$, $1/4^{\circ}$, $1/3^{\circ}$, $1/2^{\circ}$, 1° , etc. As shown previously, the choice of 1° was made for practicality, especially for basin-wide dynamics, but smaller increments are useful for regional analysis. We use the well-studied impact of El Niño on the California Current System (CCS) as an example [15,82]. Using the **Latitude** and **Longitude** slider to focus on the California coast, we can locate one of the permanent meanders in the CCS, defined by [3,21] to be at approximately 40°N and 127°W . As described in Chapter 4, the effect of El Niño on this permanent meander appears in EOF 1 as a warm oval in the aforementioned area. For longitude range selection, the 1° increment (Fig. 5.12) shows the warm oval between 127° - 125°W , while the $1/12^{\circ}$ increment (Fig. 5.13) narrows the longitudes to a more exact range at 127° - 125.5°W . Unlike the previous example, locating 125.5°W on a map with a small longitude range is much easier. However, an integer degree is still preferable as it is more common and thus comparable to the longitude values used in studies that detail regional CCS dynamics during El Niño [3,21].

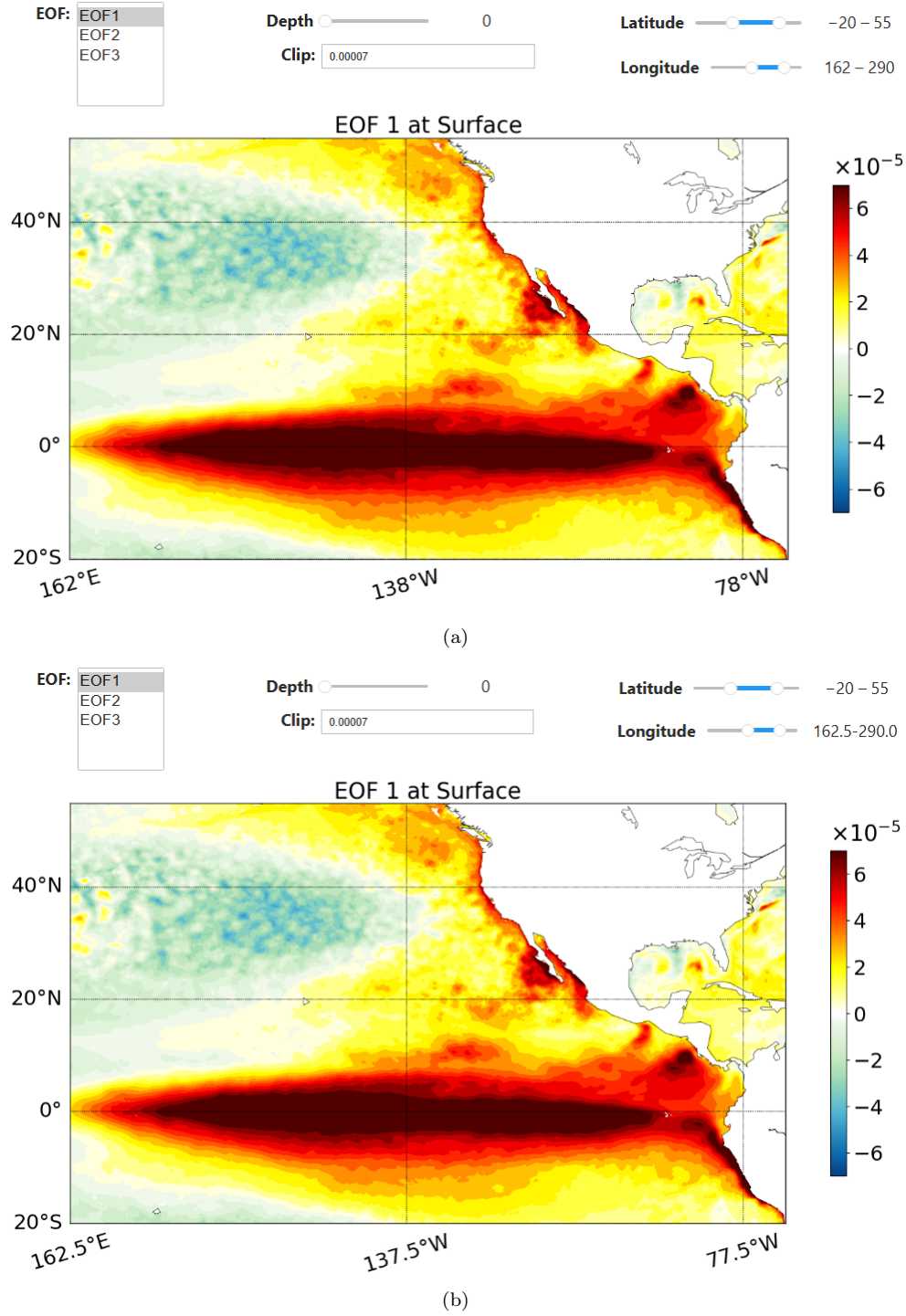


Figure 5.11: Surface view of EOF 1 cropped to capture El Niño's eastern Pacific warming. The latitude range of 20°S - 55°N and eastern longitude boundary at 70°W were selected to follow [2]. Western longitude was chosen to align with the westernmost extent of El Niño's eastern tropical Pacific warming. Figures differ in longitude increments, where (a) is the typical increment of 1°, and (b) uses an increment of 1/12° that matches the data's spatial resolution.

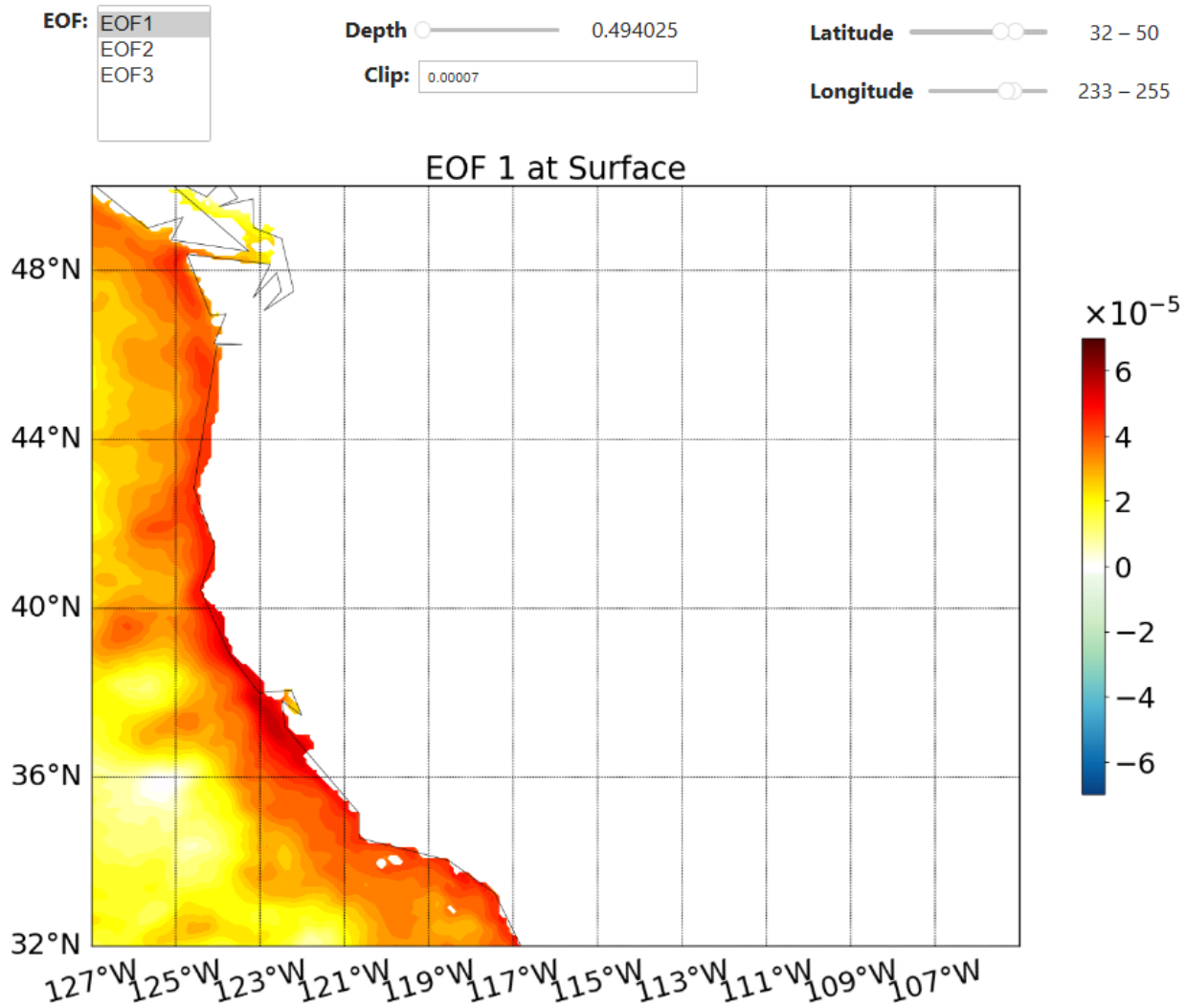


Figure 5.12: Surface view of EOF 1 cropped to show the U.S. West Coast for interpreting El Niño influence on the California Current System.

The depth scan effectively identifies areas of large variability by latitude and longitude at each depth level. It provides the most intuitive interface for general audiences because results are displayed on familiar map projections. Users can easily relate to locations and identify signals by cropping the map to their region of interest and scanning through ocean depths.

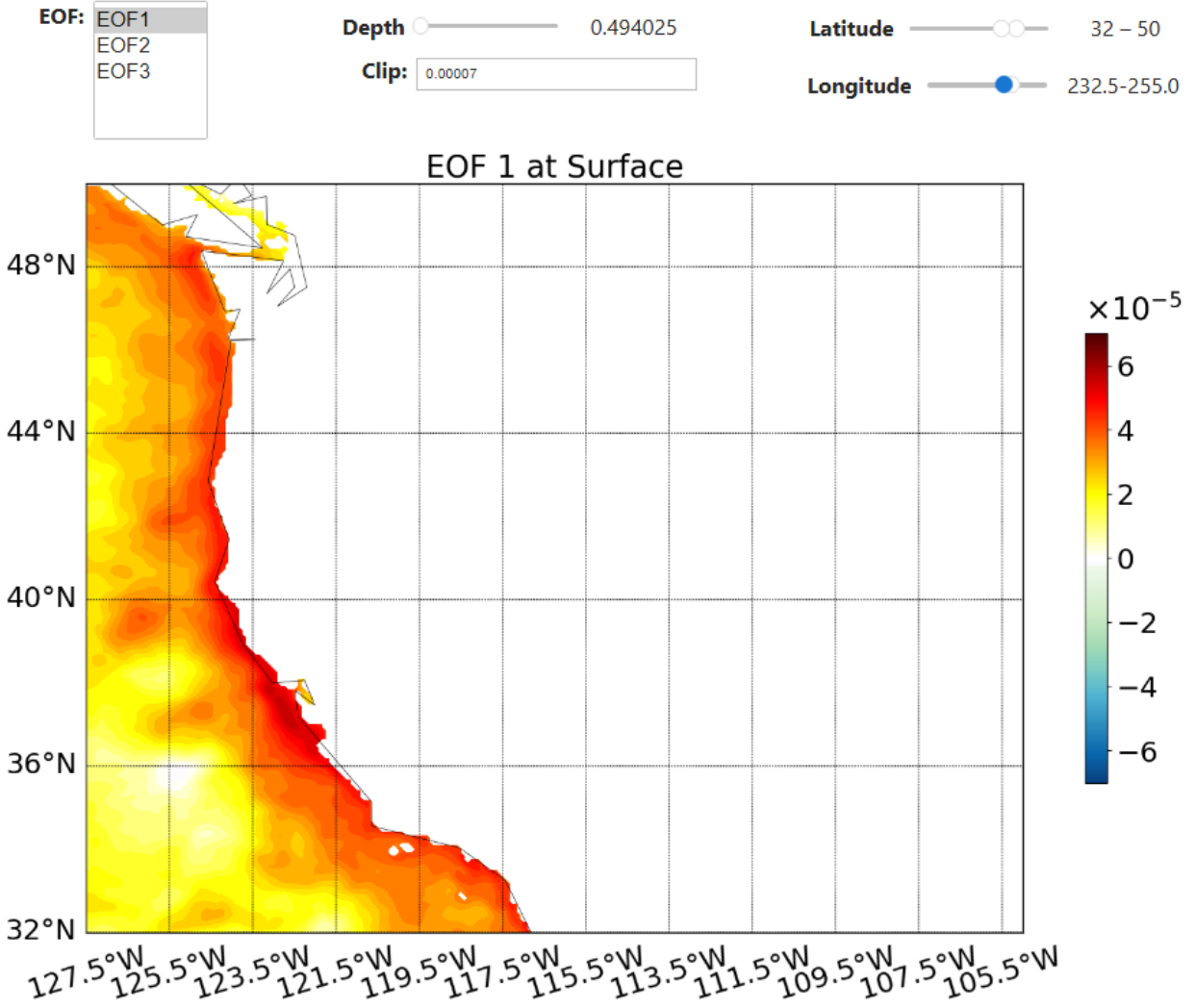


Figure 5.13: Same as Fig. 5.12 , but for a **Longitude** slider that increments by $1/12^\circ$. Step size is changed to identify precise coordinates for the permanent meander at 40°N and 127°W [3].

5.3.1.2 Latitude

The **Latitude** cross-section interface (Fig. 5.14) provides a zonal cut across all latitudes in 1° increments that can be cropped in depth and longitude. This interface features minor modifications from the previous depth cross-section. The **Latitude** control uses a single-button slider for scanning, while the **Depth** slider becomes a range slider with two buttons, similar to the **Longitude** range slider. The original XSLICE used range sliders for all dimension variables, which created consistency across the **Cut** interfaces (Fig. 5.5 and

5.6). However, for the cross-section/scanning variable, it was often unclear which button to interact with to scan through the cross-sections. This version updates the interface to follow a general design principle: variables used for cropping employ range sliders, while scanning variables use single-button sliders. As with the previous interface, range sliders initialize at their largest domain range. Since subsurface analysis with zonal cross-sections tends to favor the equator [6,7,14], the **Latitude** slider initializes at the equator for quick comparison with existing studies.

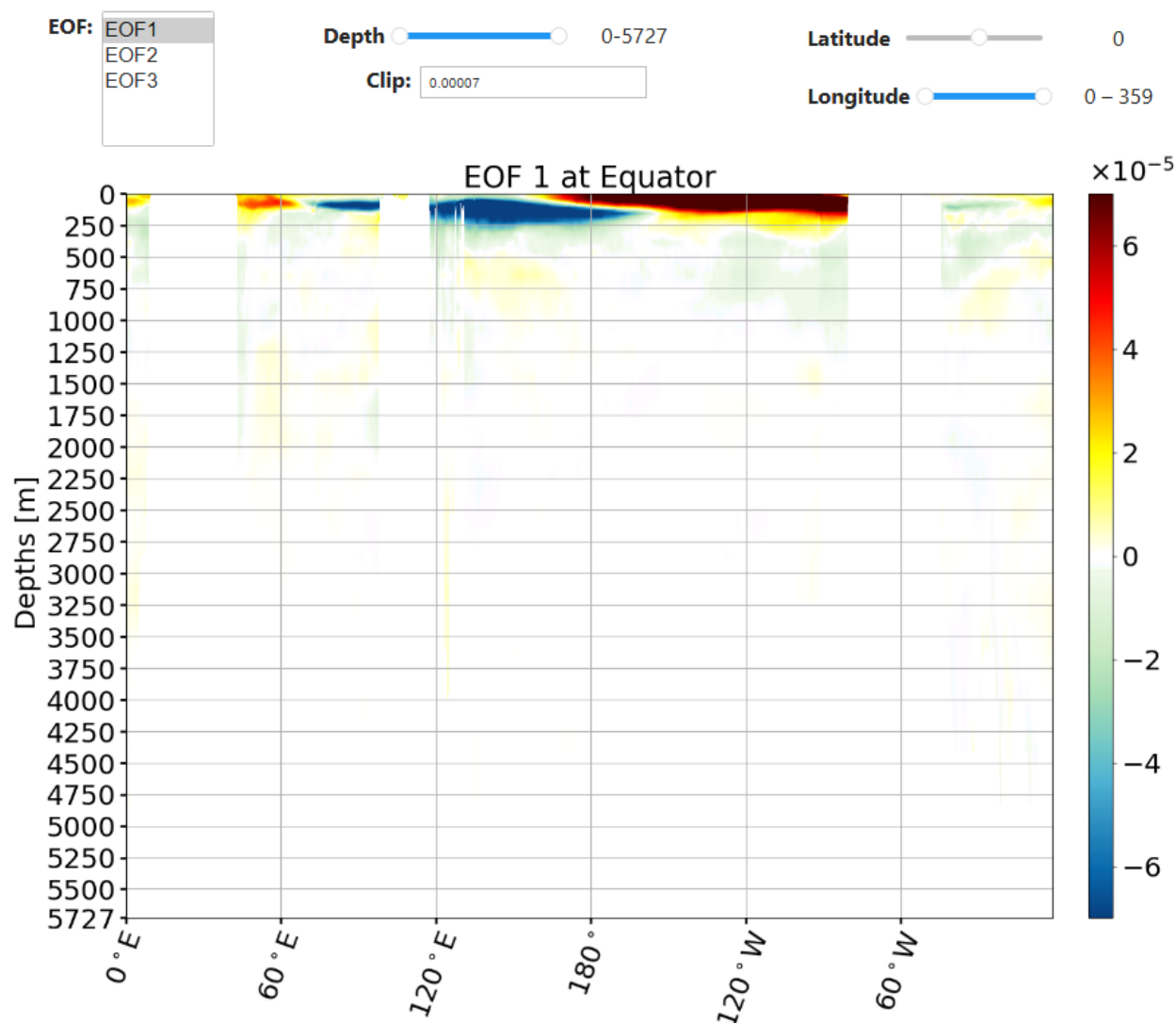


Figure 5.14: Initial interface for the latitude cross-section of XSLICE for GLORYS. In its initial state, the interface shows EOF 1 for the entire longitude-depth domain at the equator. The equatorial cross-section displays the subsurface zonal dipole signal of El Niño as warming in the east and cooling in the west.

The initial figure (Fig. 5.14) immediately reveals the vertical structure of El Niño and its global teleconnections without requiring depth layer scanning as in the **Depth** interface. The Indian Ocean displays the positive Indian Ocean dipole (pIOD), a phenomenon related to El Niño characterized by warm western and cool eastern tropical Indian Ocean waters. The Pacific contains El Niño’s subsurface dipole signature, but it is unclear from this initial figure where the exact boundaries of this feature lie.

Cropping the depth from surface to 380 meters and maintaining the small clip value of 0.00007 produces Fig. 5.15. Using this figure, we can identify the domain of El Niño equatorial dipole. The dipole’s cool half goes from 25 - 250 meters at 120°E - 170°W, and the warm half from surface to approximately 175 meters from 162°E to the South American coast. The depth cropping provides a better approximation of the dipole’s vertical extent, while a small clip value helps define the overall vertical structure.

Figure 5.16 demonstrates how a larger clip value is used to analyze the details of climate variability. The strongest warming occurs between 25 - 120 meters across the longitude band $\sim$ 140° - 80°W, while the strongest cooling appears at 50 - 160 meters between 120°E - 170°W. Identifying the depths that contain intense signals provides insight into a climate mode’s subsurface mechanisms. For El Niño, these cold and warm subsurface signals are concentrated at the thermocline, resulting from vertical thermocline displacement that amplifies variability in this region. This thermocline variation represents one of the key mechanisms associated with El Niño [114].

Selecting a different EOF exemplifies how vertical thermocline displacement varies among ocean phenomena. As explained in Chapter 3, EOF 3 contains the signature for the North Pacific meridional mode (NPM). Figure 5.17 shows that the strong equatorial subsurface warming is confined to the central Pacific. Unlike EOF 1, the thermocline variability is no longer basin-wide, suggesting that vertical thermocline displacement may be more likely in the central Pacific for the NPM. Further investigation of the NPM dynamics indicates

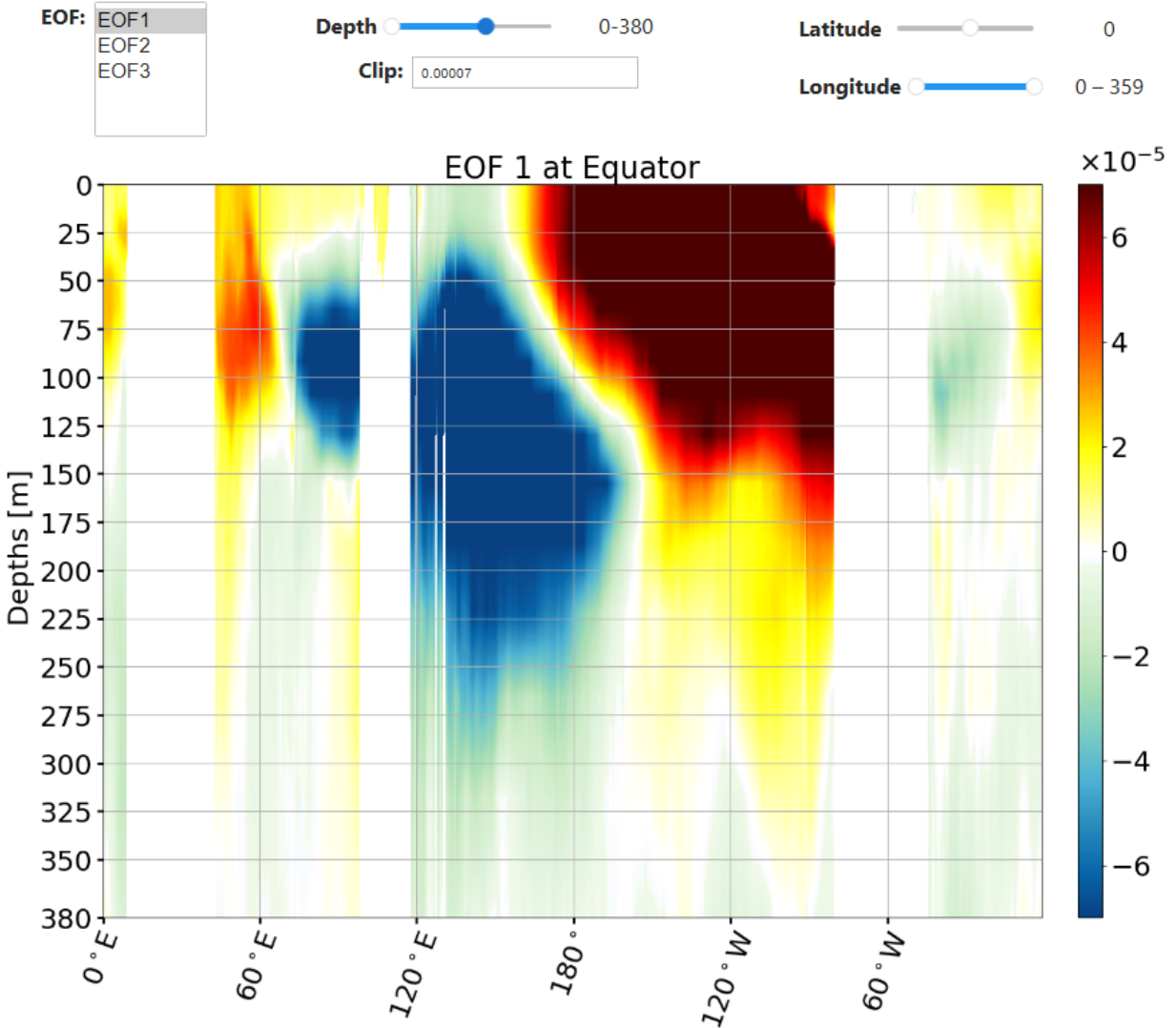


Figure 5.15: Same as Fig. 5.14 but limited to the first 380 meters to better analyze and quantify the El Niño dipole.

that central Pacific variability is not solely explained by thermocline variation but also by the trade wind charging mechanism [10,16]. Through the **Clip** and **EOF** selection tools, we are able to analyze the dynamics of each climate mode and explore their distinct mechanisms.

The **Latitude** slider creates visualizations that impart vital information about subsurface dynamics. Combining multiple zonal cuts produced from this interface illustrates how subsurface processes evolve in longitude and depth as they extend poleward. In particular, this research uses various latitude cuts along the U.S. West Coast (see Fig. 4.7) to examine each

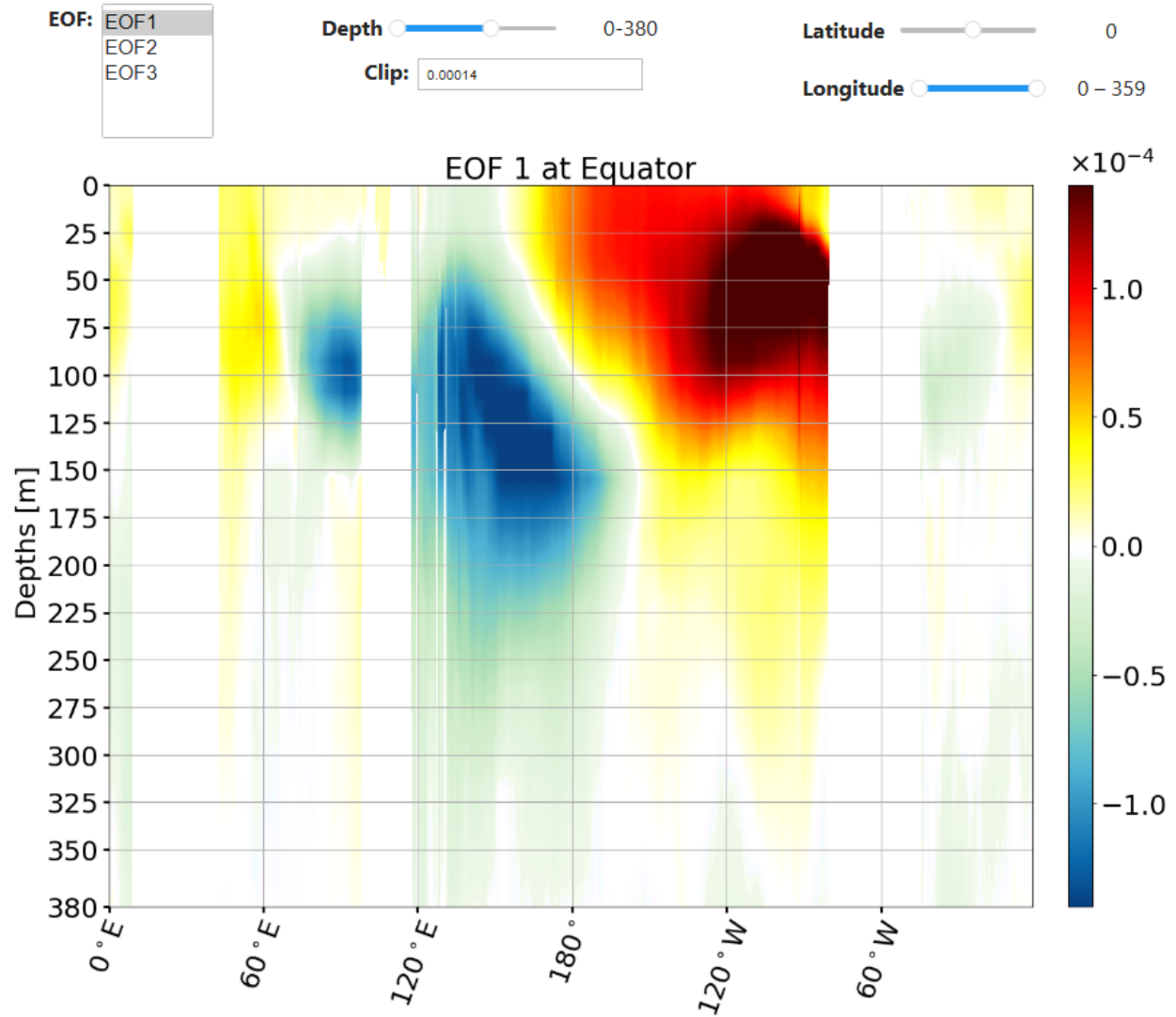


Figure 5.16: Same as Fig. 5.15 with a larger clip value to identify the areas with stronger variability in the El Niño dipole.

phenomenon's effects on the CCS. As an example, two cuts at 40°N and 37°N are shown in Fig. 5.18 using the **Latitude** cross-section tool. Both cuts exhibit distinct behaviors, with the northern cut (Fig. 5.18a) spanning a smaller longitude range than the southern cut (Fig. 5.18b). The progressive narrowing of the warming signal toward the coast from 37°N to 40°N reflects the behavior of El Niño's coastally trapped waves as they propagate poleward through the CCS [15].

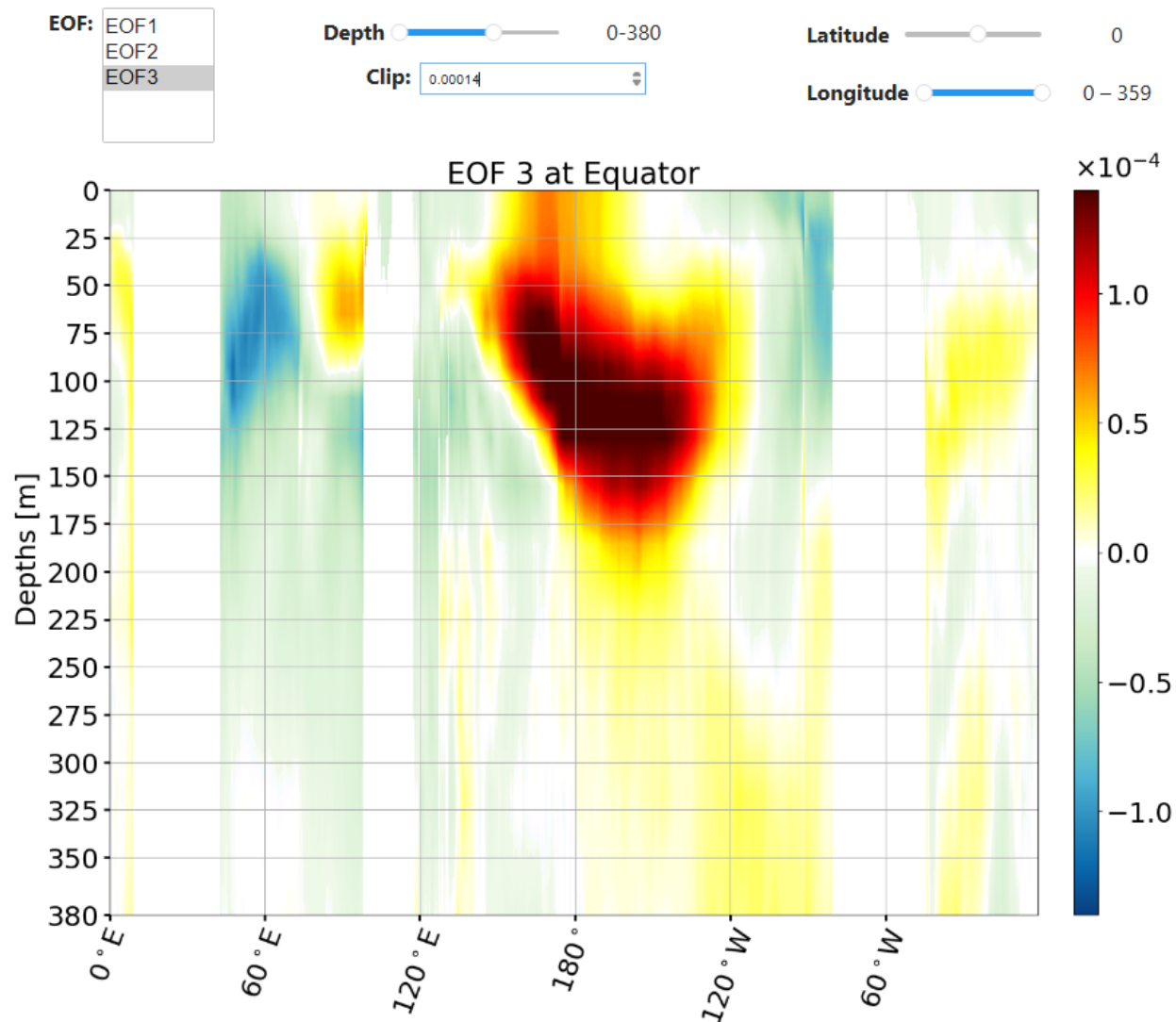


Figure 5.17: The equatorial cross-section of EOF 3 for the first 380 meters. Clip value is set to 0.00014 to display the maximum warming in the thermocline caused by the trade wind charging mechanism of the North Pacific meridional mode.

Figures generated from the **Latitude** interface are less intuitive than those from the **Depth** interface. Depth cuts incorporate a map for each depth level, making them easily understandable for general audiences. On their own, zonal cross-sections can be difficult to visualize on a 2D map. Three-dimensional visualizations address this limitation by placing each zonal cut on a 2D map, allowing audiences to understand each cut's location relative to coastlines. Figure 5.19 uses the cuts generated previously from the **Latitude** interface in Fig. 5.18 , along with additional cross-sections on the California coast. This single 3D figure displays El

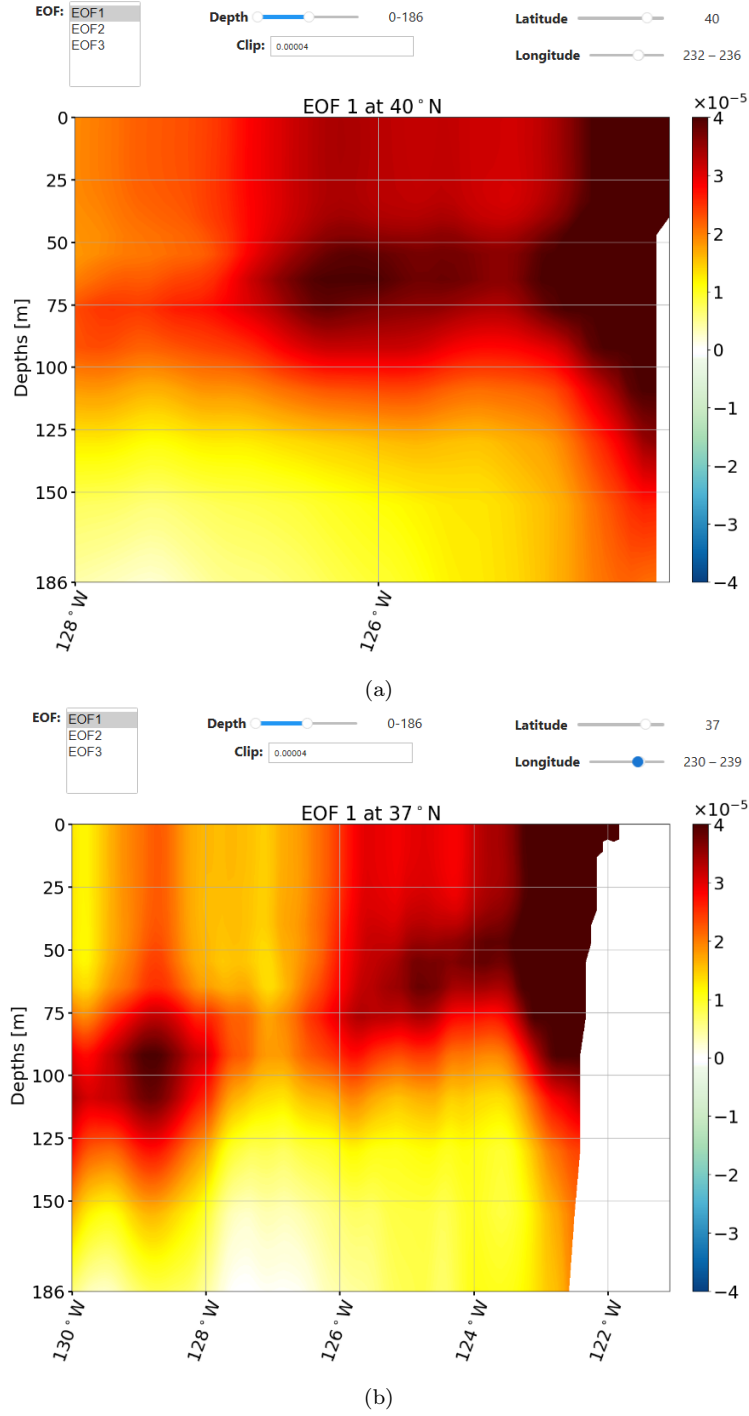


Figure 5.18: Two zonal cross-sections taken at (a) 40 °N, and (b) 37°N to examine the progressive narrowing of El Niño's warming toward the California coast.

Niño's effects on the CCS and latitude-dependent behaviors while maintaining interpretable geographic context.

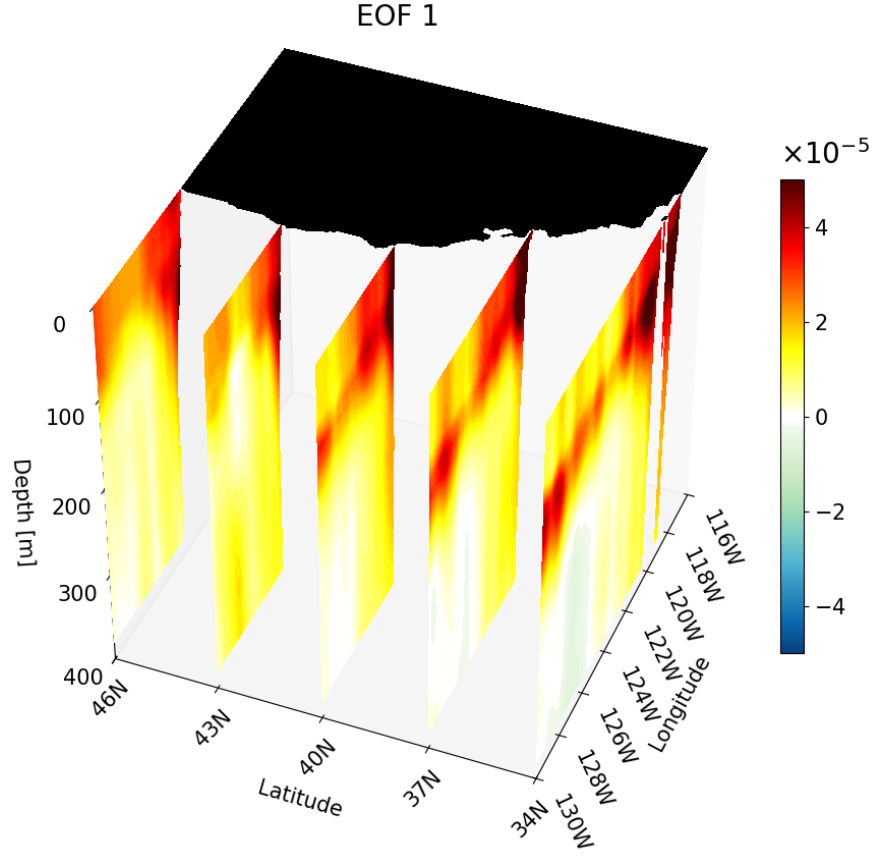


Figure 5.19: Multiple zonal cross-sections visualized along the U.S. West Coast from 34° - 46°N. Each cut is chosen to illustrate the changing warming pattern along the coast to interpret El Niño's effects on the CCS. These cross-sections include the cuts from Fig. 5.18 to demonstrate the application of the resulting cross-sections in a 3D figure.

Throughout this study, the **Latitude** interface served as the primary tool for investigating subsurface dynamics. This interface enabled scanning of zonal cuts to identify and analyze latitudes with strong signals, while the cropping feature allowed for signal enhancement by focusing on specific depths and longitudes within each zonal cut (Fig. 2.10). For each zonal cut produced by XSLICE for GLORYS, we quantified the depth and longitude ranges of several ocean dynamics; by aggregating multiple of these zonal cuts into a single 2D figure (Fig. 4.7), we tracked how these depth and longitude ranges vary with latitude. The zonal cuts from these multi-cross-section figures were further utilized to create geographically interpretable 3D visualizations, as demonstrated in Figs. 4.6(a) and (b). Beyond providing quantitative insights into subsurface structure, these multi-cross-section 2D and 3D figures

proved useful for recognizing vital subsurface mechanisms associated with various climate modes. While the **Latitude** interface produces less intuitive outputs compared to **Depth**, the resulting figures are crucial for understanding subsurface dynamics, establishing it as the most frequently used XSLICE for GLORYS cross-section interface in this research.

5.3.1.3 Longitude

The **Longitude** interface is useful for describing the latitude and depth range of a dynamic, particularly for processes like equatorial upwelling (Figs. 2.4 and 2.5) that span multiple latitudes. The interface is initialized with EOF 1 across the entire latitude-depth domain at the dateline. The meridional cut in Fig. 5.20 transects the El Niño dipole, showing warming and cooling centered around the equator. The dipole’s cooling lies below warming, indicating that the eastern Pacific cooling reaches deeper layers than the western Pacific warming. For this date line cross-section, the latitude extent of the dipole’s cooling is broader (10°S - 20°N) than the warming (5°S - 5°N). We take multiple meridional cuts to explore how the latitudinal range changes for various longitudes across the Pacific.

Each meridional cross-section is cropped to the first 554 meters from 10°S - 20°N to enhance the details of the dipole signature across the basin from 140°E - 100°W in 40° increments. To visualize all longitude cuts together and ease interpretation, the multiple meridional cuts are presented as a 3D figure (Fig. 5.21). This visualization shows that eastern Pacific warming experiences the largest latitudinal extent from 10°S - 12°N at its easternmost cut (100°W) and narrows to a smaller latitude range with western longitudes. The subsurface cooling maintains a relatively consistent latitudinal range until interrupted by continental boundaries. The warming signal shows varying depth reach, all extending from the surface to 100 meters at the date line, 225 meters at 140°W , and 125 meters at 100°W . The cooling signal reaches greater depths, spanning 50 - 325 meters at the dateline and 25 - 350 meters

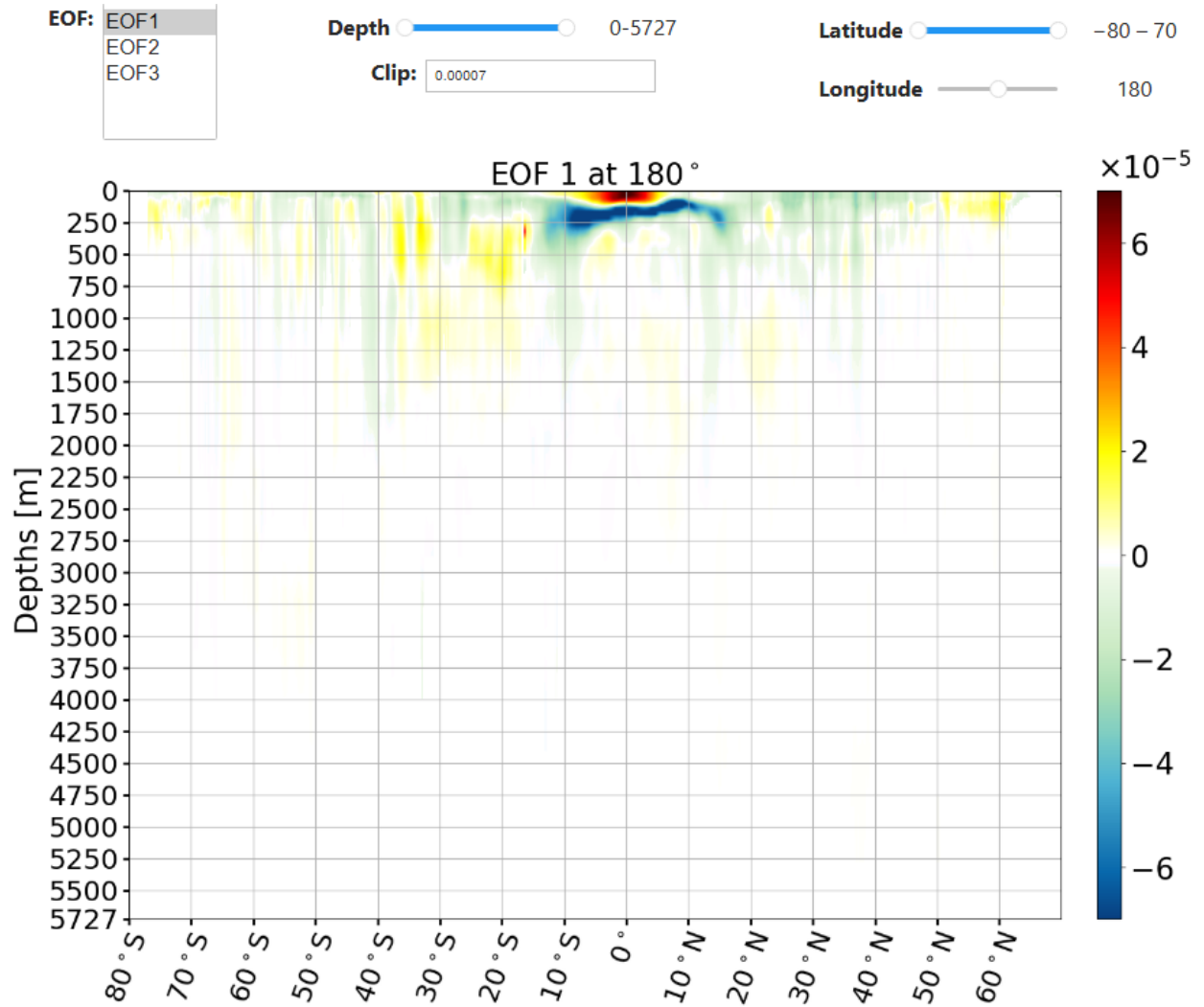


Figure 5.20: Initial interface for the longitude cross-section of XSLICE for GLORYS. In its initial state, the interface shows EOF 1 for the entire latitude-depth domain at the date line. The cross-section displays the subsurface dipole signal of El Niño around the equator with warming in shallower waters than the cooling.

at 140°W. This pattern of deeper western cooling compared to shallower eastern warming is consistent with the thermocline sloping upward from west to east.

Among the three cross-sections, the longitude dimension is perhaps the hardest to interpret. We typically read latitudes and depths vertically and longitudes horizontally, as in the depth and zonal cross-sections. In contrast, meridional cuts present latitudes horizontally and depth vertically (Fig. 5.20), resulting in a less intuitive visualization compared to the other cuts. This challenge is further amplified when presenting multiple meridional cuts in a

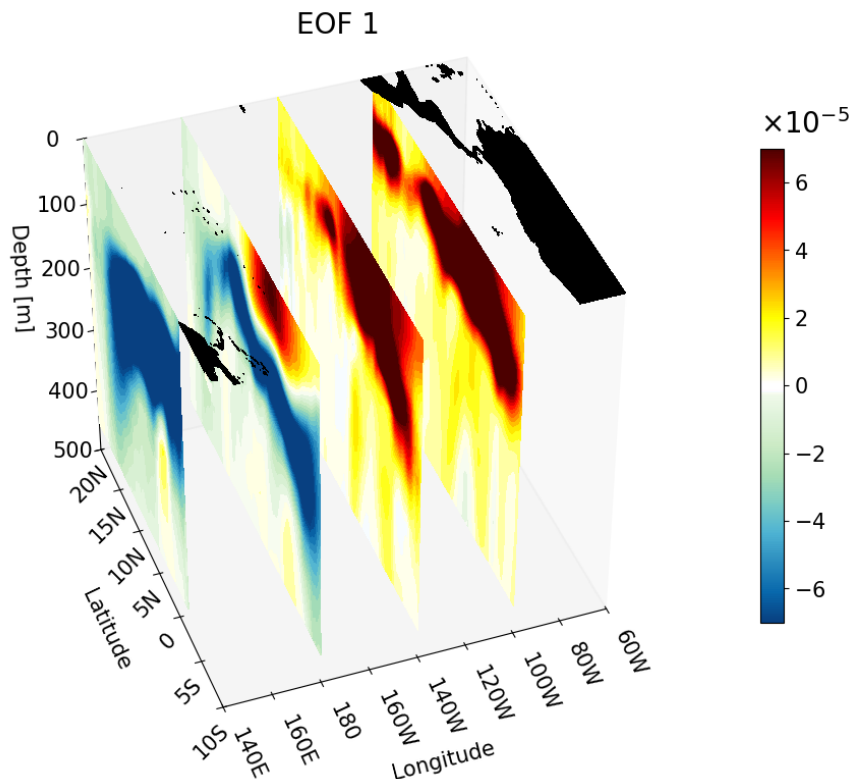


Figure 5.21: Multiple meridional cross-sections taken from 140°E - 100°W in 40° increments to display the El Niño dipole across the Pacific. Slices were chosen using the **Longitude** interface and are displayed in a 3D figure for easier interpretation.

2D figure. For better geographical understanding, we recommend creating 3D figures that display multiple meridional cuts simultaneously to supplement these 2D figures. As with the **Latitude** interface, the **Longitude** interface can be used to choose the appropriate longitudes, latitudes, and depths for multi-cross-section 2D and 3D figures such as Fig. 5.21.

5.3.2 Creation of the Interface for Comparison: Plotting Multiple Figures from 3D Datasets

The original XSLICE contained a feature called **Difference Plot** that compared two years of climate data by subtracting the two datasets. In XSLICE for GLORYS, we incorporate a **Compare** interface that plots multiple EOFs for a cross-section. We demonstrate this

interface using the depth cross-section because it is the most straightforward to interpret, and essential for preprocessing 2D EOFs. For the depth cross-section, the initial **Compare** interface (Fig. 5.22) uses the same starting depth, latitude, and longitude values as the **Depth** interface (Fig. 5.8). In our study, we compare either 3D EOFs to each other (as in Chapters 2 and 4) or 3D EOFs to 2D EOFs (as in Chapter 3). For the latter comparison, our interface enables visualization of the first three 2D and 3D EOFs. To demonstrate the comparison functionality for 3D EOFs, the interface is initialized with only the 3D EOFs displayed.

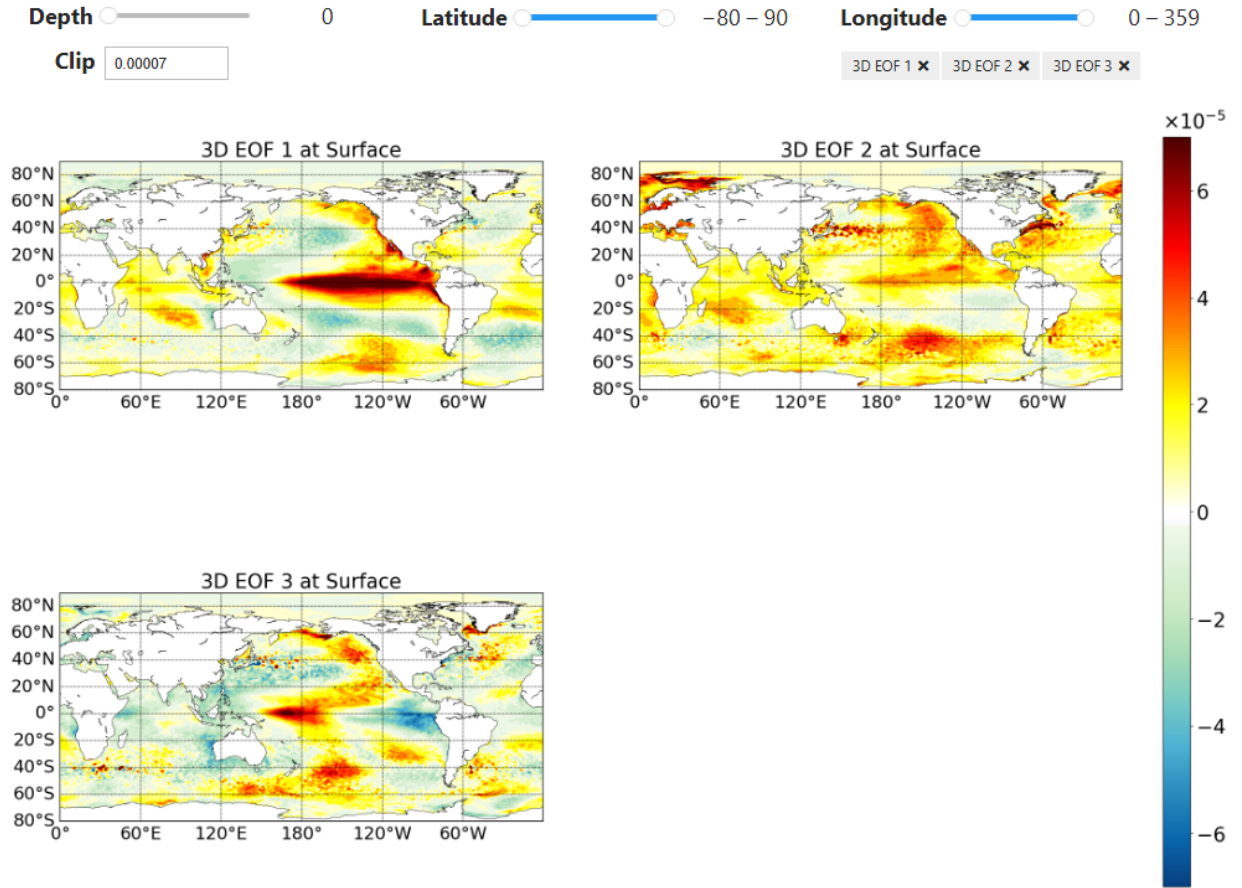


Figure 5.22: **Compare** interface for the depth cross-section. The initial dimension and clip values are the same as the **Depth** interface of Fig. 5.8. The **EOF** tags below the longitude range slider specify the EOFs that are visualized, and the order they are displayed. The top two figures are 3D EOF 1 and 2, and the bottom left figure is 3D EOF 3.

The interface allows users to select which EOFs to compare by adding or removing **EOF** input tags. These tags appear as gray boxes reading “3D EOF 1,” “3D EOF 2,” and “3D EOF 3” with an ‘x’ next to each (Fig. 5.22). Clicking the ‘x’ removes the **EOF** tag and its corresponding figure. Figure 5.23 shows the result of removing the 3D EOF 1 and 3 tags, leaving only 3D EOF 2. Clicking to the right of the last tag opens a text box where users can type the desired EOF for visualization. Alternatively, after opening the text box, users can click the downward arrow to access a drop-down menu of all EOFs to choose from. Both the text box and menu are shown in Fig. 5.23. After adding an **EOF** tag using either method, the interface will update with the additional figure, and the layout of the interface will update to accommodate the total number of figures.

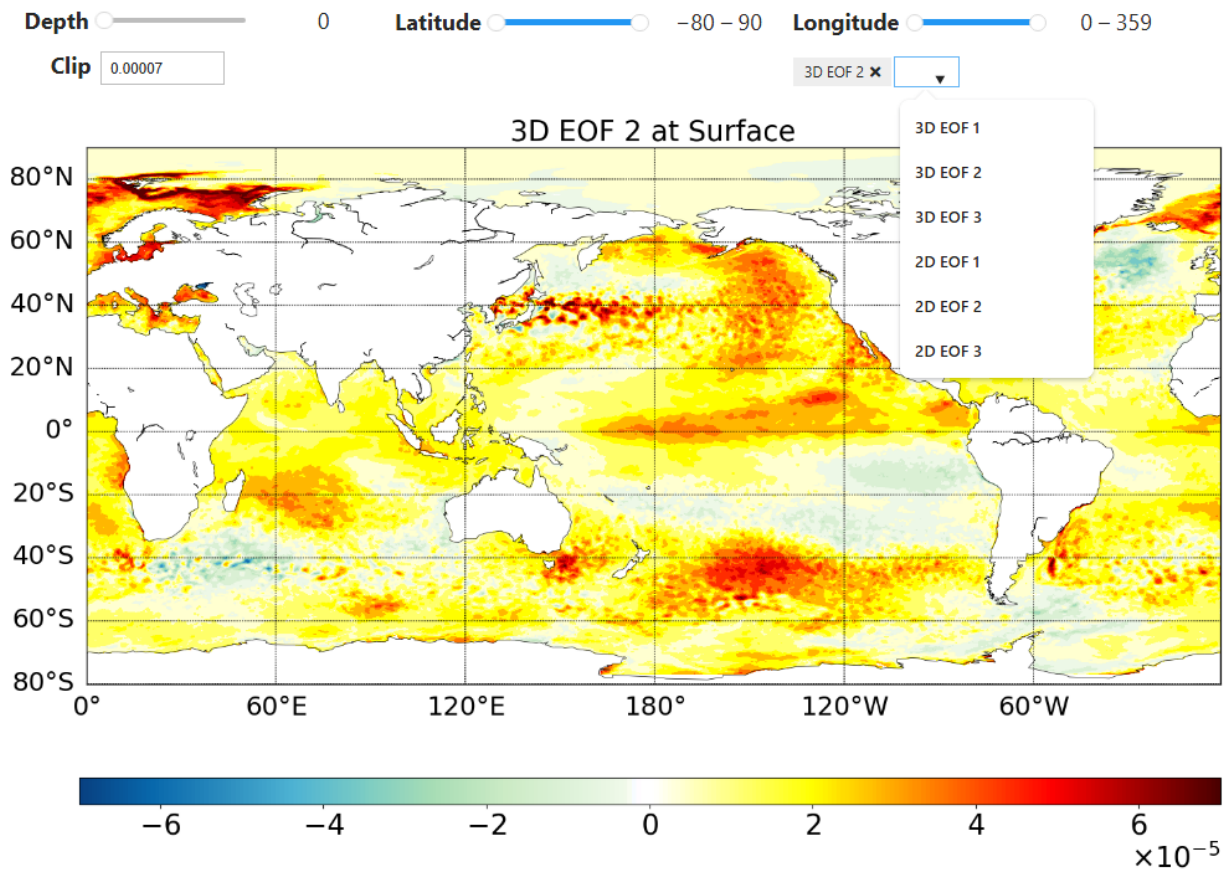


Figure 5.23: The **Compare** interface after removing 3D EOF 1 and 3D EOF 3 from the **EOF** tags, leaving only 3D EOF 2 to be displayed. A text box is displayed next to the 3D EOF 2 tag, ready for user input. A drop-down menu is displayed, providing the **EOF** tags that can be added for visualization.

The **Compare** interface dynamically adjusts its layout to maximize visualization space based on the number of tags being displayed. For the full set of 6 EOFs (three 2D EOFs and three 3D EOFs), the optimal configuration uses 2 columns and 3 rows as shown in Fig. 5.24. When displaying 3 - 4 figures, the bottom row of Fig. 5.24 is removed, displaying the EOFs in a 2×2 grid. Since depth cut figures require more horizontal than vertical space, we limit layouts to two columns for 3 - 6 figure visualizations. Adding more columns would significantly reduce individual figure sizes (Appendix Fig. C.2).

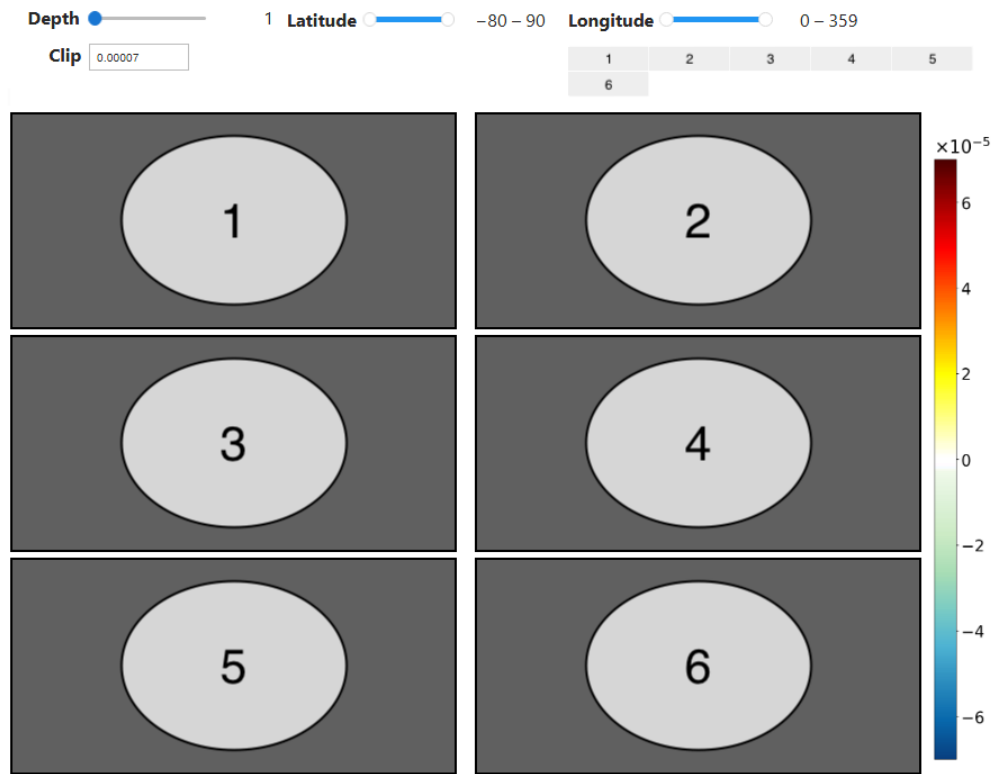


Figure 5.24: A 6-figure layout for the **Compare** interface. Figure placement follows tag order: 1, 2, 3, 4, 5, and 6. Figures are placed accordingly in three rows and two columns to maximize space. Layouts with 3 - 4 figures follow the same configuration but omit the bottom row.

For these 2-column arrangements, the colorbar is positioned on the right to prevent overlap during cropping interactions. The interface automatically resizes figures during cropping to maximize their display area within the layout constraints. For example, longitude cropping elongates each figure as demonstrated in Fig. 5.26. Bottom-positioned colorbars in 2-column

arrangements interfere with the lower row during such interactions, making right-side placement necessary for 3 - 6 figure layouts.

The layout in Fig. 5.24 works optimally for 3 - 6 figures and can accommodate additional visualizations by adding rows. However, this configuration is suboptimal for visualizations with 1 - 2 figures. In the 2-figure case, instead of side-by-side placement, figures are stacked vertically to better accommodate their greater need for horizontal space (Fig. 5.25). A single figure expands to fill the entire visualization area (Fig. 5.23). For 1-column layouts, the horizontal space is no longer divided between figures as in the 2-column layout, and cropping interactions can utilize the full width with minimal vertical resizing. This allows colorbars to be repositioned at the bottom of the interface without causing overlap issues in 1 - 2 figure configurations.

The **EOF** tag order directly controls figure placement within the interface layout. Users can easily reorder figures by dragging **EOF** tags to new positions. The program reads tags from left to right and arranges the corresponding plots row by row. Figure 5.24 illustrates this layout system with six tags. Reading from left to right, the first and second tags generate figures in the top row, the third and fourth tags populate the second row, and the fifth and sixth tags fill the third row. As mentioned previously, this arrangement applies to visualizations of 3 - 6 figures. We demonstrate the figure reordering using the initial 3-figure state in Fig. 5.22. If a user swaps the 3D EOF 2 and 3D EOF 3 tag positions, the 3D EOF 3 figure will move to the top-right position while 3D EOF 2 shifts to the bottom-left (Appendix Fig. C.1). This reordering feature facilitates comparison by allowing custom visualizations based on user preference. During this study, this feature was used to place 2D EOF 2, and later 2D EOF 3, beside 3D EOF 3 to assess their matching climate dynamics through multiple depth layers (Fig. 5.28).

To compare the first three 3D EOFs, we cropped each figure using the **Longitude** slider to focus on the Pacific (Fig. 5.26). With the resulting visualization, we identified the climate

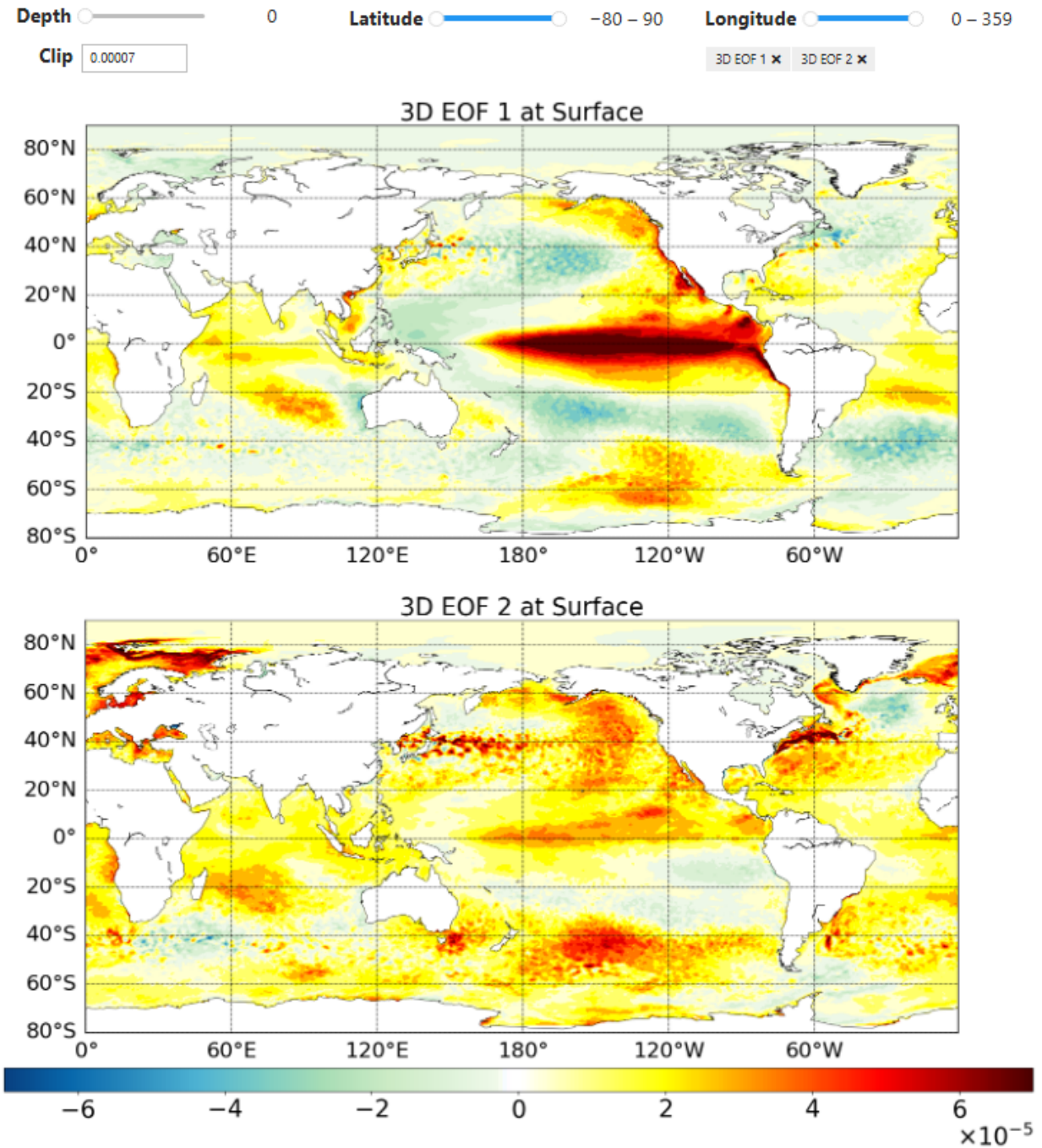


Figure 5.25: Two-figure layout displaying (top) 3D EOF 1 and (bottom) 3D EOF 2 at the surface. The figures are stacked vertically to maximize space for comparison. The colorbar is positioned at the bottom for the same reason.

modes in the Pacific as: El Niño in 3D EOF 1, warming coupled with marine heat waves and Central Pacific (CP) El Niño in 3D EOF 2, and the NPM in 3D EOF 3. Due to the well-documented connection and pattern similarities between the NPM and CP El

Niño [10, 90–92], 3D EOFs 2 and 3 were placed next to each other in the top row of Fig. 5.26 for easier comparison. Scanning through depths reveals each phenomenon’s distinct subsurface patterns. While 3D EOF 1 displays El Niño’s dipole, 3D EOFs 2 and 3 both contain central tropical Pacific warming that extends into the eastern tropical Pacific in different forms. The subsurface slice at 77 meters (Fig. 5.26) shows that 3D EOF 2 favors the eastern tropical Pacific with maximum warming in the east at 130° - 80° W; 3D EOF 3 has a line of maximum warming along the equator between 145° E - 110° W that is disrupted by strong cooling on either side of the basin. Though there is a relationship between the NPM and CP El Niño, the comparison tool allows us to discern similarities and differences in structure between the two patterns. While the **Compare** interface serves to analyze and relate 3D EOFs, it was initially developed for 2D versus 3D EOF comparisons as described in Section 3.3. This analysis required a multi-step process: (i) preprocessing 2D EOFs to ensure consistent positive patterns throughout depth layers, (ii) identifying spatial patterns in 2D EOFs at each depth layer, and (iii) comparing 2D and 3D results. The preprocessing step is essential due to the nature of 2D EOF calculations. Since 2D computations are separated by layer, they produce discontinuous pattern signs and representations across depth layers. For instance, 2D EOF 2 may display a warming signal globally at the surface but a cooling signal at 1 meter (Fig. 5.27). To correct this inconsistency, the 1-meter layer is multiplied by -1, producing a warming signal consistent with the surface.

To simplify the preprocessing step that corrects 2D EOFs, we utilize the **Compare** interface for depth cross-sections. Depth cuts are preferred over zonal or meridional cross-sections for this preprocessing because the latter produce uninterpretable figures due to depth-dependent sign discrepancies. Appendix Fig. C.3 demonstrates the result of an equatorial cross-section for an unprocessed 2D EOF 1. From this figure, it is unclear which depths require correction or what oceanic dynamics are present.

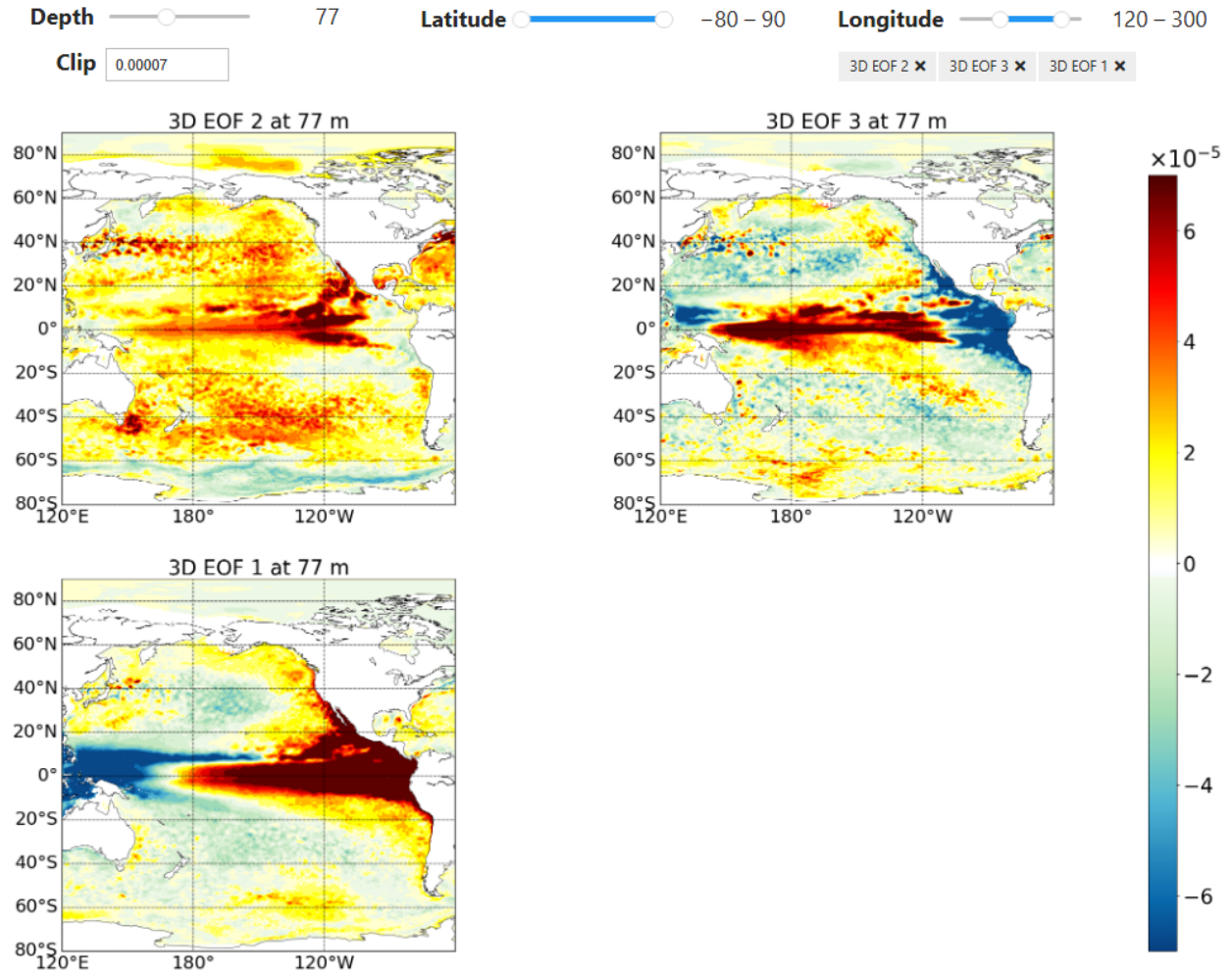


Figure 5.26: The first three 3D EOFs at 77 meters cropped to focus on the entire Pacific. The top row displays (left column) 3D EOF 2 and (right column) 3D EOF 3. Bottom row is 3D EOF 1. 3D EOF 2 and 3 are placed in the top row to explain the differences and similarities between the tropical Pacific warming from CP El Niño and the NPMM, respectively. 3D EOF 1 is placed in the bottom row for separate analysis of El Niño.

Determining which depth layers require sign correction involves scanning through each 2D EOF depth cut and comparing results to known dynamics and adjacent layers. Since 3D EOFs maintain consistency throughout the water column, when a 2D EOF resembles a 3D EOF at a given layer, the 3D results provide a reference for the expected sign. By combining the interface's comparison capabilities with depth scanning functionality, we can efficiently correct 2D EOFs and proceed to subsequent analysis steps.

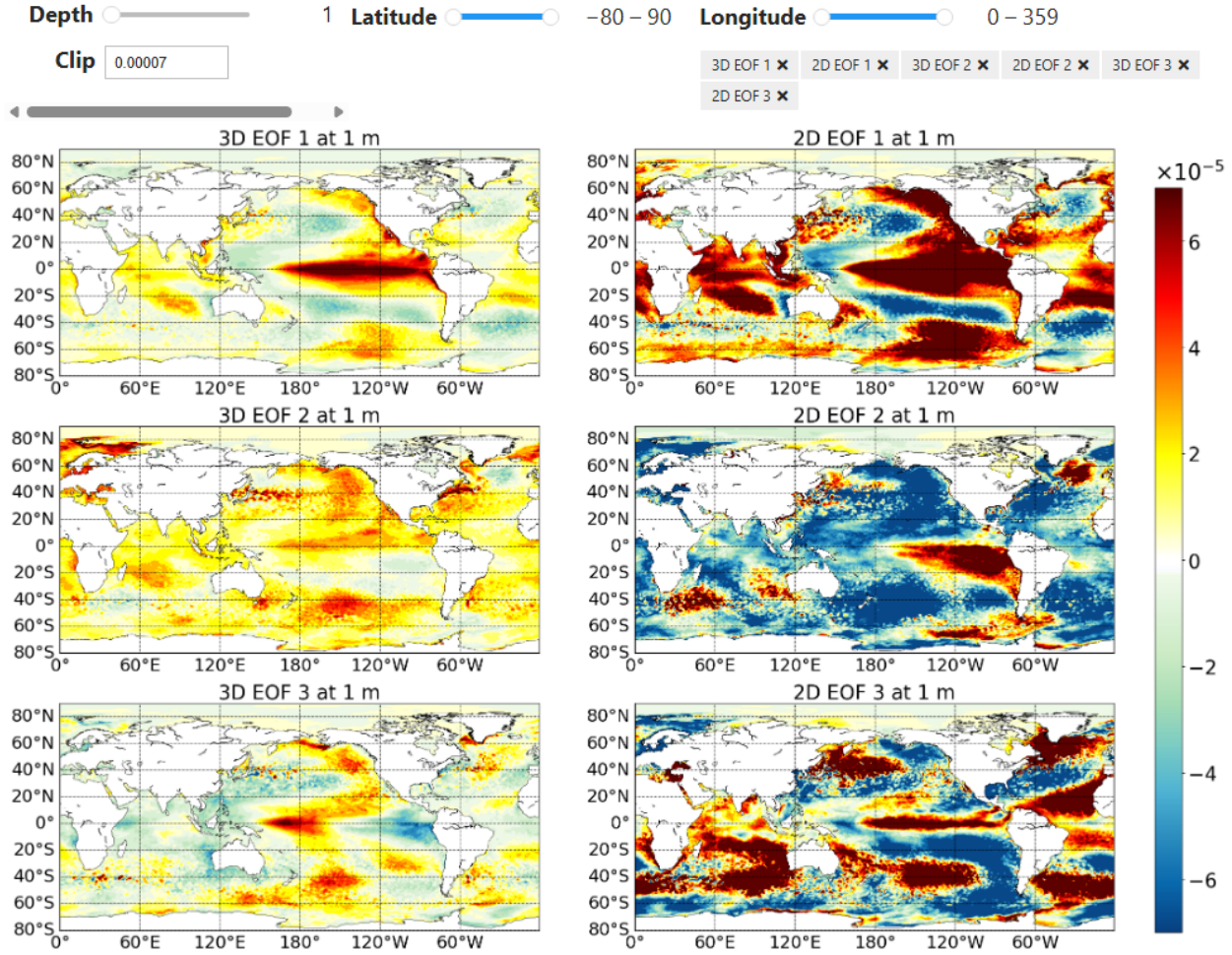


Figure 5.27: The **Compare** interface during the process of comparing 3D EOFs and 2D EOFs to correct for consistent sign representation in 2D EOFs. The first three (left column) 3D EOFs and (right column) 2D EOFs are shown for the global domain. The EOFs are illustrated at 1 meter, displaying a layer with incorrect signs in 2D EOF 2 and 2D EOF 3.

The **Depth Comparison** tool proved helpful in recognizing and deconstructing a climate mode's spatial pattern across the various EOFs. A visual comparison of 3D EOF 3, 2D EOF 2, and 2D EOF 3 suggested the NPMM was present in the three EOFs. Figure 5.28 displays the NPMM in 3D EOF 3 at 222 meters as two warm bands that straddle the equator: the northern band lies between 120°E - 180°, and the southern band is between 180° - 120°W. However, the individual 2D EOFs capture only portions of this pattern. 2D EOF 2 contains only the southern band, and 2D EOF 3 contains only the northern band. This depth layer comparison across multiple EOFs facilitated a straightforward analysis of how the NPMM

structure manifests within each 2D EOF, demonstrating the value of examining subsurface layers collectively.

The **Compare** interface presents a colorbar range dilemma when visualizing 2D and 3D EOFs simultaneously. Both EOFs have different value ranges: 2D EOFs are 10^{-4} while 3D EOFs are 10^{-5} . Since the interface produces only one colorbar for all EOFs, selecting an appropriate clip value becomes problematic. Choosing a small clip value may represent 3D EOFs well, but limits the details in 2D EOFs (Fig. 5.27). Conversely, using a large clip value may resolve this issue for 2D EOFs but produce no color saturation for 3D EOFs. Determining an appropriate clip value, therefore, requires a compromise in saturation for both EOF types.

For Fig. 5.28, the clip value was chosen to highlight the maximum warming in 3D EOF 3 without artificially amplifying the warming in the 2D EOFs to the point where the overall structure is lost. The NPM in 3D EOF 3 of Fig. 5.28 shows the trade wind charging warming in the tropical Pacific. Though not as apparent as in Fig. 3.6, this pattern remains recognizable. While the 2D EOFs exhibit strong warming signals that blend together to obscure some details, the NPM can still be decomposed using these figures.

The **Compare** interface was only demonstrated for the depth cross-section, but can be applied to other dimensions. The layout and tags remain consistent across all interfaces (not shown). Given the ability to remove all but one tag and produce figures similar to those in Section 5.3.1, one might question why a separate interface was created to plot individual EOFs. Why not use the **Compare** interface from the beginning? The implementation of **Compare** for each cross-section is not trivial. Each cut requires reshaping and creating new custom layouts when an EOF is added or discarded to ensure clear, interpretable figures. Consequently, adapting this interface for other datasets is complex. Using the **Compare** interface as the foundation for XSLICE for GLORYS would contradict our goal of developing

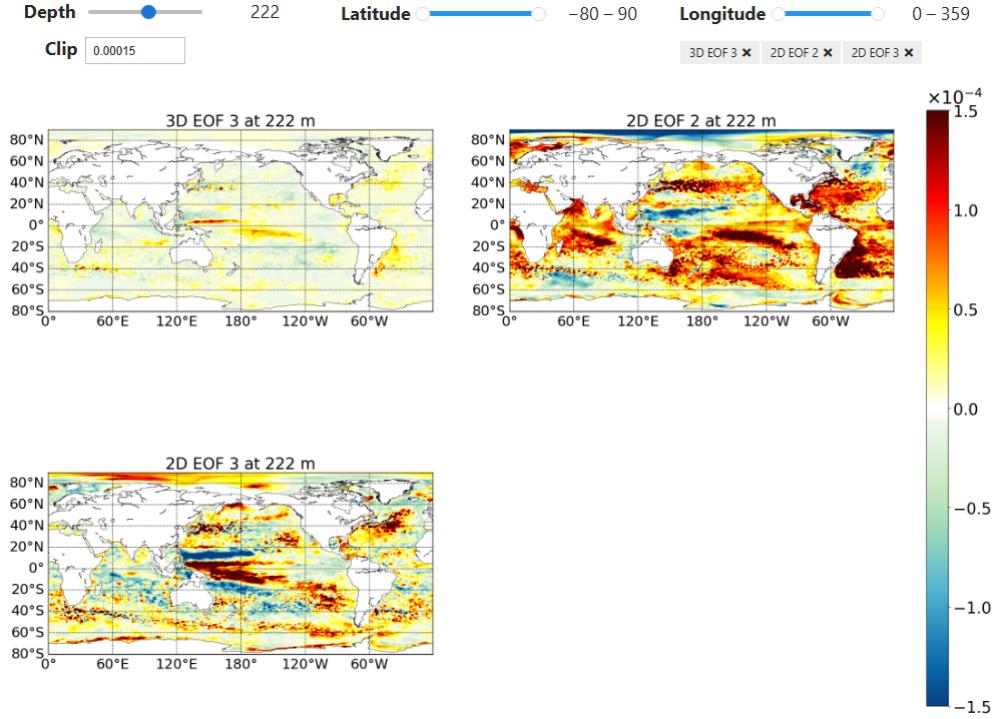


Figure 5.28: 3D EOF 3, 2D EOF 2 and 2D EOF 3 visualized at 222 meters for comparison. EOFs are displayed at this depth to identify the NPM in 3D EOF 3 and the split representation in 2D EOF 2 and 2D EOF 3.

a simple and flexible visualization tool. Providing the **Cross-Section** interfaces ensures the XSLICE for GLORYS remains easily adaptable for future applications with various datasets.

5.4 Summary and Discussion

This chapter describes the custom visualization tools developed to facilitate ocean data analysis. These tools include the custom colorbar and the four interfaces of XSLICE for GLORYS. Each section outlines the motivation, development process, functions, and applications of these tools. For clarity, the descriptions are accompanied by figures with example Python code and screen captures of specific tool interactions.

A tailored colorbar was created to effectively represent and enhance climate patterns in our EOF results. The initial iteration used contrasting warm and cold colors to highlight

respective temperature signals. While this approach made it easy to identify spatial patterns of equatorial upwelling and El Niño in the 3D EOF analysis (Chapter 2), the high contrast inadvertently visually amplified weak signals. To address this issue, the colorbar was updated to incorporate lighter cool tones and transitional colors, creating seamless gradients from cold to warm signals. The resulting colorbar maintained an appropriate representation of strong signals while accurately depicting weaker patterns. This updated colorbar was implemented with various data ranges and colormap lengths to generate figures in Chapters 3 and 4, and was employed in the XSLICE for GLORYS visualization tool.

To dynamically visualize various cross-sections from our results, we modified the XSLICE cross-section visualization tool, creating XSLICE for GLORYS. This version was designed to be easily adaptable to other datasets, allowing us to apply it to the various results in our study. To accomplish this, we simplified the original XSLICE to its main functions of visualizing, cropping, and scanning 3D data. Additionally, each cross-section interface was developed individually rather than as part of a larger interface, as in the original XSLICE. This ensured XSLICE for GLORYS maintained a simple but functional user interface with a straightforward, adaptable Python code.

XSLICE for GLORYS contains a total of 4 interfaces: the **Depth**, **Latitude**, and **Longitude** interfaces that visualize their respective cross-sections, and the **Compare** interface, which plots multiple EOFs at once for comparison. The latter is a new interface developed for this study and not a part of the original XSLICE. All interfaces include slider buttons similar to the original XSLICE that scan cross-sections and crop figures. Clip values were added as a feature to manipulate colorbar bounds and emphasize spatial patterns. We used the resulting figures to:

- Identify climate modes and quantify their 3D structure.
- Explore subsurface patterns and their associated mechanisms.

- Identify depths, latitudes, and longitudes with notable signals for later multi-figure 2D or 3D visualization.
- Compare and contrast spatial patterns in multiple 2D and 3D EOFs.
- Correct 2D results to display sign consistency throughout depth layers for later analysis.

XSLICE for GLORYS was built with the goal of visualizing to inform. For this reason, all resulting figures are sized to fill the interface display, and care was taken in creating an evolving layout for visualization with multiple figures. We prioritized figure visibility by maximizing space to detect smaller signals prior to cropping. Consequently, figures created directly from XSLICE for GLORYS may not be immediately publishable, but can be refined for publication.

This chapter focuses on the visualization tools used for the analysis and not the tools that produce polished, publishable figures. Therefore, it does not describe how to produce multi-figure 2D or 3D visualizations. Instead, we demonstrate the important preliminary steps that utilize these tools as building blocks for final figures: refining the colors for a colorbar, choosing appropriate colorbar limits, selecting specific cross-sections for 2D and 3D visualizations, and refining each cross-section through cropping. The 3D figures (Figs. 5.19 and 5.21) are included as examples of publication-ready figures created from this process.

The descriptions and examples in this chapter illustrate the value each tool provided for our analysis. The guide for creating a custom colorbar is useful for future customizations and for understanding which key aspects to manipulate to influence visual interpretation. The application of our custom colorbar demonstrated how it successfully illustrates ocean dynamics. The examples of XSLICE for GLORYS proved that a simple tool could serve as an exploratory research tool to gain insights into ocean dynamics and perform further quantitative analysis. The code is simple and adaptable without compromising functionality, making it an accessible, effective tool for other researchers. The user-friendly interface makes

it approachable as an informative tool for general audiences and could facilitate climate science education.

Bibliography

- [1] T. Bui, M. Ramirez, D. Lafarga, Y. T. Song, E. Foufoula-Georgiou, and S. S. P. Shen, “XSLICE: An efficient visualization and diagnosis tool for 4-dimensional climate dynamics,” *Journal of Atmospheric and Oceanic Technology*, vol. 42, no. 9, pp. 1185 – 1199, 2025. [Online]. Available: <https://journals.ametsoc.org/view/journals/atot/aop/JTECH-D-24-0113.1/JTECH-D-24-0113.1.xml>
- [2] B. C. Weare, A. R. Navato, and R. E. Newell, “Empirical orthogonal analysis of Pacific sea surface temperatures,” *Journal of Physical Oceanography*, vol. 6, no. 5, pp. 671 – 678, 1976. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/6/5/1520-0485_1976_006_0671_eoaops_2_0_co_2.xml
- [3] L. R. Centurioni, J. C. Ohlmann, and P. P. Niiler, “Permanent meanders in the California Current System,” *Journal of Physical Oceanography*, vol. 38, no. 8, pp. 1690 – 1710, 2008. [Online]. Available: <https://journals.ametsoc.org/view/journals/phoc/38/8/2008jpo3746.1.xml>
- [4] D. Amaya and National Center for Atmospheric Research Staff, “The climate data guide: Meridional modes and their indices,” 2025, <https://climatedataguide.ucar.edu/climate-data/meridional-modes-and-their-indices>. Accessed July 16, 2025.
- [5] M. A. Alexander, I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, “The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans,” *Journal of Climate*, vol. 15, no. 16, pp. 2205 – 2231, 2002. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/15/16/1520-0442_2002_015_2205_tabtio_2.0.co_2.xml
- [6] K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata, “El Niño Modoki and its possible teleconnection,” *Journal of Geophysical Research: Oceans*, vol. 112, no. C11, 2007. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JC003798>
- [7] H.-Y. Kao and J.-Y. Yu, “Contrasting Eastern-Pacific and Central-Pacific types of ENSO,” *Journal of Climate*, vol. 22, no. 3, pp. 615 – 632, 2009. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/22/3/2008jcli2309.1.xml>
- [8] J.-S. Kug, F.-F. Jin, and S.-I. An, “Two types of El Niño events: Cold tongue El Niño and warm pool El Niño,” *Journal of Climate*, vol. 22, no. 6, pp. 1499 –

- 1515, 2009. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/22/6/2008jcli2624.1.xml>
- [9] J. C. H. Chiang and D. J. Vimont, “Analogous Pacific and Atlantic meridional modes of tropical atmosphere–ocean variability,” *Journal of Climate*, vol. 17, no. 21, pp. 4143 – 4158, 2004. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/17/21/jcli4953.1.xml>
- [10] D. J. Amaya, “The Pacific meridional mode and ENSO: a review,” *Current Climate Change Reports*, vol. 5, no. 4, p. 296–307, 2019. [Online]. Available: <https://doi.org/10.1007/s40641-019-00142-x>
- [11] C. H. Gregory, C. Artana, S. Lama, D. León-FonFay, J. Sala, F. Xiao, T. Xu, A. Capotondi, C. Martinez-Villalobos, and N. J. Holbrook, “Global marine heatwaves under different flavors of ENSO,” *Geophysical Research Letters*, vol. 51, no. 20, p. e2024GL110399, 2024, e2024GL110399 2024GL110399. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL110399>
- [12] K. E. Trenberth and L. Smith, “The vertical structure of temperature in the tropics: Different flavors of El Niño,” *Journal of Climate*, vol. 19, no. 19, pp. 4956 – 4973, 2006. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/19/19/jcli3891.1.xml>
- [13] J.-Y. Yu, H.-Y. Kao, T. Lee, and S. T. Kim, “Subsurface ocean temperature indices for Central-Pacific and Eastern-Pacific types of El Niño and La Niña events,” *Theoretical and Applied Climatology*, vol. 103, no. 3, pp. 337–344, 2011. [Online]. Available: <https://doi.org/10.1007/s00704-010-0307-6>
- [14] K. Xu, C. Zhu, and J. He, “Linkage between the dominant modes in Pacific subsurface ocean temperature and the two type ENSO events,” *Chinese Science Bulletin*, vol. 57, no. 26, pp. 3491–3496, 2012. [Online]. Available: <http://link.springer.com/10.1007/s11434-012-5173-4>
- [15] M. Frischknecht, M. Münnich, and N. Gruber, “Remote versus local influence of ENSO on the California Current System,” *Journal of Geophysical Research: Oceans*, vol. 120, no. 2, pp. 1353–1374, 2015. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JC010531>
- [16] B. T. Anderson and R. C. Perez, “ENSO and non-ENSO induced charging and discharging of the equatorial Pacific,” *Climate Dynamics*, vol. 45, no. 9, pp. 2309–2327, 2015. [Online]. Available: <https://doi.org/10.1007/s00382-015-2472-x>
- [17] T. Doi, T. Tozuka, and T. Yamagata, “The Atlantic meridional mode and its coupled variability with the Guinea Dome,” *Journal of Climate*, vol. 23, no. 2, pp. 455 – 475, 2010. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/23/2/2009jcli3198.1.xml>

- [18] J. Wang, S. Zhang, Y. Li, and J. Sprintall, “A positive Indian Ocean Dipole leads to an Indian Ocean basin mode that favors the development of La Niña the following year,” *Geophysical Research Letters*, vol. 52, no. 7, p. e2024GL112632, 2025. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL112632>
- [19] W. Cai, K. Yang, L. Wu, G. Huang, A. Santoso, B. Ng, G. Wang, and T. Yamagata, “Opposite response of strong and moderate positive Indian Ocean Dipole to global warming,” *Nature Climate Change*, vol. 11, no. 1, pp. 27–32, 2021. [Online]. Available: <https://doi.org/10.1038/s41558-020-00943-1>
- [20] M. Mu, R. Feng, and W. Duan, “Relationship between optimal precursors for Indian Ocean Dipole events and optimally growing initial errors in its prediction,” *Journal of Geophysical Research: Oceans*, vol. 122, no. 2, pp. 1141–1153, 2017. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JC012527>
- [21] W. J. Crawford, A. M. Moore, J. Fiechter, and C. A. Edwards, “A principal oscillation pattern analysis of the circulation variability in the California Current System associated with the El Niño Southern Oscillation,” *Journal of Geophysical Research: Oceans*, vol. 124, no. 11, pp. 8241–8256, 2019. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JC015346>
- [22] W. J. Crawford, A. M. Moore, M. G. Jacox, J. Fiechter, E. Neveu, and C. A. Edwards, “A resonant response of the California current circulation to forcing by low frequency climate variability,” *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 151, pp. 16–36, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0967064517302898>
- [23] D. Lafarga, T. Bui, Y. T. Song, T. M. Smith, and S. S. P. Shen, “A feasibility study of three-dimensional empirical orthogonal functions from the NASA JPL ocean general circulation model: Computing, visualization and interpretation,” *Tellus A: Dynamic Meteorology and Oceanography*, vol. 75, no. 1, pp. 213–230, 2023. [Online]. Available: <https://doi.org/10.16993/tellusa.3223>
- [24] T. Bui, D. Lafarga, T. Smith, Y. Song, and S. S. P. Shen, “Calculation, visualization, and interpretation of three-dimensional atmosphere-ocean coupled empirical orthogonal functions using the reanalyses data,” *Theoretical and Applied Climatology*, vol. 154, pp. 1–15, 2023. [Online]. Available: <https://doi.org/10.1007/s00704-023-04513-1>
- [25] E. N. Lorenz, “Empirical orthogonal functions and statistical weather prediction,” MIT Department of Meteorology, Statistical Forecasting Project Rep. 1, 1956.
- [26] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, “Sampling errors in the estimation of empirical orthogonal functions,” *Monthly Weather Review*, vol. 110, no. 7, pp. 699 – 706, 1982. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/110/7/1520-0493_1982_110_0699_seiteo_2_0_co_2.xml

- [27] G. R. North, “Empirical orthogonal functions and normal modes,” *Journal of Atmospheric Sciences*, vol. 41, no. 5, pp. 879 – 887, 1984. [Online]. Available: https://journals.ametsoc.org/view/journals/atsc/41/5/1520-0469_1984_041_0879_eofanm_2_0_co_2.xml
- [28] S. S. P. Shen and G. R. North, *Statistics and Data Visualization in Climate Science with R and Python*. Cambridge University Press, 2023.
- [29] S. S. P. Shen and R. C. Somerville, *Climate Mathematics: Theory and Applications*. Cambridge University Press, 2019.
- [30] M. A. Alexander, I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, “The atmospheric bridge: The influence of ENSO teleconnections on air–sea interaction over the global oceans,” *Journal of Climate*, vol. 15, no. 16, pp. 2205 – 2231, 2002. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/15/16/1520-0442_2002_015_2205_tabtio_2.0.co_2.xml
- [31] E. Di Lorenzo, G. Liguori, and N. Mantua, “Climate interpretation of the North Pacific marine heatwave of 2013-2015,” *US CLIVAR’s Variations Newsletter*, vol. 14, no. 2, pp. 13–18, 2016.
- [32] K. Takahashi, A. Montecinos, K. Goubanova, and B. Dewitte, “ENSO regimes: Reinterpreting the canonical and Modoki El Niño,” *Geophysical Research Letters*, vol. 38, no. 10, 2011. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL047364>
- [33] S. S. P. Shen, G. P. Behm, Y. T. Song, and T. Qu, “A dynamically consistent reconstruction of ocean temperature,” *Journal of Atmospheric and Oceanic Technology*, vol. 34, no. 5, pp. 1061 – 1082, 2017. [Online]. Available: <https://journals.ametsoc.org/view/journals/atot/34/5/jtech-d-16-0133.1.xml>
- [34] S. Ladjal, A. Newson, and C.-H. Pham, “A PCA-like autoencoder,” *CoRR*, vol. abs/1904.01277, 2019. [Online]. Available: <http://arxiv.org/abs/1904.01277>
- [35] A. Banerjee, W. Ding, J. Dy, V. Lyubchich, and A. Rhines, Eds., *Proceedings of the 6th International Workshop on Climate Informatics: CI 2016*, 2016. [Online]. Available: <https://doi.org/10.5065/D6K072N6>
- [36] Q. Zhang, M. W. Berry, B. T. Lamb, and T. Samuel, “A parallel nonnegative tensor factorization algorithm for mining global climate data,” in *Computational Science – ICCS 2009*, G. Allen, J. Nabrzyski, E. Seidel, G. D. van Albada, J. Dongarra, and P. M. A. Sloot, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 405–415. [Online]. Available: https://doi.org/10.1007/978-3-642-01973-9_45
- [37] D. Li, Z. Yu, F. Wu, W. Luo, Y. Hu, and L. Yuan, “The tensor-based feature analysis of spatiotemporal field data with heterogeneity,” *Earth and Space Science*, vol. 7, no. 2, p. e2019EA001037, 2020. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EA001037>

- [38] A. Cichocki, “Era of big data processing: A new approach via tensor networks and tensor decompositions,” *CoRR*, vol. abs/1403.2048, 2014. [Online]. Available: <http://arxiv.org/abs/1403.2048>
- [39] O. Shamir, “Fast stochastic algorithms for SVD and PCA: Convergence properties and convexity,” *CoRR*, vol. abs/1507.08788, 2015. [Online]. Available: <http://arxiv.org/abs/1507.08788>
- [40] V. Rokhlin, A. Szlam, and M. Tygert, “A randomized algorithm for principal component analysis,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 3, pp. 1100–1124, 2010.
- [41] N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011. [Online]. Available: <https://arxiv.org/abs/0909.4061>
- [42] J. E. Kutzbach, “Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America,” *Journal of Applied Meteorology and Climatology*, vol. 6, no. 5, pp. 791 – 802, 1967. [Online]. Available: https://journals.ametsoc.org/view/journals/apme/6/5/1520-0450_1967_006_0791_eeoslp_2_0_co_2.xml
- [43] Z. Zhang and J. C. Moore, “Empirical orthogonal functions,” in *Mathematical and Physical Fundamentals of Climate Change*. Boston: Elsevier, 2015, pp. 161–197. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128000663000061>
- [44] C. Houser, A. Smith, P. Wernette, and J. Lehner, *1.08 - Spatial Frequency Analysis and Information Synthesis for Understanding Coastal Barriers*, second edition ed., J. J. F. Shroder, Ed. Oxford: Academic Press, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780128182345000237>
- [45] S. D. Meyers, J. J. O’Brien, and E. Thelin, “Reconstruction of monthly SST in the tropical Pacific ocean during 1868–1993 using adaptive climate basis functions,” *Monthly Weather Review*, vol. 127, no. 7, pp. 1599 – 1612, 1999. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/127/7/1520-0493_1999_127_1599_romsit_2.0.co_2.xml
- [46] D. Mu, H. Yan, and W. Feng, “Assessment of sea level variability derived by eof reconstruction,” *Geophysical Journal International*, vol. 214, no. 1, pp. 79–87, 2018. [Online]. Available: <https://doi.org/10.1093/gji/ggy126>
- [47] S. S. P. Shen and R. Somerville, *Climate Mathematics: Theory and Applications*. Cambridge University Press, 2019.
- [48] NOAA National Weather Service Climate Prediction Center, “El Niño-Southern Oscillation (ENSO),” 2005, <https://www.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/enso.shtml>. Accessed May 15, 2022.

- [49] E. Acar and B. Yener, “Unsupervised multiway data analysis: A literature survey,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 6–20, 2009.
- [50] A. Vallejo-Arboleda, J. L. Vicente-Villardón, and M. Galindo-Villardón, “Canonical STATIS: Biplot analysis of multi-table group structured data based on STATIS-ACT methodology,” *Computational statistics & data analysis*, vol. 51, no. 9, pp. 4193–4205, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167947306001319>
- [51] C. Lavit, Y. Escoufier, R. Sabatier, and P. Traissac, “The ACT (STATIS method),” *Computational Statistics & Data Analysis*, vol. 18, no. 1, pp. 97–119, 1994. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0167947394901341>
- [52] A. Gittens, A. Devarakonda, E. Racah, M. Ringenburt, L. Gerhardt, J. Kottalam, J. Liu, K. Maschhoff, S. Canon, J. Chhugani *et al.*, “Matrix factorizations at scale: A comparison of scientific data analytics in Spark and C+ MPI using three case studies,” in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE, 2016, pp. 204–213.
- [53] I. Stanimirova, B. Walczak, D. Massart, V. Simeonov, C. Saby, and E. Di Crescenzo, “STATIS, a three-way method for data analysis. application to environmental data,” *Chemometrics and intelligent laboratory systems*, vol. 73, no. 2, pp. 219–233, 2004. [Online]. Available: <https://doi.org/10.1016/j.chemolab.2004.03.005>
- [54] Y. T. Song and T. Y. Hou, “Parametric vertical coordinate formulation for multiscale, boussinesq, and non-boussinesq ocean modeling,” *Ocean Modelling*, vol. 11, no. 3, pp. 298–332, 2006. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1463500305000107>
- [55] Y. T. Song, R. Gross, X. Wang, and V. Zlotnicki, “A non-boussinesq terrain-following OGCM for oceanographic and geodetic applications,” *Advances in Geosciences*, vol. 18, pp. 63–86, 2010.
- [56] Y. T. Song and D. Haidvogel, “A semi-implicit ocean circulation model using a generalized topography-following coordinate system,” *Journal of Computational Physics*, vol. 115, no. 1, pp. 228–244, 1994. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021999184711892>
- [57] A. F. Shchepetkin and J. C. McWilliams, “A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate,” *Journal of Geophysical Research: Oceans*, vol. 108, no. C3, 2003. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JC001047>
- [58] D. B. Haidvogel, H. G. Arango, K. Hedstrom, A. Beckmann, P. Malanotte-Rizzoli, and A. F. Shchepetkin, “Model evaluation experiments in the North Atlantic basin: simulations in nonlinear terrain-following coordinates,” *Dynamics*

- of Atmospheres and Oceans*, vol. 32, no. 3, pp. 239–281, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S037702650000049X>
- [59] R. X. Huang and X. Jin, “Sea surface elevation and bottom pressure anomalies due to thermohaline forcing. Part I: Isolated perturbations,” *Journal of Physical Oceanography - J PHYS OCEANOGR*, vol. 32, 2002.
- [60] J.-H. Moon, Y. T. Song, P. D. Bromirski, and A. J. Miller, “Multidecadal regional sea level shifts in the Pacific over 1958–2008,” *Journal of Geophysical Research: Oceans*, vol. 118, no. 12, pp. 7024–7035, 2013. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2013JC009297>
- [61] Y. T. Song and F. Colberg, “Deep ocean warming assessed from altimeters, gravity recovery and climate experiment, in situ measurements, and a non-boussinesq ocean general circulation model,” *Journal of Geophysical Research: Oceans*, vol. 116, no. C2, 2011. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006601>
- [62] S. Liang, X. Li, and J. Wang, “High-level land product integration methods,” in *Advanced Remote Sensing*. Boston: Academic Press, 2012, pp. 667–690. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B9780123859549000228>
- [63] G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng, “Sampling errors in the estimation of empirical orthogonal functions,” *Monthly weather review*, vol. 110, no. 7, pp. 699–706, 1982. [Online]. Available: https://journals.ametsoc.org/view/journals/mwre/110/7/1520-0493_1982_110_0699_seiteo_2_0_co_2.xml
- [64] R. H. Weisberg and L. Qiao, “Equatorial upwelling in the central Pacific estimated from moored velocity profilers,” *Journal of Physical Oceanography*, vol. 30, no. 1, pp. 105 – 124, 2000. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/30/1/1520-0485_2000_030_0105_euitcp_2.0.co_2.xml
- [65] American Meteorological Society, “Equatorial Upwelling. Glossary of Meteorology,” 2012, https://glossary.ametsoc.org/wiki/Equatorial_upwelling#:~:text=The%20rising%20of%20water%20along,either%20side%20of%20the%20equator. Accessed 1 May 2022.
- [66] K. Wyrtki and G. Eldyn, “Equatorial upwelling events in the central Pacific,” *Journal of Physical Oceanography*, vol. 12, no. 9, pp. 984 – 988, 1982. [Online]. Available: https://journals.ametsoc.org/view/journals/phoc/12/9/1520-0485_1982_012_0984_eueitc_2_0_co_2.xml
- [67] S. P. Khatiwala, F. W. Primeau, and M. Holzer, “Ventilation of the deep ocean constrained with tracer observations and implications for radiocarbon estimates of ideal mean age,” *Earth and Planetary Science Letters*, vol. 325–326, pp. 116–125, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0012821X12000672>

- [68] A. C. Naveira Garabato, G. A. MacGilchrist, P. J. Brown, D. G. Evans, A. J. S. Meijers, and J. D. Zika, “High-latitude ocean ventilation and its role in earth’s climate transitions,” *Philos. Trans. A Math. Phys. Eng. Sci.*, vol. 375, no. 2102, p. 20160324, 2017.
- [69] J. Marshall and F. Schott, “Open-ocean convection: Observations, theory, and models,” *Reviews of Geophysics*, vol. 37, no. 1, pp. 1–64, 1999. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98RG02739>
- [70] J. Koelling, D. Atamanchu, J. Karstensen, P. Handmann, and D. W. R. Wallace, “Oxygen export to the deep ocean following Labrador Sea water formation,” *Biogeosciences*, vol. 19, no. 2, pp. 437–454, 2022. [Online]. Available: <https://bg.copernicus.org/articles/19/437/2022/>
- [71] I. Yashayaev and J. W. Loder, “Further intensification of deep convection in the Labrador Sea in 2016,” *Geophysical Research Letters*, vol. 44, no. 3, pp. 1429–1438, 2017. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071668>
- [72] D. Roemmich and J. Gilson, “The global ocean imprint of ENSO,” *Geophysical Research Letters*, vol. 38, no. 13, 2011. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL047992>
- [73] L.-C. Wang, T. L. Dao, and J.-Y. Yu, “Continued weakening of the equatorial Pacific upwelling annual cycle in CMIP5 future projections,” *Scientific Reports*, vol. 12, no. 1, p. 15595, 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-19874-2>
- [74] NOAA National Weather Service Climate Prediction Center, “Cold and warm episodes by season,” 2001, https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php. Accessed May 15, 2022.
- [75] J. W. Hurrell, “Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation,” *Science*, vol. 269, no. 5224, pp. 676–679, 1995. [Online]. Available: <https://www.science.org/doi/abs/10.1126/science.269.5224.676>
- [76] J. Yu and H. Kao, “Decadal changes of ENSO persistence barrier in SST and ocean heat content indices: 1958–2001,” *Journal of Geophysical Research: Atmospheres*, vol. 112, no. D13, 2007. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2006JD007654>
- [77] B. Fox-Kemper, H. T. Hewitt, C. Xiao, G. Aðalgeirsdóttir, S. S. Drijfhout, T. L. Edwards, N. R. Golledge, M. Hemer, R. E. Kopp, G. Krinner, A. Mix, D. Notz, S. Nowicki, I. S. Nurhati, L. Ruiz, J.-B. Sallée, A. B. A. Slangen, and Y. Yu, “2021: Ocean, cryosphere and sea level change,” in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield,

- O. Yelekçi, R. Yu, and B. Zhou, Eds. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 2021, pp. 1211–1362.
- [78] Copernicus Marine Service, “Global ocean physics reanalysis,” 2021, <https://doi.org/10.48670/moi-00021>. Accessed June 2024.
 - [79] L. Jean-Michel, G. Eric, B.-B. Romain, G. Gilles, M. Angélique, D. Marie, B. Clément, H. Mathieu, L. G. Olivier, R. Charly, C. Tony, T. Charles-Emmanuel, G. Florent, R. Giovanni, B. Mounir, D. Yann, and L. T. Pierre-Yves, “The copernicus global 1/12° oceanic and sea ice glorsy12 reanalysis,” *Frontiers in Earth Science*, vol. 9, 2021.
 - [80] L. Yu and M. M. Rienecker, “Mechanisms for the Indian Ocean warming during the 1997–98 El Niño,” *Geophysical Research Letters*, vol. 26, no. 6, pp. 735–738, 1999. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999GL900072>
 - [81] F. Sun and J.-Y. Yu, “Impacts of Central America gap winds on the SST annual cycle in the eastern Pacific warm pool,” *Geophysical Research Letters*, vol. 33, no. 6, 2006. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005GL024700>
 - [82] M. G. Jacox, J. Fiechter, A. M. Moore, and C. A. Edwards, “ENSO and the California current coastal upwelling response,” *Journal of Geophysical Research: Oceans*, vol. 120, no. 3, p. 1691–1702, 2015.
 - [83] K. E. Trenberth and D. P. Stepaniak, “Indices of El Niño evolution,” *Journal of Climate*, vol. 14, no. 8, pp. 1697 – 1701, 2001. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/14/8/1520-0442_2001_014_1697_lieno_2.0.co_2.xml
 - [84] S.-W. Yeh, J.-S. Kug, B. Dewitte, M.-H. Kwon, B. P. Kirtman, and F.-F. Jin, “El Niño in a changing climate,” *Nature*, vol. 461, no. 7263, p. 511–514, 2009. [Online]. Available: <https://doi.org/10.1038/nature08316>
 - [85] N. Chen and A. J. Majda, “Simple dynamical models capturing the key features of the Central Pacific El Niño,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 42, pp. 11 732–11 737, 2016. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1614533113>
 - [86] H. Zhang, A. Clement, and P. Di Nezio, “The South Pacific meridional mode: A mechanism for ENSO-like variability,” *Journal of Climate*, vol. 27, no. 2, p. 769–783, 2014.
 - [87] Y. You and J. C. Furtado, “The South Pacific meridional mode and its role in tropical Pacific climate variability,” *Journal of Climate*, vol. 31, no. 24, p. 10141–10163, 2018. [Online]. Available: <https://journals.ametsoc.org/view/journals/clim/31/24/jcli-d-17-0860.1.xml>

- [88] J. Servain, I. Wainer, J. P. McCreary, and A. Dessier, “Relationship between the equatorial and meridional modes of climatic variability in the tropical Atlantic,” *Geophysical Research Letters*, vol. 26, no. 4, p. 485–488, 1999. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999GL900014>
- [89] B. Tian and K. Fan, “Why is the North Atlantic Oscillation more predictable in December?” *Atmosphere*, vol. 10, no. 8, 2019. [Online]. Available: <https://www.mdpi.com/2073-4433/10/8/477>
- [90] I. Richter, M. F. Stuecker, N. Takahashi, and N. Schneider, “Disentangling the North Pacific meridional mode from tropical Pacific variability,” *npj Climate and Atmospheric Science*, vol. 5, p. 94, 2022. [Online]. Available: <https://doi.org/10.1038/s41612-022-00317-8>
- [91] A. Capotondi, M. Newman, T. Xu, and E. Di Lorenzo, “An optimal precursor of northeast Pacific marine heatwaves and Central Pacific El Niño events,” *Geophysical Research Letters*, vol. 49, no. 5, p. e2021GL097350, 2022, e2021GL097350. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL097350>
- [92] C.-Y. Lin, J.-Y. Yu, and H.-H. Hsu, “CMIP5 model simulations of the Pacific meridional mode and its connection to the two types of ENSO,” *International Journal of Climatology*, vol. 35, no. 9, pp. 2352–2358, 2015. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.4130>
- [93] S.-W. Yeh, B. P. Kirtman, J.-S. Kug, W. Park, and M. Latif, “Natural variability of the central Pacific El Niño event on multi-centennial timescales,” *Geophysical Research Letters*, vol. 38, no. 2, 2011. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL045886>
- [94] C. H. Gregory, C. Artana, S. Lama, D. León-FonFay, J. Sala, F. Xiao, T. Xu, A. Capotondi, C. Martinez-Villalobos, and N. J. Holbrook, “Global marine heatwaves under different flavors of ENSO,” *Geophysical Research Letters*, vol. 51, no. 20, p. e2024GL110399, 2024, e2024GL110399. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL110399>
- [95] M. T. Julianto, S. Dhimas, A. Sopaheluwakan, S. Nurdianti, and P. Septiawan, “Identification of global warming contribution to the El Niño phenomenon using empirical orthogonal function analysis,” *Agromet*, vol. 35, no. 1, p. 11–19, 2021.
- [96] H.-Y. Kao and J.-Y. Yu, “Contrasting Eastern-Pacific and Central-Pacific Types of ENSO,” *Journal of Climate*, vol. 22, no. 3, pp. 615–632, 2009. [Online]. Available: <http://journals.ametsoc.org/doi/10.1175/2008JCLI2309.1>
- [97] A. Capotondi, P. D. Sardeshmukh, E. Di Lorenzo, A. C. Subramanian, and A. J. Miller, “Predictability of us west coast ocean temperatures is not solely due to ENSO,” *Scientific Reports*, vol. 9, no. 1, 2019. [Online]. Available: <https://www.nature.com/articles/s41598-019-47400-4>

- [98] D. J. Vimont, M. A. Alexander, and M. Newman, “Optimal growth of Central and East Pacific ENSO events,” *Geophysical Research Letters*, vol. 41, no. 11, pp. 4027–4034, 2014. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL059997>
- [99] B. T. Anderson, R. C. Perez, and A. Karspeck, “Triggering of El Niño onset through trade wind-induced charging of the equatorial Pacific,” *Geophysical Research Letters*, vol. 40, no. 6, pp. 1212–1216, 2013. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/grl.50200>
- [100] M. Alexander, H. Seo, S.-P. Xie, and J. Scott, “ENSO’s impact on the gap wind regions of the eastern tropical Pacific ocean,” *Journal of Climate*, vol. 25, pp. 3549–3565, 2012.
- [101] D. J. Amaya, M. G. Jacox, J. Dias, M. A. Alexander, K. B. Karnauskas, J. D. Scott, and M. Gehne, “Subseasonal-to-seasonal forecast skill in the California Current System and its connection to coastal kelvin waves,” *Journal of Geophysical Research: Oceans*, vol. 127, no. 1, p. e2021JC017892, 2022, e2021JC017892 2021JC017892. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JC017892>
- [102] M. L. Batteen, “Wind-forced modeling studies of currents, meanders, and eddies in the California Current system,” *Journal of Geophysical Research: Oceans*, vol. 102, no. C1, pp. 985–1010, 1997. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96JC02803>
- [103] M. G. Jacox, E. L. Hazen, K. D. Zaba, D. L. Rudnick, C. A. Edwards, A. M. Moore, and S. J. Bograd, “Impacts of the 2015–2016 El Niño on the California Current System: Early assessment and comparison to past events,” *Geophysical Research Letters*, vol. 43, no. 13, p. 7072–7080, 2016.
- [104] J. Feng, Z. Wu, and X. Zou, “Sea surface temperature anomalies off Baja California: A possible precursor of ENSO,” *Journal of the Atmospheric Sciences*, vol. 71, no. 5, pp. 1529 – 1537, 2014. [Online]. Available: <https://journals.ametsoc.org/view/journals/atsc/71/5/jas-d-13-0397.1.xml>
- [105] J. G. Acker and G. Leptoukh, “Online analysis enhances use of NASA earth science data,” *Eos, Transactions American Geophysical Union*, vol. 88, no. 2, pp. 14 – 17, 2007. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007EO020003>
- [106] Worldview, “Welcome to Worldview,” *NASA Earth Science Data and Information System (ESDIS)*, 2024, <https://worldview.earthdata.nasa.gov>. Accessed September 10, 2025.
- [107] C. A. Smith, G. P. Compo, and D. K. Hooper, “Web-based reanalysis intercomparison tools (WRIT) for analysis and comparison of reanalyses and other datasets,” *Bulletin of the American Meteorological Society*, vol. 95, no. 11, pp. 1671 – 1678, 2014. [Online]. Available: <https://journals.ametsoc.org/view/journals/bams/95/11/bams-d-13-00192.1.xml>

- [108] J. Pierret and S. S. P. Shen, “4D visual delivery of big climate data: A fast web database application system,” *Advances in Data Science and Adaptive Analysis*, vol. 09, no. 03, p. 1750006, 2017. [Online]. Available: <https://doi.org/10.1142/S2424922X17500061>
- [109] S. S. P. Shen, J. Pierret, I. Dorado *et al.*, “4DVD visualization and delivery of the 20th century reanalysis data: methods and examples,” *Theoretical and Applied Climatology*, vol. 142, pp. 243–254, 2020. [Online]. Available: <https://doi.org/10.1007/s00704-020-03288-z>
- [110] Panoply, “Panoply netCDF, HDF and GRIB Data Viewer,” *NASA Goddard Institute for Space Studies*, 2024, <https://www.giss.nasa.gov/tools/panoply/>. Accessed September 10, 2025.
- [111] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [112] D. W. Behringer, M. Ji, and A. Leetmaa, “An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system,” *Monthly Weather Review*, vol. 126, pp. 1013–1021, 1998.
- [113] P. Podrzaj, “A brief demonstration of some Python GUI libraries,” in *Proceedings of the 8th International Conference on Informatics and Applications ICIA2019*, 2019, pp. 1–6.
- [114] N. Chen and A. J. Majda, “Simple dynamical models capturing the key features of the Central Pacific El Niño,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 42, pp. 11 732–11 737, 2016. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.1614533113>

Appendix A

Partition Input Method for Computing 3-Dimensional Empirical Orthogonal Functions for Big Space-Time Data: An Example of High-Resolution Ocean Assimilation Output

Empirical orthogonal functions were computed layer-by-layer from the Global Ocean Physical reanalysis (GLORYS) boreal winter ocean temperatures in Chapter 3. The calculation was done for the global domain for all 50 depth layers from surface to 5727 meters, using December-January-February (DJF) boreal winter mean of GLORYS ocean temperature data from 1993/1994 boreal winter to 2020/2021. Each layer produces 28 2D EOFs and 28 PCs.

A correlation analysis was employed using the resulting 2D PC with Central Pacific El Niño indices, meridional mode indices, and 3D PCs. The meridional mode indices are computed

using methods outlined in Section 3.3.2. The two Central Pacific El Niño indices used are the El Niño Modoki Index (EMI) and the Trans-Niño index (TNI). The EMI is derived by [6] to capture the structure of Central Pacific El Niño through three areas: A (10°S - 10°N, 165°E - 140°W), B (15°S - 5°N, 110° - 70°W), and D (10°S - 20°N, 125° - 145°E).

$$EMI = SST A_A - 0.5[SST A_B] - 0.5[SST A_D] \quad (A.1)$$

From [83] the TNI is defined as:

$$TNI = SST A_{1+2_N} - SST A_{4_N} \quad (A.2)$$

Where N refers to the normalization of the sea surface temperature anomalies in the Niño 1+2 (0° - 10°S, 90° - 80°W) and Niño 4 (5°S - 5°N, 160°E - 150°W) region. TNI is further normalized to have unit standard deviation.

The correlation analysis between 2D PC 3 and the various time series resulted in similar correlations. The starting layer for significant correlations varied, some starting as early as 22 meters and the latest at 34 meters. To show the progression of spatial patterns from the surface to the starting correlation layer, Fig. A.1 was created. From this figure we observe the evolution of the central tropical Pacific from cold to warm.

All correlations also exhibit a local maximum at 40 meters. Figure A.1 demonstrates how 2D EOF 3 progresses to a north Pacific meridional mode (NPMM) signal at 40 meters, supporting the maximum correlation. Additional layers after 40 meters were included to show the continued evolution of central tropical Pacific warming that progresses to resemble the NPMM and Central Pacific El Niño at 56 meters (Fig. A.1(r)).

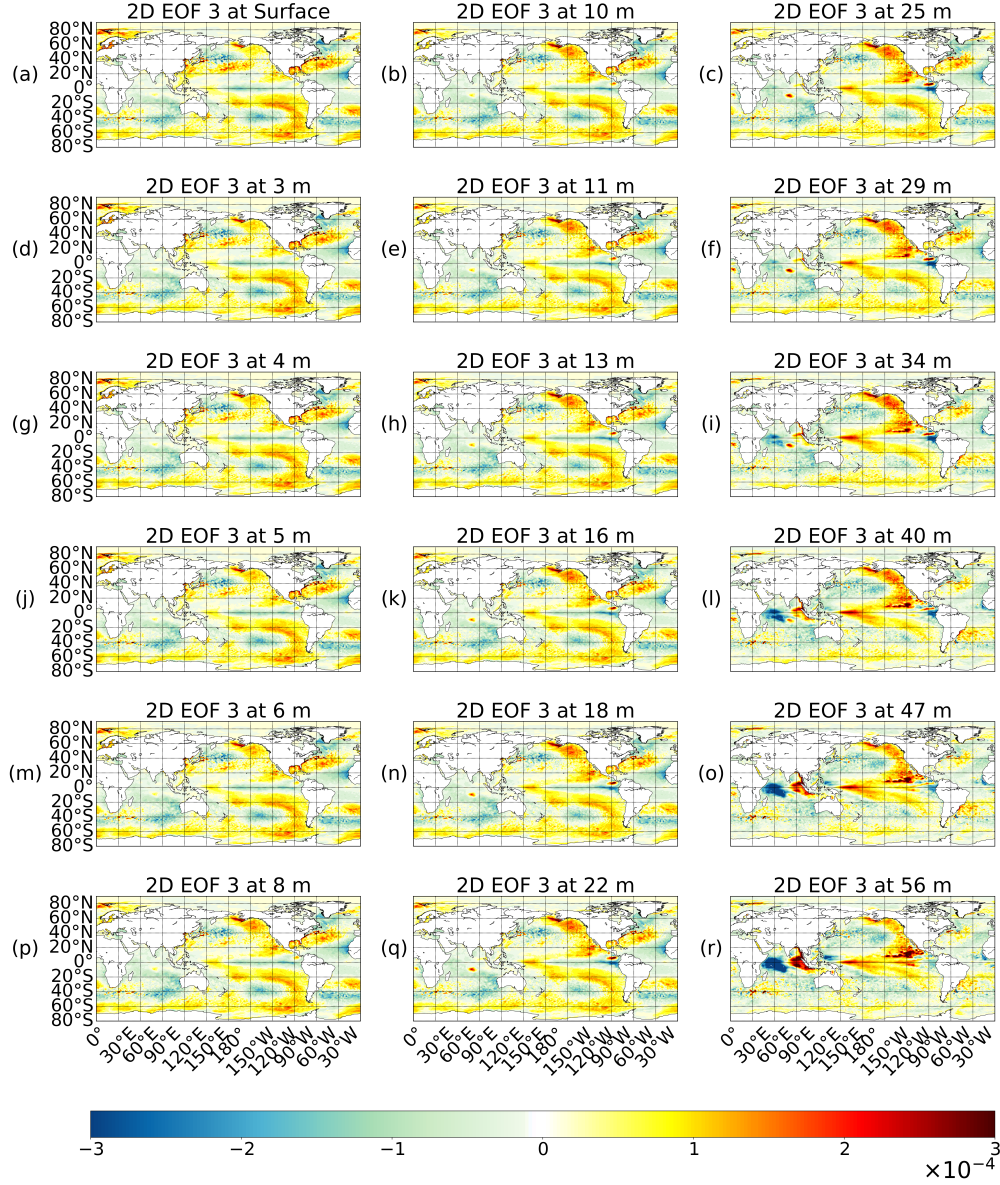


Figure A.1: 2D EOF 3 for the first 56 meters. EOFs are displayed to be read top to bottom, then left to right to simulate the progression of depth layers. Depths were chosen to be displayed at the first 8 meters (left column), 10 - 22 meters (middle column), and 25 - 56 meters (right column). These depths are chosen to show the progression of central tropical Pacific warming into NPM and Central Pacific El Niño features.

Appendix B

Four-dimensional North Pacific Variability and Regional Impacts from a High-Resolution Ocean Reanalysis

The Fig. B.1 contains results for a three-dimensional calculation of the E index using a 45° rotation of the two leading EOFs following the methods outlined in [32]. The E index captures the Eastern Pacific (EP) El Niño pattern. Figure B.1 are zonal cross-sections of the E pattern taken at 40°N and 37°N to analyze the impact of EP El Niño on the U.S. West Coast, specifically on the permanent meanders in the California Current System. The permanent meanders latitudes were defined according to [3]. The cross-sections extend from surface to 222 meters and use a smaller colorbar range than Fig. 4.7 to better visualize the meander at 40°N . The 40°N cut shows consistent warming in the first 50 meters between $\sim 128.5^\circ - 127.5^\circ\text{W}$ and maximum warming at 75 - 100 meters between $\sim 129^\circ - 127^\circ\text{W}$. The cut at 37°N has consistent warming in the first 30 meters between $\sim 126^\circ - 125^\circ\text{W}$ and reaches its maximum at the thermocline (60 - 80 meters) at $\sim 126^\circ - 125.5^\circ\text{W}$.

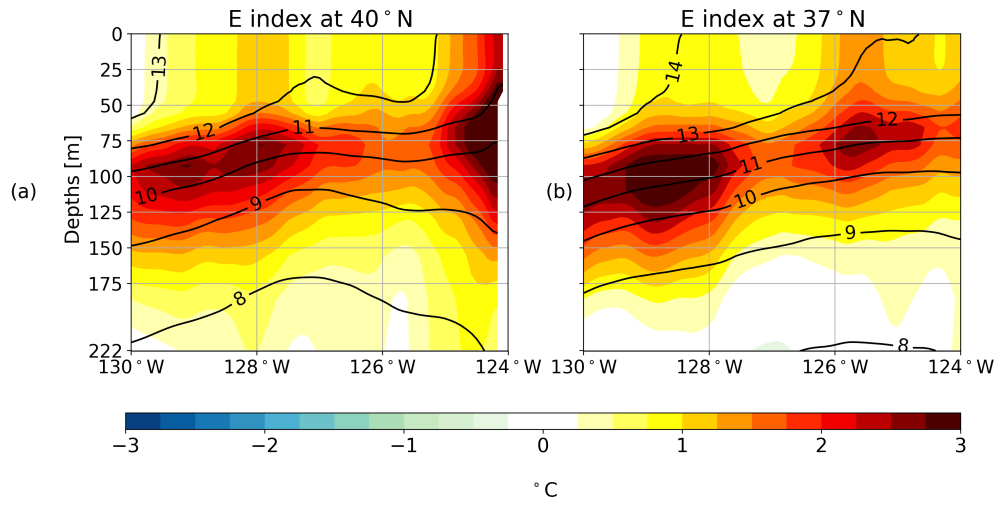


Figure B.1: Zonal cuts for the E index at (a) 40°N and (b) 37°N. These cuts are the same as in Fig. 4.7 (e) and (g) but with a smaller depth and color bar range to emphasize the effects of El Niño on the meanders in the California current system. Longitudes start at 130°W to match Fig. 4.7, but end at 124°W to eliminate whitespace from the land data.

Appendix C

Development of 3D Visualization Tools for Ocean Data

Figure C.1 displays the **Compare** interface of XSLICE for GLORYS with the first three 3D EOFs computed globally from GLORYS data. The interface uses **EOF** tags (gray boxes reading “3D EOF 1,” “3D EOF 2,” and “3D EOF 3”) to control the arrangement of figures, with each figure positioned according to its tag order. Initially, the **EOF** tags appear in numerical sequence: 3D EOF 1, 3D EOF 2, and 3D EOF 3. This places the corresponding figures as follows: 3D EOF 1 in the upper left, 3D EOF 2 in the upper right, and 3D EOF 3 in the lower left. Users can reorder the figures by dragging the **EOF** tags to their desired positions. Figure C.1 shows the result of switching the 3D EOF 2 and 3D EOF 3 tags. This repositions 3D EOF 3 in the upper right and 3D EOF 2 in the lower left.

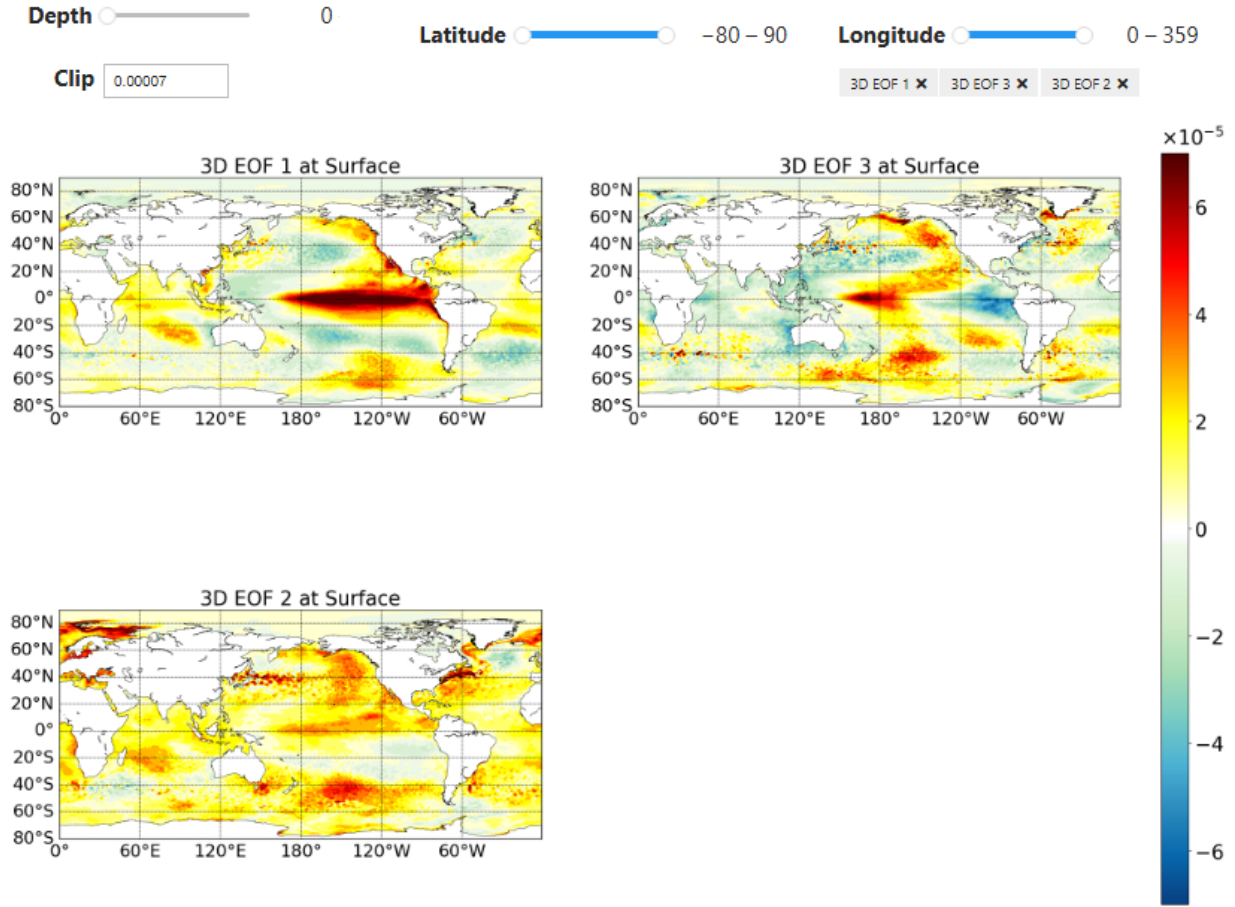


Figure C.1: The **Compare** interface displayed to demonstrate the interaction of reordering **EOF** tags. The top two figures reflect the positions of the first two tags, and the bottom figure reflects the position of the third tag.

The **Compare** interface layout was developed to optimize visualization space. Since depth cross-sections require more horizontal than vertical space, visualizations with three or more figures are placed in a layout with two columns and variable rows. Using additional columns significantly reduces individual figure size. Figure C.2 displays the first three 3D EOFs in a three-column layout. Compared to the 2×2 layout employed in the interface, the figures in the three-column layout are much smaller, which compromises the ability to visually analyze the spatial patterns.

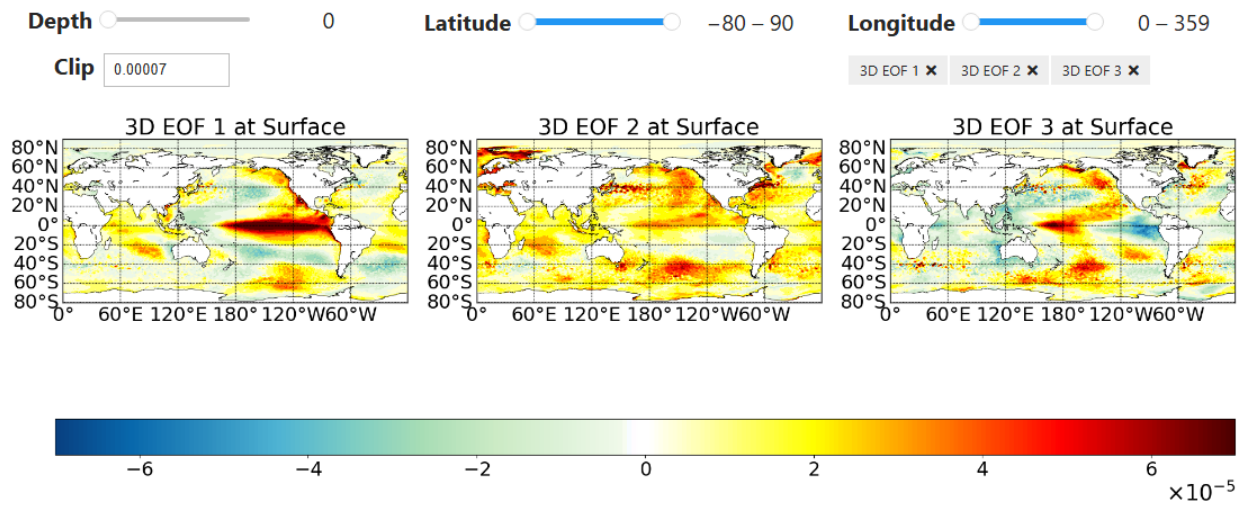


Figure C.2: A three-figure layout with three columns for the **Compare** interface. Each EOF figure is resized to fit the visualization space and are fairly small. To ensure the interface produces a large interpretable figure, a two-column layout was used instead.

Since 2D EOF calculations are performed separately for each depth layer, they produce discontinuous pattern signs across depth layers. This discontinuity requires a preprocessing step to correct the signs before meaningful analysis can occur. The correction process involves scanning through each 2D EOF depth layer and comparing the pattern to known ocean dynamics, 3D EOFs, and the previous 2D EOF layer. When a sign correction is needed, the entire layer is multiplied by -1.

This preprocessing requirement means that all zonal and meridional cross-section figures for 2D EOFs are not immediately useful for signal analysis. Figure C.3 demonstrates this problem by displaying 2D EOF 1 without sign correction. From the figure, it is unclear

which depths require correction or what oceanic dynamics are present. This limitation not only demonstrates the importance of the sign correction step but also justifies the motivation to prioritize the depth cross-section implementation in the **Compare** interface over zonal and meridional cross-section implementations. Among the three cross-section types, only the depth cross-section produces interpretable figures for the sign correction procedure.

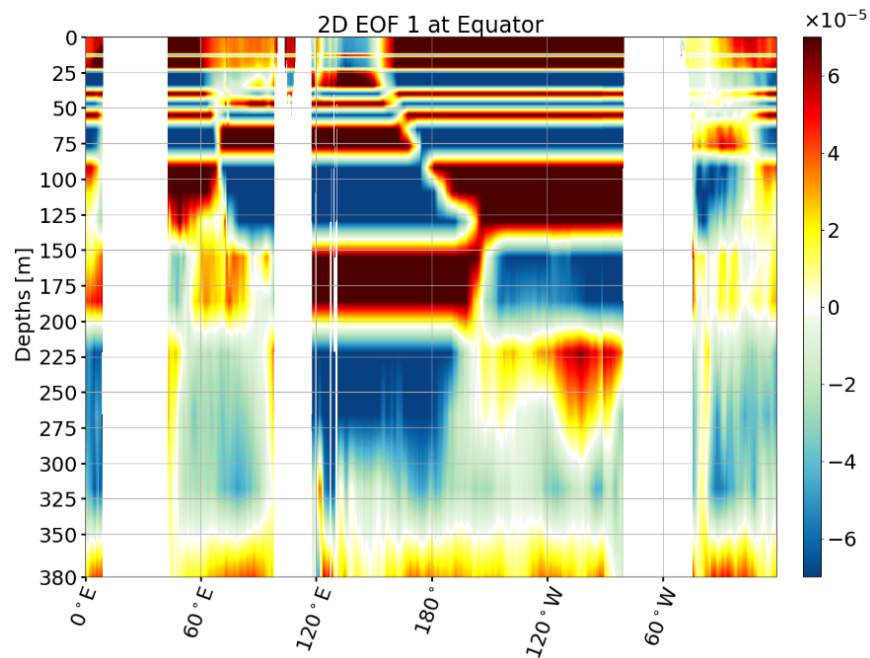


Figure C.3: Unprocessed 2D EOF 1 for a zonal cross-section at the equator. The layer signs have not been corrected, resulting in inconsistent patterns across depth layers. After proper sign correction, the cross-section reveals the El Niño dipole pattern shown in Fig. 3.10(d).