Calculation, Visualization, and Interpretation of Three-Dimensional Air-Ocean Coupled Empirical Orthogonal Functions Using the Data of General Circulation Models

Thomas Q. Bui[1], Danielle Lafarga[1†], Thomas M. Smith[2†], Yuhe T. Song[3†] and Samuel S.P. Shen[1*†]

[1*]Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA.
[2]NOAA/STAR/SCSB and CISESS/ESSIC, University of Maryland, 5825 University Research CT, College Park, MD 20740, USA.
[3]NASA Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA.

*Corresponding author(s). E-mail(s): sshen@sdsu.edu;
Contributing authors: tbui6742@sdsu.edu; dlafarga@sdsu.edu; tom.smith@noaa.gov; yuhe.t.song@nasa.gov;
†These authors contributed equally to this work.

Abstract

This paper presents 3-dimensional (3D) empirical orthogonal functions (EOFs) for two cases: (A) an atmospheric 3D EOF using the atmospheric model data from the surface of approximately 1,000 millibar [mb] (i.e., 1.0 bar, or 1,000 hPa in the metric unit) atmospheric pressure level to the 1.0 mb level about 50 kilometers from the sea level, and (B) a coupled atmosphere-ocean 3D EOF when the top layer of an ocean model is taken into account. These 3D EOFs help describe the 3D structure of the large-scale climate dynamic patterns, such as the Hadley Cells, Walker circulation, and the vertical profiles of the El Niño Southern Oscillation (ENSO). The 3D EOF tool developed in this research allows people to quantitatively describe and visually display air-ocean dynamics such as El Nio and the Hadley cells, particularly over

the tropical Pacific. Computationally, 3D EOFs are built on a 3D covariance matrix that naturally take the correlation between different pressure levels. In this way, 3D EOFs can help answer questions that cannot be answered by the conventional 2D EOFs computed layer-by-layer.

Keywords: Empirical Orthogonal Functions, Covariance, Hadley Cell, El Nio

1 Introduction

Empirical orthogonal functions (EOFs) are usually defined as orthonormal eigenvectors of a spatial covariance matrix. The first a few EOFs often have physical interpretations, such as illustrating spatial patterns of climate variability (Shen and Somerville (2019), Zhang and Moore (2015)). The eigenvalue associated with an EOF quantifies the variance of the climate system described by this EOF mode (Shea (2013)). Often the first few modes generally describe physically meaningful patterns in the dataset, and the latter modes do not (Liang et al (2012)). In physics, they are typically regarded as noise. However, mathematically, they can still be useful, such as data reconstruction (Shen et al (2017)).

In the existing literature, EOFs are typically computed in 2D space such as in Shen et al (2017), i.e., EOFs are computed from data of a layer of climate model output. However, climate dynamics, such as El Niño Southern Oscillation (ENSO), have vertical structures. The atmospheric temperature at one layer is correlated with another. Thus, the 3D covariance should be considered. The purpose of this paper is to compute EOFs in 3D space and to describe the climate dynamics represented by the EOFs. Our 3D EOFs can help quantify the interactions between sea surface temperature and the atmospheric temperature at different pressure layers up to 1.0 mbar.

Covariance is a measure of how two random variables interact with one another. This paper concerns how the temperature at one data point varies with the temperature at a separate data point. The monthly 3D climate model used has a large number of grid boxes. Computationally, to handle the eigenvalue of a smaller matrix, eigenvalues are computed from a temporal covariance matrix.

This paper will quantify and illustrate the spatial patterns given by the general atmospheric circulation model with the 5 meter sea surface temperature data serving as a boundary condition in comparison to the spatial patterns of the atmosphere only. The EOFs will be calculated for both air-ocean coupled data and air only data, in order to demonstrate the significance of air-ocean coupling. Due to a much higher heat capacity and much larger density of water compared to the air, one would expect that coupling water into the atmospheric model will significantly impact the spatial patterns depicted in the lower atmosphere. For this reason, one would expect that El Nio, which is known to be a significant factor of sea surface temperature variance, would account for much

more variance in the air-ocean coupled model than the air-only model. These are true and are quantitatively substantiated by our numerical results.

The remainder of this paper will be organized as such: Section 2, Data, which will describe the datasets and formatting of the space-time matrix; section 3, Methods, explaining the mathematics for the EOF compuations; section 4, Results, which goes in detail about the spatial patterns and quantifies their variances; and section 5, Conclusions and discussion.

2 Data

2.1 Temperature data from the NOAA GODAS model and the NOAA NCEP 20th Century Reanalysis V3

The data used for these calculations are two climate models: An atmospheric model and an oceanic model. The first is the air temperature data from the NOAA-CIRES-DOE Twentieth Century Reanalysis Version 3.0 (20CRV3) (Slivinski and coauthors (2019)). The 20CRV3 dataset has no missing data. The second is the potential temperature data from the National Centers for Environmental Prediction's (NCEP) Global Ocean Data Assimilation System (GODAS)(Behringer and Leetmaa (1998)). In the GODAS dataset, the missing data in the land region are represented with NaN.

The 20CRV3 air temperature dataset is a single netCDF file downloaded from https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html. The data covers the time from January 1836 to December 2015. The temperature data are in Kelvin and on a $1^\circ \times 1^\circ$ latitude-longitude grid ranging from $90^\circ S - 90^\circ N$ and from $0^\circ E - 359^\circ E$. The resulting data is an array with dimension given $2160 \times 28 \times 181 \times 360$. Here, 2160 is the total number of months from January 1836 to December 2015; 28 is the number of atmospheric pressure levels used in the 20CRV3, and the pressure levels are depicted in Table 1; 181 is the number of latitude coordinates from $90^\circ S - 90^\circ N$; and 360 is the number of longitude coordinates from $0^\circ E - 359^\circ E$.

Table 1: Twenty eight (28) pressure levels in mb of the 20CRV3 model

1000	975	950	925	900	850	800
750	700	650	600	550	500	450
400	350	300	250	200	150	100
70	50	30	20	10	5	1

The GODAS dataset of the ocean water potential temperature in Kelvin consists of a series of netCDF files, downloaded from https://psl.noaa.gov/data/gridded/data.godas.html. Each file is for a specific year with twelve months. The spatial resolution is a $\frac{1}{3}^\circ \times 1^\circ$ latitude-longitude grid ranging from $74.5^\circ S - 64.5^\circ N$ and $0.5^\circ E - 359.5^\circ E$. Each GODAS files is a data array of dimension $12 \times 40 \times 418 \times 360$,

2.2 Formulation of the space-time data

The GODAS data for all years ranges from January 1980 - December 2021. The dataset is still being updated. However, the 20CRV3 data ranges from January 1836 - December 2015. The overlap between the two datasets is from January 1980- December 2015, which is used for our EOF computing. The 20CRV3-GODAS combined space-time data matrix is created by concatenating each flattened month of the GODAS data to each flattened month of the 20CRV3 data. This leads to a $N \times Y$ space-time matrix in which N represents the total number of data points $N = (28 \times 181 \times 360) + (418 \times 360) = 1,974,960$; and Y represents the number of time steps. Since the data is measured monthly over the course of 36 years from 1980 to 2015, thus $Y = 36$ for any given month, say, January. We compute EOFs for each month, i.e., a set of EOFs for January, another for February, etc.; all together we compute 12 sets of EOFs.

For a given layer i of atmosphere at a given time t, the data form a matrix in the following format.

$$l_{i,t} = \begin{pmatrix} T_{-90,0} & T_{-90,1} & \dots & T_{-90,359} \\ T_{-89,0} & T_{-89,1} & \dots & T_{-87,359} \\ T_{-88,0} & T_{-88,1} & \dots & T_{-88,359} \\ \dots & \dots & \dots & \dots \\ T_{90,0} & T_{90,1} & \dots & T_{90,359.5} \end{pmatrix}_{181 \times 360} . \tag{1}$$

Here, 181 is the number of zonal grid lines, and 360 are the number of meridional grid lines, at the $1^\circ \times 1^\circ$ latitude-longitude resolution.

The data for the top layer of the ocean water are organized in a similar manner:

$$l_{w,t} = \begin{pmatrix} T_{-74.5,0.5} & T_{-74.5,1.5} & \dots & T_{-74.5,359.5} \\ T_{-74.167,0.5} & T_{-74.167,1.5} & \dots & T_{-74.167,359.5} \\ T_{-73.833,0.5} & T_{-73.833,1.5} & \dots & T_{-73.833,359.5} \\ \dots & \dots & \dots & \dots \\ T_{64.5,0.5} & T_{64.5,1.5} & \dots & T_{64.5,359.5} \end{pmatrix}_{418 \times 360} \tag{2}$$

where w indicates water, t is the discrete time in months, 418 is the number of the zonal grid lines, and 360 is the number of meridional grid lines. The spatial resolution is approximately $0.3325^\circ \times 1^\circ$ latitude-longitude. The matrix covers the ocean from 74.5°S to 64.5°N.

These matrices are flattened and stacked upon one another column-wise to create the following space-time data matrix for January:

$$X = \begin{pmatrix} l_{1,1980} & l_{1,1981} & \cdots & l_{l,2015} \\ l_{2,1980} & l_{2,1981} & \cdots & l_{2,2015} \\ l_{3,1980} & l_{3,1981} & \cdots & l_{3,2015} \\ \cdots & \cdots & \cdots & \cdots \\ l_{28,1980} & l_{28,1981} & \cdots & l_{28,2015} \\ l_{w,1980} & l_{w,1981} & \cdots & l_{w,2015} \end{pmatrix}_{1974960 \times 36} \tag{3}$$

This matrix has 1,974,960 rows, corresponding to the 1,974,960 grid boxes, and 360 columns, corresponding to 36 years from 1980 to 2015.

When considering only the atmospheric data, we use only the 20CRV3 dataset and exclude GODAS water data. For consistency, the same time frame $1980-2015$ is used for the EOF computations even though the entire temporal range of 20CRV3 is from $1836-2015$. This results in a slightly smaller space-time data matrix with dimensions $1,824,480 \times 36$.

3 Methods

3.1 Compute the weighted space-time anomaly matrix

From the initial raw space-time data matrix $X_{N\times Y}$, the climatology and standard deviation are computed, in order to compute the standardized anomalies from which the EOFs are computed. The climatology and standard deviation are $N \times 1$ vectors. The climatology is the arithmetic mean of the row data over time, i.e.,

$$\mu = [\mu_i]_{N\times 1} = \left[\frac{1}{Y}\sum_{t=1}^{Y} X_{it}\right]_{N\times 1} \tag{4}$$

with t representing each time step and Y being the total number of time steps.

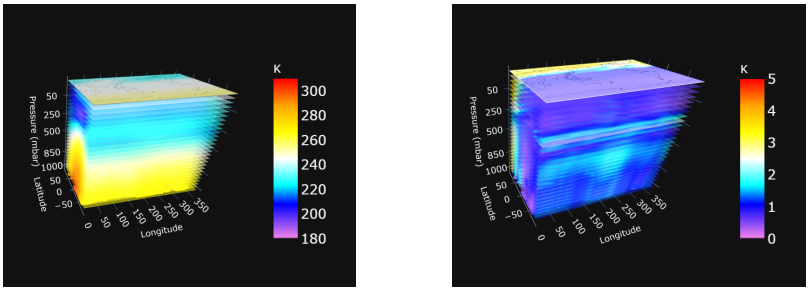
The standard deviations for each row are defined as follows:

$$S = [\sigma_i]_{N\times 1} = \left[\frac{1}{Y}\sum_{t=1}^{Y} (X_{it}-\mu_i)^2\right]_{N\times 1}^{\frac{1}{2}} \tag{5}$$

The standardized anomaly space-time matrix is computed in the following way:

$$\mathbf{T}_{N\times Y} = [T_{it}]_{N\times Y} = \left[\frac{X_{it}-\mu_i}{\sigma_i}\right]_{N\times Y} \tag{6}$$

The 3D grid boxes have different volumes, relevant to latitude and thickness of each layer. The air and water grid boxes have different heat capacity. These differences directly influence the data and their interpretation of climate dynamics. To accurately incorporate embed climate dynamics in our 3D



(a) January Climatology (b) January Standard Deviation

Fig. 1: The climatology and standard deviation for January shown above have been used to calculate the standardized anomalies

EOFs, we must take these differences into account. It is well known that 2D EOF computing for a large region requires an area factor $\sqrt{\cos(\phi_i)}$, where ϕ_i is the latitude of the centroid of the 2D grid box i. In this paper, we introduce a 3D weight factor that includes latitude, longitude, layer thickness, heat capacity, and density:

$$w_f = \sqrt{c_p \rho \Delta z \Delta\theta \Delta\phi \cos\phi}. \tag{7}$$

Here, c_p is the heat capacity which is 4.812 for water and 1.005 for air. This term is used to account for the large difference in heat capacity between air and water. The symbol ρ is the density of the material, applied not only because water has a higher density than air, but the density of the air changes with the pressure and altitude. The rest is applied to take account of the geometrical dimensions of each grid box. The data point is located at the centroid of a small volume of air or water, determined by

$$R_E^2 \Delta\phi \Delta\theta \Delta z. \tag{8}$$

Here, R_E is the radius of Earth, Δz is the thickness of a 3D grid box, $\Delta\theta$ is the longitude extent of the grid box, and $\Delta\phi$ is the latitude extent of the grid box. These three quantities $\Delta\phi, \Delta\theta, \Delta z$ are the spatial resolution of the study domain when we use the 3D gridded data. The function $\cos\phi$ is due to the spherical geometry of the Earth as the areas near the poles are smaller than the areas near the equator for the boxes of the same latitude and longitude extents. The expression

$$c_p \rho R_E^2 \Delta z \Delta\theta \Delta\phi \cos\phi X_{it} \tag{9}$$

is the heat energy inside the small volume of mass, while

$$\mathrm{C}_{th} = c_p \rho R_E^2 \Delta z \Delta\theta \Delta\phi \cos\phi \tag{10}$$

is the thermal mass of the volume. Therefore, our covariance matrix is weighted by thermal mass, a physical interpretation of our weighted covariance formulation. The weight factor of our space-time data is the square root of the thermal mass for every 3D grid box.

The Earth's radius R_E is served as a scaling factor and will not affect the eigenvalue problem for our EOF calculation. Thus, R_E will be dropped from here in this paper.

The data for air temperature in a climate model are usually organized by pressure levels in [mbar] instead of meters. In addition, the height will change over time because having a constant pressure with changing temperature would result in a change of altitude. To keep calculations simple, the altitude h_i of the top surface of 3D grid box i is approximated using the following hypsometric equation (Shen and Somerville (2019))

$$h_i = \frac{-RT}{g} \ln \frac{P_i}{P_{i-1}} + h_{i-1} \qquad i = 1, 2, ..., 28, \tag{11}$$

where h_{i-1} is the altitude of the bottom of 3D grid box i; P_i and P_{i-1} are the top and bottom pressures of 3D grid box i; $g = 9.8\ m/s$ is the acceleration of gravity; $R = 287.055\ J/(kgK)$ is the gas constant, and T is the temperature in Kelvin; 28 is the number of layers used in our atmospheric model.

The thickness for layer i is then

$$\Delta z_i = h_i - h_{i-1} = \frac{-RT}{g} \ln \left(\frac{P_i}{P_{i-1}} \right) \qquad i = 1, 2, \cdots 28, \tag{12}$$

Because the temperature changes over time, the temperature used to calculate the air height for each data point is the climatology or mean temperature over the 36 year time span. The initial pressure, P_0 is assumed to be 1013.25 mbar and the initial height, h_0, is assumed to be 0 meters.

The air density for each grid box is calculated through the equation

$$\rho_{\theta,\phi,p} = \frac{p}{0.2869 T_{\theta,\phi,p}} \tag{13}$$

using the climatology of each data point as the temperature $T_{\theta,\phi,p}$ and p to be the pressure at the given point. The water density is assumed to be a constant 1000 kg/m^3.

The weighted space-time anomaly matrix is then defined

$$\mathbf{T}_w = \left[T_{i,t} \sqrt{c_{p_i} \rho_i \Delta z_i \Delta\theta_i \Delta\phi_i \cos\phi_i} \right] \tag{14}$$

where $T_{i,t}$ are the entries of the standardized space-time anomaly matrix $\mathbf{T}_{N \times Y}$.

The temporal covariance matrix for the model data is

$$\mathbf{C} = \frac{1}{Y}[\mathbf{T}_w]^T[\mathbf{T}_w], \tag{15}$$

in which $[\mathbf{T}_w]^T$ is the transpose of the $\mathbf{T}_w$. Since $Y << N$, this greatly reduces the size of the matrix for computing eigenvectors and eigenvalues, which is usually done for a spatial covariance matrix.

For the computation of the eigenvectors, the $\frac{1}{Y}$ factor may be ignored as it is a scalar and the eigenvectors will be normalized in the computation. The eigenvectors are denoted

$$\overrightarrow{v_k} \qquad k = 1, 2, ..., Y. \tag{16}$$

These are principal components for the weighted anomalies.

The EOFs are the normalized projections of the space-time anomaly matrix

$$E_k = \frac{\mathbf{T}_w \overrightarrow{v_k}}{\|\mathbf{T}_w \overrightarrow{v_k}\|} \tag{17}$$

These are the geometric EOFs and PCs which are orthonormal vectors. To reveal the physical properties of the EOFs, the weight factors should be removed. They can be done in the following way:

$$\Psi_{k,i} = \frac{E_{k,i}}{\sqrt{c_{p_i}\rho_i \Delta z_i \Delta\theta_i \Delta\phi_i \cos\phi_i}}. \tag{18}$$

Here, $\Psi_{k,i}$ is the ith component of the kth physical EOF.

3.2 The theory for the weights

The eigenvalue problem of the weighted covariance function in a 3-D space is defined as follows:

$$\int_V \langle T_a(\mathbf{r}, t) T_a(\mathbf{r}', t)\rangle \Psi_n(\mathbf{r}', t) dW(\mathbf{r}') = \lambda_n \Psi_n(\mathbf{r}), \tag{19}$$

in which V is the 3-Dimensional domain under investigation which can be in the atmosphere, ocean, or atmosphere and ocean, $\mathbf{r}$ is the position vector for any spatial point inside V and is determined by latitude, longitude and elevation of the point, t is the time, and λ_n is the eigenvalue that corresponds to $\Psi_n(\mathbf{r})$.

Here, $T_a(\mathbf{r}, t)$ is the temperature anomaly field, $\langle\cdot\rangle$ stands for the expected value, $\Psi_n(\mathbf{r})$ are continuous eigenfunctions, $W(\mathbf{r})$ is the weight for the covariance function $\langle T_a(\mathbf{r}, t) T_a(\mathbf{r}', t)\rangle$ and is specified by

$$dW(\mathbf{r}) = c_p \rho \cos\phi dz d\theta d\phi, \tag{20}$$

The integral problem may be discretized as such:

$$\sum_{j=1}^{N} \langle T_a(\mathbf{r}_i, t) T_a(\mathbf{r}_j) \rangle \Psi_n(\mathbf{r}_j) \Delta W(\mathbf{r}_j) = \lambda_n \Psi_n(\mathbf{r}_i)$$

This equation can be re-organized into the following symmetric form between i and j:

$$\sum_{j=1}^{N} \left\langle \sqrt{\Delta W(\mathbf{r}_i)} T_a(\mathbf{r}_i, t) \sqrt{\Delta W(\mathbf{r}_j)} T_a(\mathbf{r}_j, t) \right\rangle \sqrt{\Delta W(\mathbf{r}_j)} \Psi_n(\mathbf{r}_j) \quad (21)$$

$$= \lambda_n \sqrt{\Delta W(\mathbf{r}_i)} \Psi_n(\mathbf{r}_i) \quad (22)$$

Denote

$$\mathbf{T}_w(\mathbf{r}_i, t) = \mathbf{T}_a(\mathbf{r}_i, t) \sqrt{\Delta W(\mathbf{r}_i)} \quad (23)$$

as the weighted temperature anomalies and denote

$$E_n(\mathbf{r}_i) = \Psi_n(\mathbf{r}_i) \sqrt{\Delta W(\mathbf{r}_i)} \quad (24)$$

as the geometric EOFs. Therefore, the computational procedure for $E_{k,i}$ (see Eqs. (17) and (18)) in the previous subsection is substantiated.

The geometric EOFs are computed through 3D covariance shown above. They possess the orthonormal properties:

$$\sum_{i=1}^{N} E_n^2(\mathbf{r}_i) = 1 \quad \text{(Normality)} \quad (25)$$

$$\sum_{i=1}^{N} E_m(\mathbf{r}_i) E_n(\mathbf{r}_i) = 1 \quad \text{when } m \neq n, \quad \text{(Orthogonality)}. \quad (26)$$

These two properties are often used to check numerical results of EOFs, because correct EOFs necessarily satisfy these conditions.

4 Results

4.1 Air-Ocean coupled EOFs

This method has been used to calculate EOFs and PCs for all months, but for the results section, this paper will primarily focus on the EOFs from January due to the presence of El Nio Southern Oscillation (ENSO) and the northern polar vortex bomb.

Figure 2 shows EOF1 of January temperature, ranging from the sea surface up to 10 mbar. From Figure 4, the first mode explains nearly 16% of the total variance. El Nio is a signal in the eastern Pacific (Jadhav et al (2015)), but explains less variance for the higher atmosphere. ENSO explains much of the

temperature variation when the sea surface data of the GODAS data is coupled with the 20CRV3. EOF1 shows a strong eastern Pacific El Nio pattern, rising in the atmosphere (see Fig. 5). As shown in Figs. 5a and 5b, the ENSO pattern weakens higher in the atmosphere. The cross sections in Figs. 5c and 5d no longer have the dominant ENSO shape as the lower atmosphere, but there still exists a warm bubble in the eastern Pacific from the El Nio signal. Beyond 100 mbar, the ENSO signal disappears. From Fig. 6b, a strong separation exists between the stratosphere and the troposphere. In Fig. 5f, there is another signal present. It is a region that seems to split over two main regions: the northern North America and the northeastern Asia. This may be the transition of the vortex split into two daughter vortices described in Matthewman et al (2009). Figure 6 has two vertical cross sections. Figure 6a is a zonal average from $180°E - 230°E$ and clearly shows a phenomenon symmetric about the equator. The cold regions that extend from approximately 15° to 40° are Hadley cells. The Hadley cells are formed as warm air shoots up from the equator, it cools and circulates downward and outward (Johanson and Fu (2009)). As shown in Fig. 6a, there is a major upward motion of warm air near the equator, splitting around 800 mbar and circulating back down into colder air. Figure 6b is a zonal average and clearly shows the effects of the El Nio signal rising through the atmosphere. The signal remains the strongest up until approximately 750 mbar in the primary red region and continues to rise up until approximately 150 mbar. This parallels the vertical cross section of the Walker circulation with westerly anomaly winds in the lower atmosphere reaching to the 800 mbar, and easterly anomaly winds in the higher atmosphere reaching up to 100 mbar (Bell et al (2014)). PC1 (see Fig. 3) shows oscillations approximately every three to four years, trending upward. This may be the result of ENSO as the El Nio signal is present in EOF1.

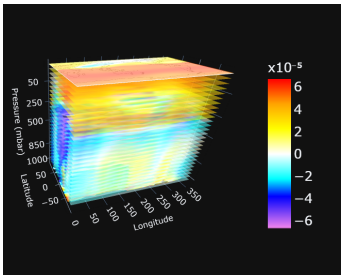


Fig. 2: January Air and Ocean EOF1

EOF2 in Fig. 7 shares similarities with EOF1 in terms of overall shape, but the dominant effects are inverted. From Figure 3, the second eigenvalue of the modes explains approximately 11% of the total variance. Just as in EOF1, ENSO is also present in EOF2. Figure 8 demonstrates a different pattern for EOF2 than from EOF1. In Fig. 8a and Fig. 8b, the warm El Nio signal grows as it reaches higher altitudes. This contrasts the shrinking signal in EOF1.

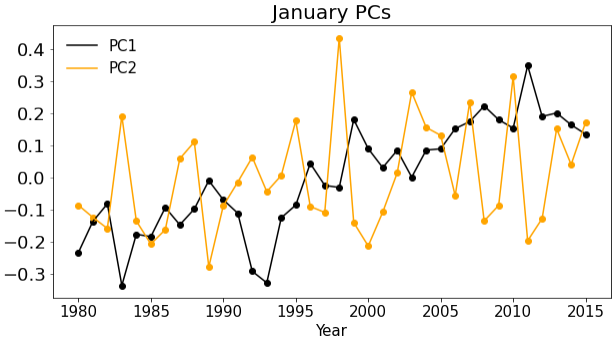


Fig. 3: January PC1 and PC2

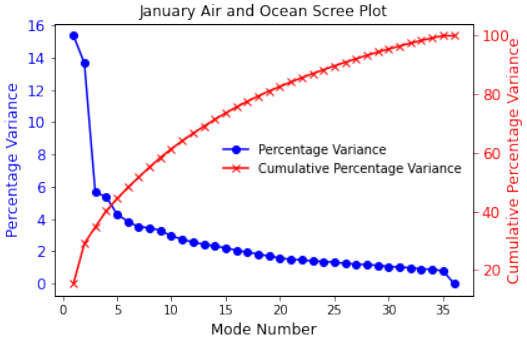


Fig. 4: January Air and Ocean Scree Plot

In Fig. 8c and Fig. 8d, there are two main cold signals in the northern and southern Pacific. This greatly contrasts with Fig. 5c and Fig. 5d in which the notable signal was the warm region in the eastern Pacific. In Fig. 8e, the 50 mbar cross section clearly resembles an inversion of the 250 mbar cross section (Fig. 8d). In Fig. 8f, the 10 mbar cross section has notable signals in the northern hemisphere. Unlike the northern polar vortex signal from Fig. 5f, these signals are mostly above northern Eurasia, east Asia, and Central America. Figure 9 shows the vertical cross sections for EOF2. Fig. 9a shows a Hadley in the central Pacific. Figure 9b shows a vertical cross section of the zonal average from $20°S - 10°N$. In contrast to Fig. 6b, Fig. 9b does not have one narrowing El Nio signal rising. As shown in Fig. 8, the warm El Nio signal grows and spreads across the Pacific. PC2 (see Fig. 3) follows a similar pattern to PC1 with oscillations occurring every three to four years. As EOF2 similarly demonstrates ENSO in the lower atmosphere, PC2 may be explaining the ENSO phenomenon as well.

Through comparing EOF1 and EOF2 of the air-ocean coupled model, there is clear modal mixing in the computations of the EOFs. The El Nio signal

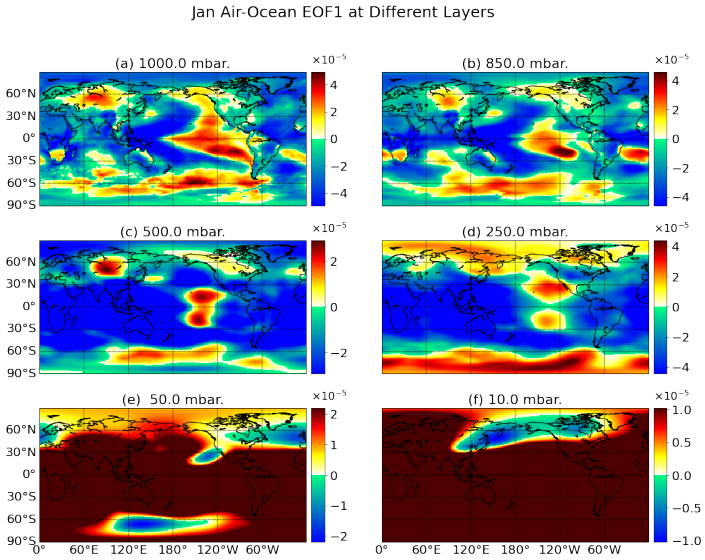


Fig. 5: The following layers are all subsections of the same EOF

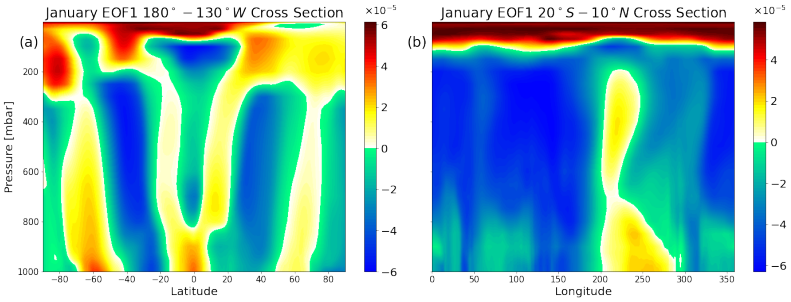


Fig. 6: These cross sections are proportional to pressure, not physical space

appears in both EOF1 and EOF2. This explains the relative closeness in percentage variance between modes 1 and 2. Thus, by adding the two mode percentage variances, one can conclude that El Nio accounts for approximately 30 % of the total variance of air and ocean temperature when coupled. Likewise, Hadley cells are present in the first two modes of the air-ocean model as well. The focus on the first two modes is primarily because in Fig. 4, the first two modes each account for over 12 % of the percentage variance while EOF3 and the following modes regard for less than 6 % of the percentage variance, a considerable drop in sizes of the eigenvalues. The air-only model results will

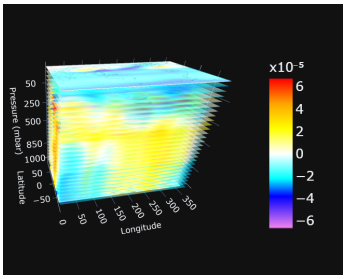


Fig. 7: January Air and Ocean EOF2

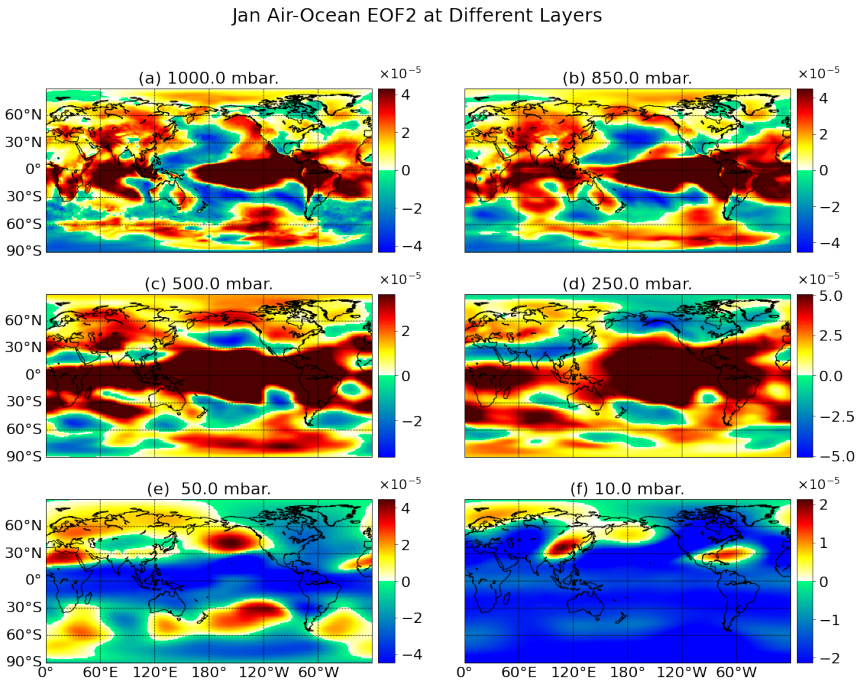


Fig. 8: The following layers are cross sections of EOF2

also only demonstrate the first two modes, but as the scree plot will illustrate, there will be considerable differences.

4.2 Air Only EOFs

The same EOF computations have been done to only the 20CRV3 air dataset. The $\Delta\phi$ and $\Delta\theta$ factors may be ignored for this specific set of calculations

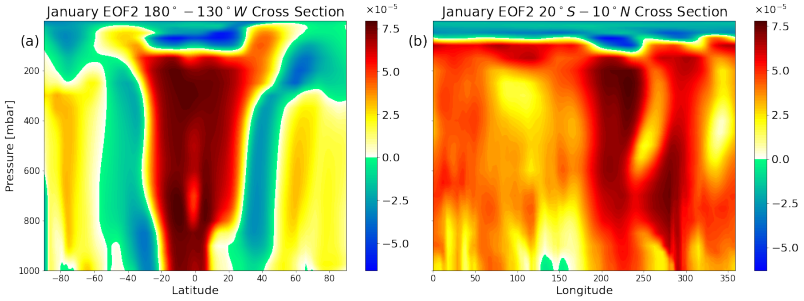


Fig. 9: Vertical Cross Sections of January Air-Ocean Coupled EOF2

since all of the data is on a $1° \times 1°$ grid. Likewise, the c_p was also ignored for computations since all of the data is atmospheric. In excluding the surface layer of the GODAS data, the overall 3D pattern seems similar to the 3D air-ocean coupled EOF Mode 1. However, further analysis of each layer and vertical cross sections will detail the significant differences in EOF Mode 1 due to the oceanic coupling.

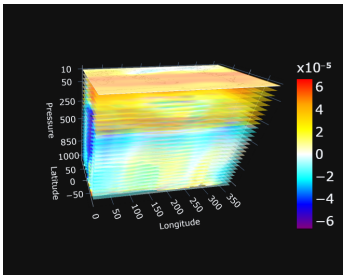


Fig. 10: January Air Temperature EOF1

Figure 10 shows EOF1 of January temperature ranging from the 1000 mbar up to the 10 mbar layer. From Figure 12, the first eigenvalue of the modes explains nearly 25% of the total variance. This is significantly higher than the first mode of the air-ocean coupled model, which only accounted for nearly 16 % of the total variance. The El Nio signal that was prominently in EOF1 of the air-ocean coupled model is no longer as strong in EOF1 of the air-only model. This is evidently clear when comparing Figs. 14b and 6b. The El Nio signal in the $20°S - 10°N$ zonal average is a much fainter green, rather than the prominent yellow. The dominant ENSO pattern in Figs. 5a and 5b are not present in Figs. 13a and 13b. Instead, there is another warming signal in the southern Pacific ocean in Figs. 13b and 13c . From Fig. 13a-c, there are strong signals in the southern Pacific and weaker signals in the northern Pacific. However, at 250 mbar, in Fig. 13d, the signals in the northern and southern

polar regions are much stronger and span all longitudes. From Fig. 14a, there is a clearer separation between the troposphere and the stratosphere than in 6. In Fig. 13f, the northern polar vortex bomb signal is present once more, but unlike in Fig. 5f, the vortex bomb is more centered over east Asia. This seems to suggest that oceanic coupling has a significant role in the northern polar vortex bomb in the North American region. The two vertical cross sections in Fig. 14 are also range from $180°E - 230°E$ and from $20°S - 10°N$. Unlike the air-ocean EOF1 vertical cross section (see Figure 6), the Hadley cells are not present in the air-only cross section of Fig. 14a. This suggests that Hadley cells are not as prominent in the atmosphere when ignoring the sea surface interactions over the central Pacific. In Fig. 14b, the El Nio signal is still present in the eastern Pacific, however, in comparison to Fig. 6b, the El Nio signal is much weaker compared to when sea surface temperature is coupled. PC1 (see Fig. 11) demonstrates a much stronger modulation than PC1 of the air-ocean coupled model (see Fig. 3) which may imply that the dynamics concerning the air-only mode 1 vary much greater than the dynamics of the air-ocean coupled mode 1.

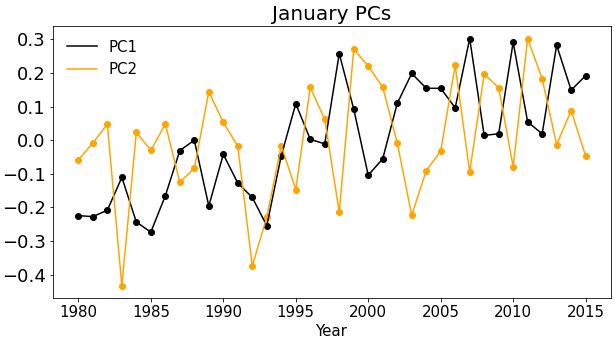


Fig. 11: January PC1 and PC2 with only 20CRV3

In Fig. 15, a major difference in the 20CRV3-only EOF2 is the much weaker separation between the troposphere and the stratosphere. From Fig. 12, the second eigenvalue of the modes explains approximately 11% of the total variance. Unlike in the air-ocean coupled model, the percentage variance difference between EOF1 and EOF2 for the air-only EOF1 and EOF2 is drastically large. In Fig. 3, the difference in percentage variance between mode 1 and mode 2 was approximately 2%. However, in Fig. 12, the difference is approximately 14% . This major difference is due to the modal mixing in the air-ocean coupled model. Unlike in EOF1, ENSO has a dominant presence in EOF2. Figure 16 is much more similar to Fig. 5 than Fig. 13. There is an ENSO pattern that weakens as it reaches the higher atmosphere. Figures 16c and 16d have the similar warm bubble in the eastern Pacific but is much larger. In Fig. 16f,

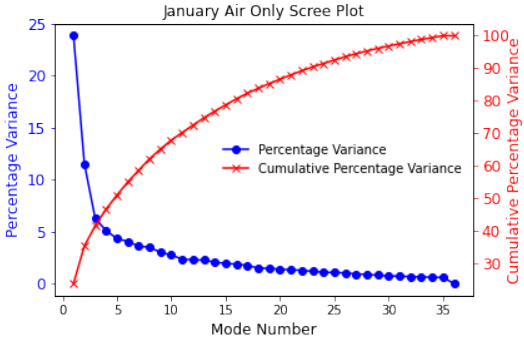


Fig. 12: January Air Only Scree Plot

there is a major cooling anomaly in the northern hemisphere over North America and Russia. The anomaly is negatively strongest over North America. The weaker separation between the troposphere and stratosphere seen in Fig.12 is apparent again in Fig. 17 with the warm equatorial air rising upwards to the top layers in Fig. 17a and the El Nio signal in Fig. 17b directly connects to the stratosphere. The warming in the lower atmosphere is much more dominant in EOF2 than in EOF1 so the separation between the troposphere and stratosphere is not as apparent. The Hadley cells which have not been seen in EOF1, are in EOF2 (see Fig. 17a). Likewise, the El Nio signal is much stronger in EOF2 (see Fig. 17b) than in EOF1 (see Fig. 14b). PC2 (see Fig. 11) has a similar upward trend to PC1, with a three to four year oscillation cycle. This oscillation pattern is the result of ENSO which explains the larger amplitude of the oscillations as ENSO varies the atmosphere more greatly.

5 Conclusions and Discussions

This paper has demonstrated the differences in heat interactions between an air-ocean coupled model and an air-only model through the computation of EOFs in 3D space. The data is first arranged into a space-time matrix as shown in section 2. The EOFs have been computed via temporal covariance which has been explained in further detail in section 3. Computing covariance in 3D space allowed the quantification of interactions of data points between different layers rather than within the same layer. This allowed the vertical cross sections to see spatial patterns in Figs. 6, 9, 13, and 16. The EOFs are organized in the order of importance based on the relative size of their associated eigenvalues. This determines the percentage variance each mode accounts for in the atmospheric model. An attempt to avoid the modal mixing was to use rotational EOFs, with the varimax rotation which commonly results in more physically interpretable patterns Shea (2013), but the results were similar and did not resolve the issue of modal mixing.

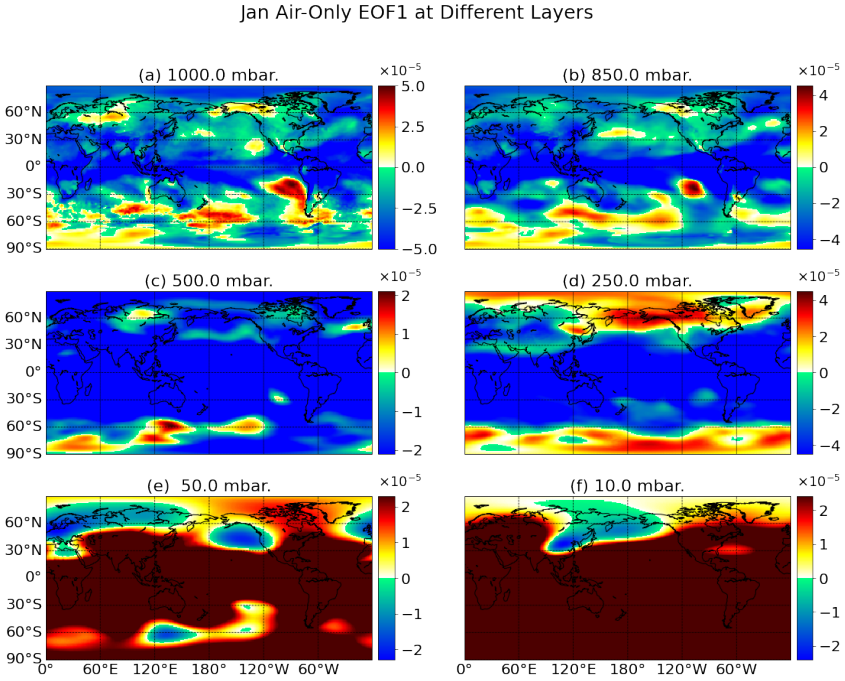


Fig. 13: The following layers are all part of the same 3D EOF1

The contrasting results between the air-ocean coupled model and the 20CRV only model have been illustrated in section 4. In the air-ocean coupled model, certain climate phenomena such as ENSO and Hadley cells are much more prominent due to the contribution of water. Due to modal mixing, ENSO accounted for approximately 30 % of the variance with the air-ocean coupled model whereas it only accounted for approximately 11 % of the variance in the air-only model. Likewise, Hadley cells were present in the first mode of the air-ocean coupled model, but not the air-only model. This suggests that water has a significant role in the formation of Hadley cells and the El Nio signal in the eastern Pacific. Although the computations have been done for all months, the results primarily focused on January because of ENSO and the northern polar anomaly.

The following results have been visualized in both 2-dimensional cross sections and 3-dimensional interactive figures which may be found at https://doi.org/10.5281/zenodo.7073022 in addition to other visualization methods such as a GIF animation.

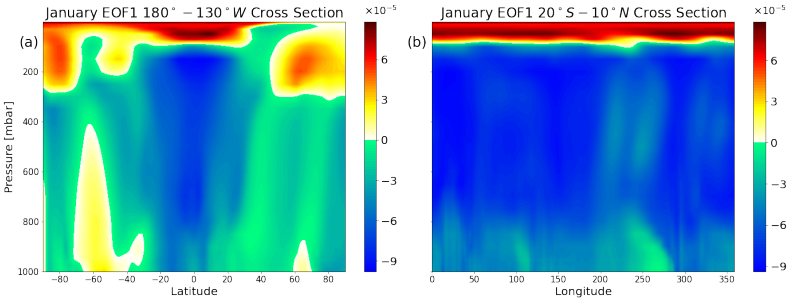


Fig. 14: Vertical Cross Sections of January Air-Only EOF1

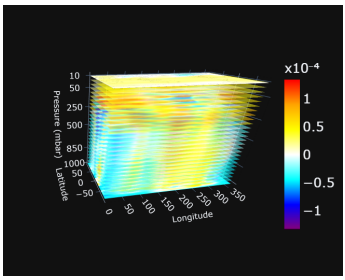


Fig. 15: January Air Only EOF2

Acknowledgments. This study is supported and monitored by The National Oceanic and Atmospheric Administration Cooperative Science Center for Earth System Sciences and Remote Sensing Technologies under the Cooperative Agreement Grant #: NA16SEC4810008.

The authors would like to thank NOAA Educational Partnership Program with Minority Serving Institutions for fellowship support for Thomas Bui and Danielle Lafarga.

Support for the Twentieth Century Reanalysis Project version 3 dataset is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Earth System Research Laboratory Physical Sciences Laboratory.

The statements contained within the manuscript/research article/poster are not the opinions of the funding agency or the U.S. government, but reflect the authors' opinions

Statements & Declarations

Funding. Thomas Bui and Danielle Lafarga were supported by the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Educational Partnership Program under Agreement No. NA16SEC4810006.

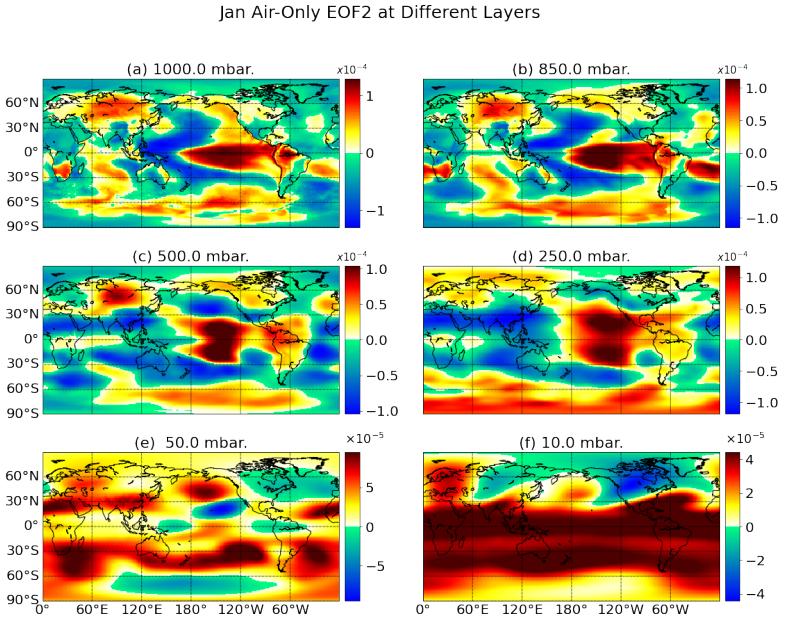


Fig. 16: The following layers are part of the same EOF

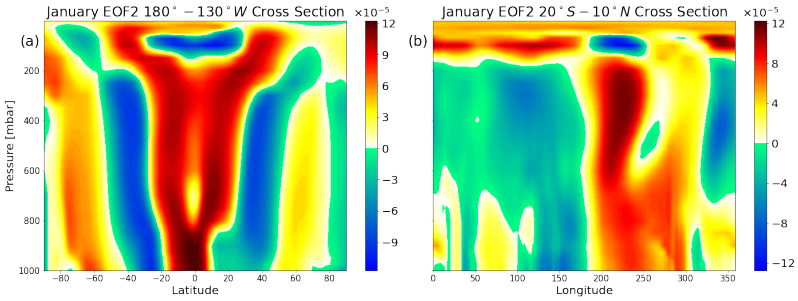


Fig. 17: Vertical Cross Section of Air Only EOF2

This publication contents are solely the responsibility of the award recipients and do not necessarily represent the official views of the U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval. Not applicable.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Availability of data. The original data used in this study are freely available. In particular, the NCEP Global Ocean Data Assimilation System (GODAS) data are provided by the NOAA PSL, Boulder, Colorado, USA, from their website
https://psl.noaa.gov/data/gridded/data.godas.html
The NOAA/CIRES/DOE 20th Century Reanalysis (V3) data are provided by the NOAA PSL, Boulder, Colorado, USA, from their website
https://psl.noaa.gov/data/gridded/data.20thCReanV3.html

Code availability. Both the computer code and its results are freely available in the GitHub repository: https://doi.org/10.5281/zenodo.7073022

Author Contributions. All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Thomas Bui, Danielle Lafarga and Samuel Shen. The first draft of the manuscript was written by Thomas Bui. All authors participated in the revision of each version of the manuscript. All authors read and approved the final manuscript.

References

Behringer MJDavid W., Leetmaa A (1998) An improved coupled model for enso prediction and implications for ocean initialization. part i: The ocean data assimilation system. Monthly Weather Review 126(4):1013 – 1021

Bell R, Hodges K, Vidale PL, et al (2014) Simulation of the global enso–tropical cyclone teleconnection by a high-resolution coupled general circulation model. Journal of Climate 27(17):6404–6422

Jadhav J, Panickal S, Marathe S, et al (2015) On the possible cause of distinct el niño types in the recent decades. Scientific reports 5(1):1–13

Johanson CM, Fu Q (2009) Hadley cell widening: Model simulations versus observations. Journal of Climate 22(10):2713–2725

Liang S, Li X, Wang J (2012) Chapter 22 - high-level land product integration methods. In: Liang S, Li X, Wang J (eds) Advanced Remote Sensing. Academic Press, Boston, p 667–690, https://doi.org/https://doi.org/10.1016/B978-0-12-385954-9.00022-8, URL https://www.sciencedirect.com/science/article/pii/B9780123859549000228

Matthewman NJ, Esler JG, Charlton-Perez AJ, et al (2009) A new look at stratospheric sudden warmings. part iii: Polar vortex evolution and vertical structure. Journal of Climate 22(6):1566–1585. https://doi.org/10.1175/2008JCLI2365.1

Shea D (2013) Analysis tools and methods. URL https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/empirical-orthogonal-function-eof-analysis-and-rotated-eof-analysis

Shen SSP, Somerville RCJ (2019) Climate Mathematics: Theory and Applications. Cambridge University Press

Shen SSP, Behm GP, Song YT, et al (2017) A dynamically consistent reconstruction of ocean temperature. Journal of Atmospheric and Oceanic Technology 34(5):1061 – 1082. https://doi.org/10.1175/JTECH-D-16-0133.1, URL https://journals.ametsoc.org/view/journals/atot/34/5/jtech-d-16-0133.1.xml

Slivinski LC, coauthors (2019) Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. Quarterly Journal of the Royal Meteorological Society 145(724):2876–2908

Zhang Z, Moore JC (2015) Chapter 6 - empirical orthogonal functions. In: Zhang Z, Moore JC (eds) Mathematical and Physical Fundamentals of Climate Change. Elsevier, Boston, p 161–197, https://doi.org/https://doi.org/10.1016/B978-0-12-800066-3.00006-1, URL https://www.sciencedirect.com/science/article/pii/B9780128000663000061