RECONSTRUCTION AND ANALYSIS OF OCEAN BIOGEOCHEMICAL AND CLIMATE VARIABILITY AT OCEAN STATION PAPA

By

Erin McKee

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2020

@Copyright by Erin McKee, 2020

Dedication Page

I dedicate this work to all of my family and friends who've helped me through this herculean task. Many thanks to everyone who has helped me with this project, including my supervisor, Mike, my committee members, and the Dal Oceanography office staff. From sleepless nights due to frustration over code or my cat acting up, to staring at figure after figure while my eyes were starting to blur there were always people I could count on to give me a break, to get me out of that head space; yes, that includes you, Sean.

To meeting new people and making lasting memories. Thank you, everyone.

Table of Contents

List of Tables

List of Figures

Abstract

Few ocean stations have been in place long enough to observe decadal ocean cycles, but Ocean Station Papa (OSP) is one of those few. With a time series that spans over 60 years, OSP is an ideal data set for linking decadal and multi-decadal changes in the climate and ocean cycles. The purpose of this study is to determine if long term (decadal) climate variations significantly relate to and potentially impact biogeochemical variables in the open ocean north Pacific surface waters. One outstanding problem is that the OSP time series has highly variable sampling effort over time for different ocean variables. Such time gaps and irregular sampling make it difficult to do standard statistical time series analyses. Hence, the first part of this thesis is a statistical reconstruction of the original OSP data onto a regular monthly time grid. This is done using a state space model and the Kalman smoother algorithm. Its central idea is to estimate missing observations in seven ocean variable time series (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen) by using empirical relationships between the variables, as well as making use of the fact that some of these variables (e.g. temperature and salinity) are available for the entire analysis period. Specifically, a period of high-density sampling is first used to establish the relationship between the variables, which is then used to reconstruct the seven ocean variable time series with the Kalman smoother algorithm. The second part of this thesis aims to relate the reconstructed OSP variables to modes of climate variability. The reconstructed OSP time series are first smoothed to remove seasonal variations. They are then compared to four climate modes (Pacific Decadal Oscillation, North Pacific Gyre Oscillation, Southern Oscillation Index, and Multivariate ENSO Index) using cross-correlation and cross-spectral analyses. The cross-correlations between the ocean state variables and climate modes show that NPGO has the greatest number of significant correlations as the leading variable. The cross-spectral analyses show that PDO has the least amount of influence on seven ocean state variables, and that NPGO is the climate mode with the most influence on the ocean state variables. Using this method, the influence of climate variability on physical, biogeochemical, and biological ocean variables could potentially be used on any ocean time series.

List of Abbreviations and Symbols Used

PDO – Pacific Decadal Oscillation

ENSO – El Niño/Southern Oscillation

NPGO – North Pacific Gyre Oscillation

OSP – Ocean Station Papa

DFO – Fisheries and Oceans Canada

NOAA – National Oceanic and Atmospheric Administration

PMEL – Pacific Marine Environmental Laboratory

SST – Sea Surface Temperature

SSS – Sea Surface Salinity

HNLC – High Nutrient Low Chlorophyll

HOT – Hawaii Ocean Time-series

BATS – Bermuda Ocean Time-series

SOI – Southern Oscillation Index

MEI – Multivariate ENSO Index

MICE – Multiple Imputation by Chained Equations

MV – Multi-Variate

STD – Standard Deviation

SSTa – Sea Surface Temperature Anomalies

SSHa – Sea Surface Height Anomalies

NEP – North Eastern Pacific

NPO – North Pacific Oscillation

CI – Confidence Interval

Φ – Dynamics matrix

V_t – Observation error

W_t – Model error

A_t – Measurement operator

Y_t – Observation at time t

X_t – Ocean state at time t

ϕ – Phase value at frequency ω

ω – Frequency

Chapter 1 – Introduction

1.1: Introduction and Overview

It is known that the climate-scale variability (referred to as climate from here on in) has an influence on the physical part of the ocean (as well as the ocean having an influence on climate), but the effects climate has on the biogeochemical variables in the ocean surface are less well characterized and understood. It is possible to observe this influence, in theory, with a data set that spans decades so the major modes of climate variability, such as the Pacific Decadal Oscillation (PDO) and the El Niño Southern Oscillation (ENSO) can be observed, but few such data sets exist in the ocean. Climate modes might have a direct influence on biogeochemical variables through changes in temperature or precipitation amount affecting species composition and/or nutrient uptake rates, or they might have an indirect influence by ways of shifting water currents or wind patterns. Using a long-term data set, this project aims to determine if there are any links between the regional climate modes and biogeochemical variables in the area. The data set that makes this possible exists at Ocean Station Papa (OSP) (50°N, 145°W) (Figure 1.1), where there has been data collection in the surface water from 1950 to the present (Freeland, 2007). The OSP data set is described in more detail later in this section.

This study plans to link and analyse changes in the surface oceanography with climatic modes using the OSP data. These insights could provide information on how the biogeochemical variables, and the base of the oceanic food web (phytoplankton), are affected by climate modes in the Pacific Ocean. Specifically, this thesis will develop approaches to make optimal use of the long term OSP time series, and to investigate links

between biogeochemical variables and climate variability. It is comprised of two parts; the first consists of the statistical reconstruction of a complete OSP biogeochemical time series, and the second consists of investigating any links between the reconstructed OSP time series and the selected climate variables.

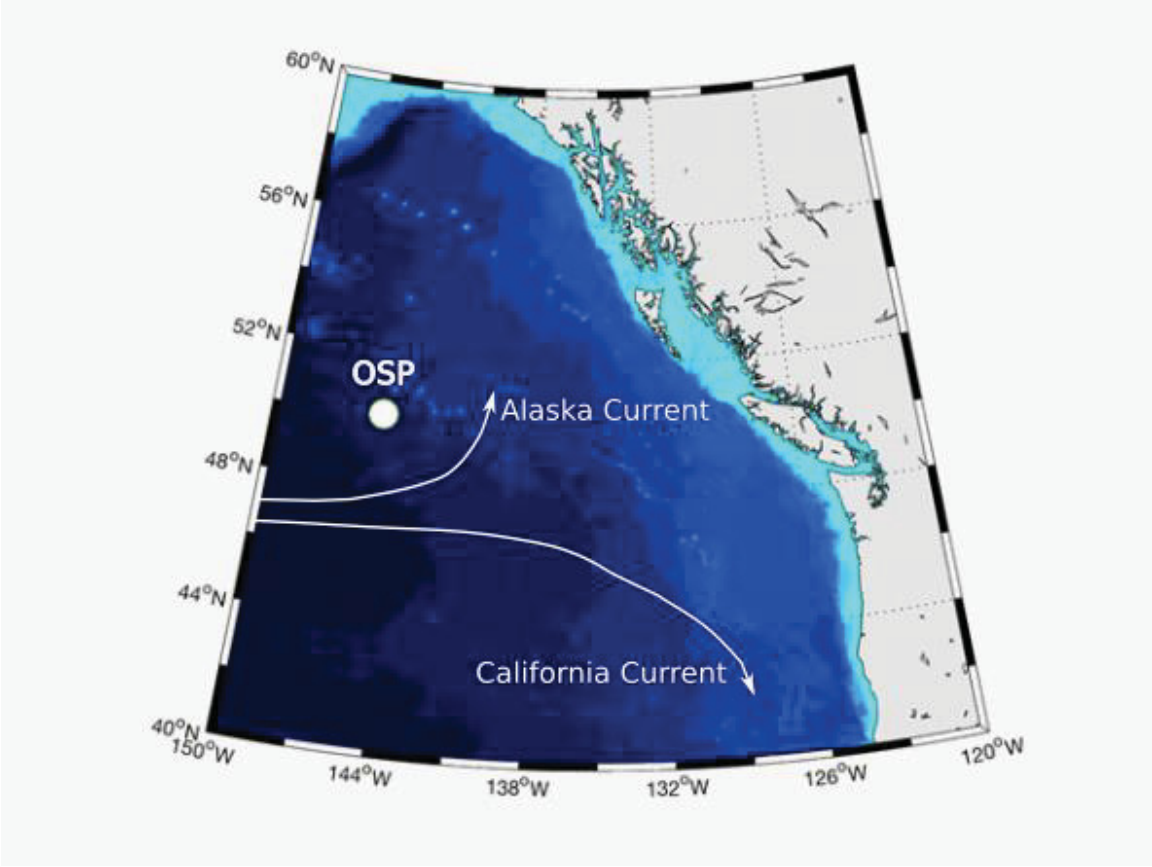

Figure 1.1: Map of Eastern Sub-Arctic Pacific ocean with Ocean Station Papa (OSP) indicated by the white dot at 50°N, 145°W. Colour scale indicates bathymetry where lighter blues indicate shallower waters than the darker blues (grey is land).

Ocean Station Papa was originally a weather station and had a ship stationed there from 1943 to 1981, where it not only took atmospheric measurements for weather prediction, but also collected a variety of oceanic data (Freeland, 2007), with systematic oceanographic observations starting in 1956. Initially the weather ships at OSP were

provided by the United States Navy, but this did not last long. Canadian occupation of the station started in 1950 (Garrett, 2006). The weather ship occupation of OSP ended in 1981 due to financial reasons. From this point on, shipboard observations at OSP by the Fisheries and Oceans Canada (DFO) involved sampling three to six times per year (but with the addition of many new variables (Freeland, 2007), marking a new era in the monitoring program). In 2007 another new era began when a National Oceanic and Atmospheric Administration (NOAA) surface mooring with the ability to make high-frequency observations (hourly, daily, weekly, etc.) was installed at OSP, primarily to study the Northeast Pacific carbon cycle (Fassbender et al., 2016). In 2013, this mooring was supplemented by two others as OSP became one of the Ocean Observatories Initiative Global Sites. To this day, the DFO continues to maintain shipboard measurements and provide ship time for mooring maintenance and OSP is considered a long-term monitoring site.

Ocean Station Papa is an important oceanic monitoring site, as it is one of the longest open ocean time series (Freeland, 2007). Long data records are essential for investigating any link between biogeochemical ocean surface variables and climate oscillations. Climatic events such as ENSO occur on a 2-7 year time scale (Santoso et al., 2017), so there should be many data sets that are capable of recording the effects they have on the surface ocean waters, but events such as PDO, North Pacific Gyre Oscillation (NPGO), and 'the blob', which occur on longer scales (decadal to centennial), can only have their cycles recorded by longer data sets. OSP is one of a few data sets that has the potential to provide an insight into the effects of climate forcing on ocean surface waters that many other oceanic data sets cannot, which addresses a question that is central to this

thesis. Note also that it is also possible for these long time series to be recorded using proxies, such as alkenones in the sediment records, though this technique is typically used for paleo-oceanography (Gould et al., 2019).

The overarching goal of this project is to link the changes in the selected biogeochemical variables in the surface waters at OSP with climate events using a statistical analysis of climate indices and reconstructed OSP ocean variables). There are a few other data sources available for this project besides the core data in the DFO archive: (i) the Pacific Marine Environmental Laboratory (PMEL), (ii) OSP data as compiled by Frank Whitney, and (iii) ocean reanalysis data products for physical variables. However, here are a few issues that need to be resolved before work can begin on the overarching goal. Firstly, the most ubiquitous feature which will be dealt with by statistical reconstruction (using the Kalman smoother algorithm) is that the OSP data time series is not continuous in the sense that it is not sampled at regular time intervals. There are times where hardly any samples were taken, and the measurement density varies greatly amongst recorded variables over 10 to 20-year time spans. When samples were taken consistently, they were often only taken two to three times a year, which provides the bare minimum data to see annual cycles. These changes in sampling protocol make it incredibly hard to do anything other than basic statistics (e.g. means, trends, histograms, characterising the average seasonal cycle) with this time series and had severely limited exploration of this rich data set. This has not, however, discouraged scientists from conducting studies on long-term trends at OSP. For example, Tabata (1989), Thomson & Tabata (1989), Freeland et al. (1997), and Whitney et al. (2007) have all looked at long-term trends at OSP using 27 to 39 years worth of data. In order to look at seasonal,

annual, or decadal cycles and link them to climate variability a time series with samples at regular intervals is needed for the more complex and informative statistical time series analyses that cannot be done when there are large gaps in the time series. Another related issue that can be solved with statistical reconstruction is the infilling or imputation of poorly sampled the ocean variables; some of the variables only have a few measurements throughout the entire 60-year time series, and others have been sampled at varying depths multiple times a year. For example, sea surface temperature (SST) and sea surface salinity (SSS) have been well sampled throughout the 60-year period, while chlorophyll and phosphate have been more sparsely sampled throughout the same period. Figure 1.2 shows an example of a well-sampled variable (SST). Figure 1.2a shows that while temperature has been measured to a depth of at least 500m from 1956 to 2017, there is a greater frequency of sampling in the shallower waters. Figure 1.2b depicts the number of times samples were taken each year and shows that the greatest temperature sampling frequency occurred from 1970-1981, where there were more than 500 samples taken per year. Figure 1.2c depicts the number of samples that were taken at each depth and shows that the top 5m of ocean is the most well-sampled depth, and that the top 100m is more frequently sampled compared to deeper water depths. Figure 1.2d is a time series of the top 15m, which clearly demonstrates that there is more sampling in the surface waters than the deeper waters.

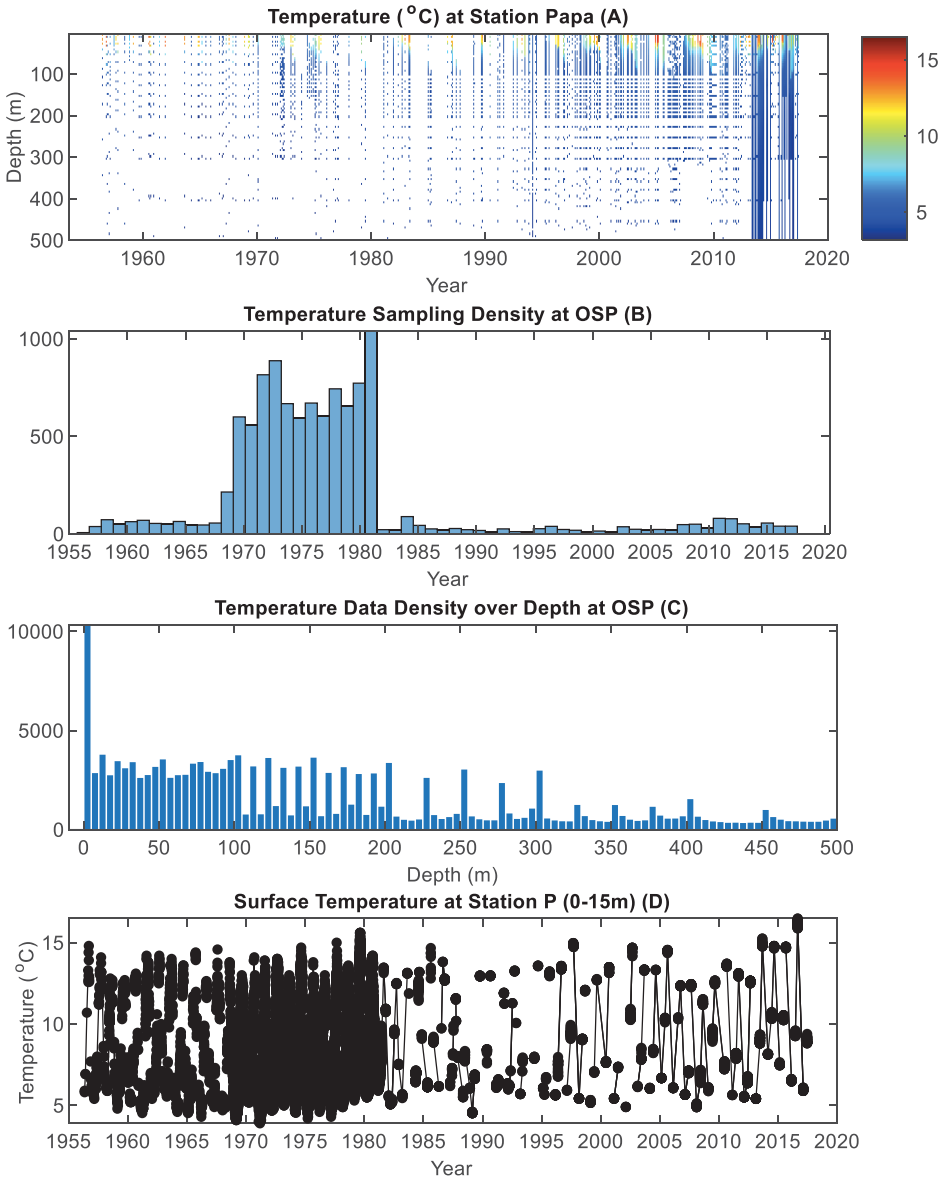

Figure 1.2:The temperature record at OSP. Panel A: The time and depth series of temperatures, with colour indicating temperature values. Panel B:The number of times per year temperature samples were taken. Panel C: The number of temperature samples taken at each depth throughout the entire time series. Panel D: Time series of mean temperature in the upper 15m.

Of the 39 variables that have been measured at OSP, the biogeochemical variables that were chosen for this project are: temperature (°C), salinity (PSU), nitrate (μmol/L) phosphate (μmol/L), silicate (μmol/L), chlorophyll (μmol/kg), and oxygen (mg/m^3). They were chosen due to their availability, coverage during the 60-year time period, and/or relevance to the goals of this study (i.e. linking surface biogeochemistry to climate). While samples have been taken from a variety of depths, the surface waters have the greatest density of samples (Figure 1.2) and so the average in the top three depth bins or 'sample buckets' (top 15m) was used. These represent surface time series for each of the seven variables. The climate modes that were chosen (PDO, ENSO, and NPGO) were based on known climate modes that originate and/or affect the sub-Arctic region where OSP is located. Once the biogeochemical variables (ocean state variables) have been reconstructed to be on a monthly time interval, a series of correlations, cross-correlations, and cross-spectral analyses will be completed to determine if and how the chosen climate modes impact the ocean state variables.

As noted, it is important for the time series to be complete (i.e. on a regular time grid, or values at every time step), since standard time series analyses methods generally require regular time intervals. There are a few advanced and specialised time series techniques that allow for irregularly sampled data, but we do not consider these here. There are also a number of imputation methods that are available to infill (gap fill) incomplete time series, ranging from mean imputation (calculating the mean value from the observations and using that value to fill in the gaps) and simple linear interpolation (which fits a straight line over data gaps) to more complex imputation methods, such as regression imputation which fills in the missing data values by regressing the missing

variable on other variables. For this project, the issue of having incomplete time series will be dealt with using a form of regression imputation using the Kalman smoother (which is a solution to a state space model that utilizes the Kalman filter, explained further in Section 2.1). This novel application of the Kalman smoother will be employed to reconstruct a complete time series for the full multivariate ocean time series with appropriately estimated error bars. The Kalman smoother can also make a prediction as to the state of the system for a period with no data points for that variable by borrowing information from other variables, a feature which is needed for this study. This means that it provides for multivariate infilling and imputation, unlike most methods that only deal with a single time series. Therefore, the Kalman smoother will be used to provide estimated values for all the selected ocean variables at each time step (monthly intervals in this project, ranging from 1956 to 2017), which will result in the complete, reconstructed time series with error bars needed for the second part of this thesis, where reconstructed ocean variables are statistically related to climate modes.

This thesis is structured as two major parts. The first part (Chapter 2) will consist of reconstructing the biogeochemical variables from OSP using the Kalman smoother. Then, the second part (Chapter 3) will examine the link between the biogeochemical variables at OSP and the climate modes to determine if climate variability has a significant influence on the surface biogeochemical variables at OSP.

1.2: Brief Overview of OSP Oceanography

Located in the sub-Arctic Pacific Ocean, OSP is located in a High Nitrogen, Low Chlorophyll (HNLC) region, which means there is some limiting nutrient/biological need that is preventing the local phytoplankton from using all of the available nutrients (Whitney et al. 2005). In fact, there have been studies done on whether fertilising HNLC regions with iron (the most common known limiting nutrient) can mitigate ocean acidification (Maldonado & Price, 1999), and on whether iron really is the limiting factor in these regions or if a vitamin (B_{12}) is the cause (Koch et al. 2011). There have also been many studies on whether increasing the productivity of these regions would result in an increase in the amount of atmospheric carbon dioxide being drawn into the ocean (Aumont & Bopp, 2006, Salter, et al., 2012).

The surface waters in the sub-Arctic Pacific are highly stratified by a steep halocline, and are often isolated from the nutrient rich deep waters (Haug, et al., 1999), but there is evidence that the northern Pacific Ocean is becoming less saline (Freeland, 2013), which may cause the halocline to weaken.

There have been many oceanographic studies in the OSP region. These include a study about 'The Blob'; a massive "blob" of warmer-than-usual water off the Pacific coast that has persisted since fall 2013 (Bond et al., 2015, Cavole et al., 2016, Whitney, 2015)), as well as a study that links ocean climate and ecosystem change with the North Pacific Gyre Oscillation (and initially identifies the NPGO) (Di Lorenzo et al., 2008). Other processes of interest in the region include the permanent stratification (Haug et al., 1999), the utilization of iron by plankton communities in the sub-Arctic Pacific Ocean

(Maldonado & Price, 1999), and whether iron fertilization can mitigate ocean acidification (Cao & Caldeira, 2010).

There are also many studies done directly at the OSP sampling site, not just in the surrounding region. C.S. There have been many studies over the past decades (Wong et al., 1999, 2010; Wong, 1978) looking at carbon dynamics at OSP (carbon fluxes, isotopic composition, relationships between phytoplankton, nutrients, and carbon, etc.). Another study of the biology at OSP was conducted by Frost (1987), who investigated the annual cycle of phytoplankton and zooplankton in the ecosystem. Physical studies include the annual cycle and depth penetration of waves (Alford et al., 2012), eddy circulation and ocean metabolism (Pelland, 2016), estimating secular changes in steric sea level (Thomson and Tabata, 2009), quantifying how surface waves drive upper ocean turbulence (D'Asaro et al., 2013), how the winter mixed layer in the Northeast Pacific has changed (Freeland et al., 1997), and many others. These studies, however, don't use the full suite of available data from OSP; the studies typically only cover a time span of about a month to a couple of years at most. There are some studies that do cover a greater time span, such as the evaluation the drivers of marine carbon cycling at OSP which used 7 years' worth of data (Fassbender et al., 2016), as well as some other long term studies (Di Lorenzo et al., 2009, Crawford & Peña, 2016 , and Freeland, 2013). These studies have their own methods of completing their times series, such as using a regional ocean model hindcast, locating all (published and unpublished) data available from previous research projects, or focusing on anomalies, spatial maps, and trends. Another way of dealing with inconsistent sampling at OSP is to exclude data from periods with sporadic sampling, like Peña & Varela (2007) did. When looking at decade's worth of data from OSP, some

researchers take data from the high density sampling, weather ship period and compute the long-term average annual cycle. The resulting averages are then interpolated to determine the baseline monthly average (Peña & Varela, 2007, Whitney et al., 2007, Wong et al., 2007). At other long-term open ocean time series stations (Hawaii Ocean Time-Series (HOT), Bermuda Atlantic Time-Series (BATS)), which have both been in operation since 1988, researchers typically use linear interpolation to fill in any gaps in the time series (Bingham & Lukas, 1996, Malmstsrom et al., 2010, Krause et al., 2009). Linear interpolation is a suitable method of imputation for these time series as they take regular, monthly (or better) samples. This is one of the issues that this project hopes to address – making full use of this valuable long-term, multi-variable data set. To do so, we aim to reconstruct and analyse the full time series with as many relevant biogeochemical variables as possible and make these long-term data are available for more complex studies.

1.3: Climate Modes

With climate change becoming a more accepted concept (especially for non-scientists) these days and being the cause for many protests or even being a deciding factor for many in federal elections around the world, there seems to still be a general misunderstanding of what is climate change and what is climate variability. There has been a semantic change to the term 'climate change' since data were first being collected at OSP. In the 1960s, the term climate change (which now often implies long term changes in the climate due to human activity) was used for what would be described today as climate variability (inconsistencies and anomalies in the climate, some natural in origin and some anthropogenically driven) (Rohli & Vega, 2018). Climate modes are

specific recurring patterns of climate variability. One of the climate modes often referenced in the popular media is the El Niño/Southern Oscillation (ENSO). An El Niño or La Nina can alter temperature and precipitation patterns enough to affect food production (agriculture and fishing) in many countries, particularly those bordering the Pacific Ocean (Santoso et al., 2017). There are other climate modes that are not typically as well known to the public that have a longer cycle as compared to ENSO's 3-5 year cycle. These longer-term climatic patterns may be the reason why people will say that the local climate was quite different when they were young, or 'back in my day', and blame climate change for those differences.

That being said, just as with ENSO events, longer term climate modes, such as the Pacific Decadal Oscillation (PDO), also alter weather patterns around the world. Figure 1.3 depicts some of the biological and physical phenomenon in the ocean and their space and time scales, ranging from micro patches to biogeographical provinces. The dotted oval labelled 'F' ('El Niño' type events) most closely represents the spatial and temporal scales which are examined in this thesis. These changes in weather and ocean patterns can have effects that are harder to determine due to the longevity of the event, and the fact they occur together with other climate variations. For example, PDO was only officially recognised as a climate mode after looking at 70 years of catch history of the Pacific salmon (Mantua & Hare, 2001). The changes in the Pacific salmon catch history is just one of the reasons why understanding how these climate modes affect the biogeochemical variables (nutrients for the bottom of the oceanic food web) in the ocean surface is of interest to this study.

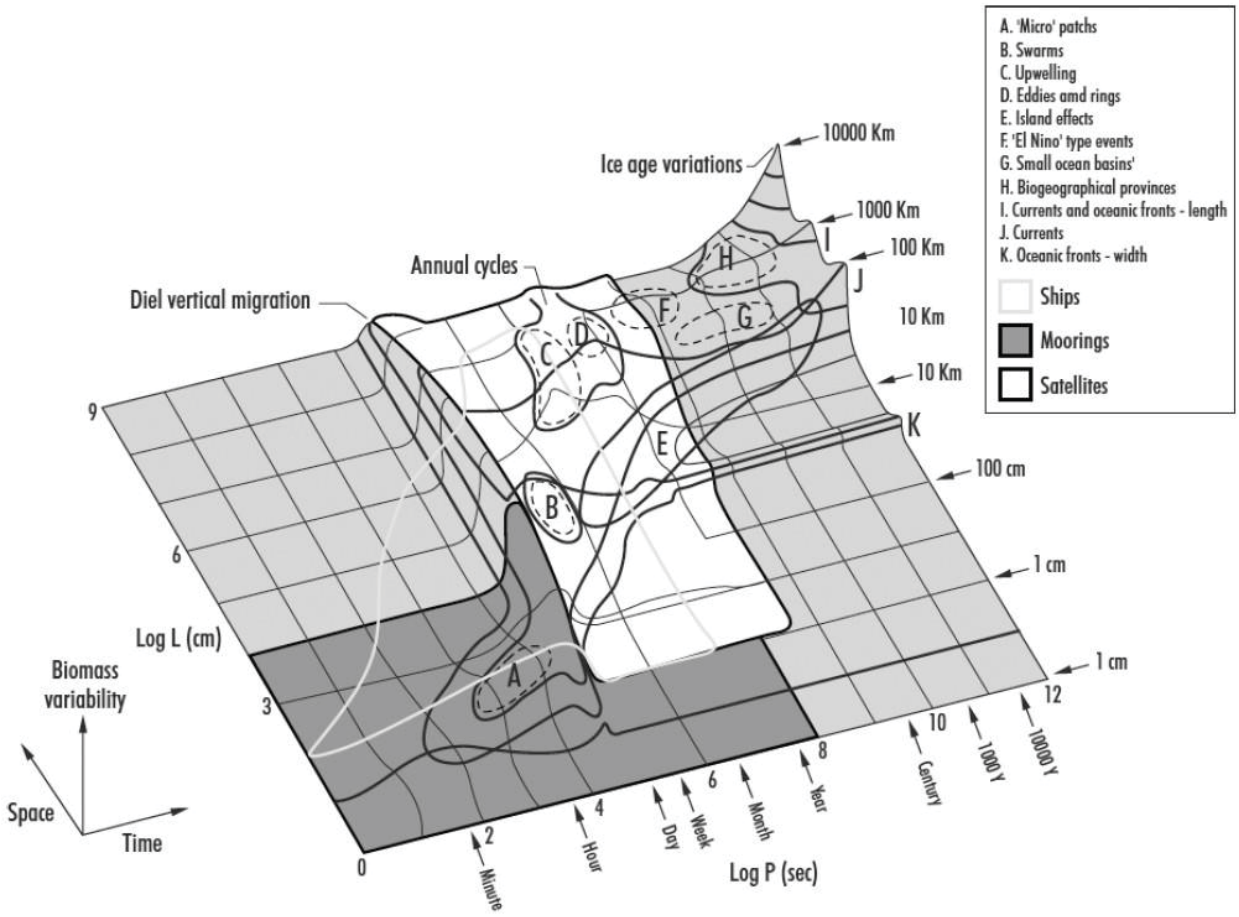

Figure 1.3: Stommel diagram from Kaiser et al., 2005 showing time, space, and biomass scales of biological and physical phenomneom within the ocean. The height of the vertical surface denotes the energy or power in the frequency bands. Shading denotes collection methods used. Area of interest for this study is the dotted circle "F'.

Due to the goals of our study (a complete monthly time series from 1956 to 2017) and the variable and inconsistent samplings rates at OSP, ocean reanalysis products will also be used in this study since these are available on the monthly basis for physical variables. Note that the relationship between the variables will still be calculated using the DFO OSP data, but the ocean reanalysis products for temperature and salinity at OSP will be used to provide a monthly foundation from which other time series will be reconstructed (see Section 2.3 for details). Based on the amalgamation of the information

provided by ocean models, atmospheric forcing fluxes, and ocean observations, ocean reanalyses are data assimilation products based historical reconstructions of the ocean systems (Balmesda et al., 2012). The ORAS4 reanalysis obtained from Advancing Reanalysis (Smith, 2010) was used in this study as it provides monthly sea surface (10m) temperature (Figure 1.4) and salinity data based on daily surface fluxes of heat, momentum, and fresh water. The seasonal cycle of ORAS4 temperature in Figure 1.4 is the dominant feature in the time series. There also appears to be an increase in SST in the last few years in Figure 1.4. There are data at every, monthly, time interval, whereas in Figure 1.2d there are gaps in the time series.

The reconstructed time series of seven ocean state variables (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen will be compared with the climate modes PDO, NPGO, SOI, and MEI (SOI and MEI are both ENSO indices) to determine how much effect these climate modes have on ocean variables mentioned above. Studying the interaction between climate variability and ocean variables is not a new concept, and there have been many studies done on it. A few findings include: the spatial pattern of the salinity signature of the PDO differed from the temperature signature (Overland et al., 1999); anomalous sperm whale sightings at OSP are correlated with ENSO events (Diogou et al., 2019); and studying oxygen and nutrient trends and their relation to PDO, NPGO, and the North Pacific Index (Stramma et al. 2020). The nitrate variable being used is actually a nitrate plus nitrite data set (just called nitrate for simplicity as nitrite is negligible compared to nitrate), but there are also separate data sets for nitrate and nitrite. Oxygen also has multiple, distinct, data sets in the OSP archive with different units (ml/L, µmol /L, µmol /kg). The oxygen data recorded in µmol/kg was

the one chosen for this study. Some of these other data sets will be used to verify the accuracy of the reconstruction made in this project.

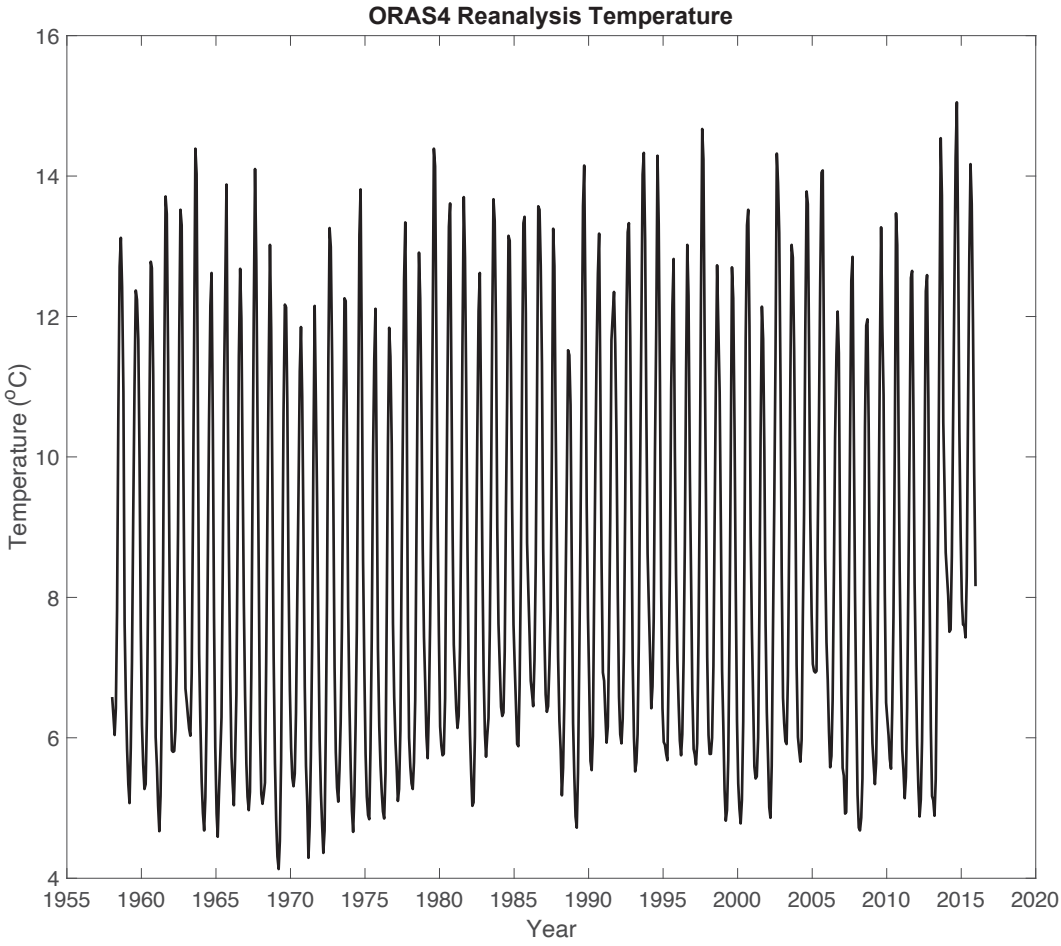

Figure 1.4: Temperature time series from ORAS4 reanalysis. Temperature is documented on a regular (monthly) timescale.

Chapter 2 – Statistical Reconstruction and Application

2.1: Statistical Reconstruction Method

In order to analyse the OSP data and connect it to any trends and variability in relevant climate data, the gaps or data voids in the time series need to be filled (imputated). The reason is that some of the statistical analyses that will be of interest (e.g. cross-spectral analysis) require the time series to have values at every time step for all variables, and that the different time series need to be the same length. Note that there are a number of ways to imputate time series data that have significant gaps in them, ranging from linear interpolation for simpler systems, to multiple imputation by chained equations (MICE) for more complex systems, and even reanalysis products based on numerical modelling (Storto & Masina, 2016), which can be seen as the gold standard. For this study, linear interpolation will not work as there are large gaps with durations greater than the important cycles in the data (seasonal and decadal), that cannot be resolved and will hence be mis-represented for these time periods. As for reanalysis based on numerical modelling and data assimilation, while it works for physical variables where the relationships are well known, it will not be nearly as accurate for biogeochemical variables where most of the relationships and influencing factors are not well understood. As well, there are a lot fewer biogeochemical data than physical data. From a practical perspective, re-analysis is a complex exercise requiring a numerical modelling and data assimilation system and requires more computational time and effort than is available for this study.

Therefore, we focus on statistical reconstruction method using the Kalman smoother. The Kalman smoother is a solution algorithm for a state space model which is

suitable for linear and Gaussian systems which contain statistical noise and error, and it is particularly suitable for determining features such as inter-annual variability and seasonal cycles. The output of the Kalman smoother produces estimates of the unknown variables over time (along with their errors), and hence allows for a complete gap-free time series to be reconstructed, as will be demonstrated below. To the author's knowledge, this feature of the state space model/Kalman smoother has not been used explicitly to carry out imputation.

The Kalman smoother (state space) equations are:

$$X_t = \Phi\, X_{t-1} + W_t \quad (1)$$

$$Y_t = A_t\, X_t + V_t \quad (2)$$

where:

- X_t : ocean state at time t (e.g. temperature, salinity, nitrate, etc.).
- Φ : dynamics matrix.
- W_t: model error.
- Y_t : observations at time t (key quantity, comprises the OSP observations).
- A_t : measurement operator.
- V_t : observations error.

The main assumptions are that W_t and V_t are both normally, and independently distributed with a 0 mean and variances of the system noise covariance and observation error covariance, respectively, and that they are independent of each other. To check the assumptions, we check the one-step ahead prediction errors (the innovations). They should be normal, independent, and identically distributed. The goal of the state space

model is to estimate the ocean state (X_t) monthly over the entire analysis period (1956-2017). To do this, the OSP observations (Y_t) are needed, and as well as specification of the dynamics matrix (Φ) and the measurement operator (A_t), as well as error statistics for the model and observations. In these equations, X_t describes the ocean state comprising temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen; Φ relates the ocean state variables from one time (month) to the next time and will be estimated externally using a training data set; W_t, which has the same units as X_t, describes the model error, and is an estimated input in the form of its variance (or standard deviation); Y_t comes from the observed OSP data; A_t is an input which indicates which observations are to be used in the analysis at each time step; and V_t, which has the same units at Y_t, describes the observation errors and is also an estimated input in terms of its variance.

It is with the state space model described by Eqs. 1 and 2 that the gaps within the OSP data set can be filled. While there are approaches that use the Kalman smoother when there are no variable observations at an analysis time, the availability of monthly temperature and salinity data negate the need to use these methods. Equation 1 predicts what the ocean state will be, and Eq. 2 determines which observations will be used at each time step, through A_t to refine the estimate of the ocean state. Given that Y_t, A_t, and V_t are known (or will be determined from the data), and Φ and W_t are also assumed to be known (or can be estimated, again from the data), the only unknown is X_t, which estimates the mean ocean state (the output) and its variance (the errors) at any given time. Hence, the output product of the Kalman smoother is a reconstructed time series with error bars. The procedure for this will be illustrated below. Details of the Kalman smoother algorithm can be found in Appendix A.1.

The dynamics matrix, Φ, is the matrix that describes the linear time evolution of the ocean system from one month to the next. Its size is determined by the number of variables that comprise the ocean state (here a 7 x 7 matrix). In this project it is derived by a multivariate regression (explained in Section 2.2) using a training data set with a time step of one month. A multivariate regression is applied to this training data to determine the dynamics matrix (see Section 2.2). Specifically, a six year section of the time series in the late 1970s with a high density of samples was used to estimate Φ (details in Section 2.2). We will assume this Φ is valid for the entire time period of interest (i.e. the relationships between the variables do not change over time). The values in the matrix indicate the relationship that each variable has with itself and every other variable for a monthly time step.

2.2: Example of Statistical Reconstruction Using Kalman Smoother

A simple application to some of the OSP surface data will be done in this section to provide a better understanding as to how the Kalman smoother is applied for time series reconstruction. For illustration, a simple three variable (temperature, salinity, and nitrate) example is used based on an OSP surface ocean data set provided by Frank Whitney (which is also partly included in DFO archive) from the years 1956 to 2014, as well as the physical reanalysis time series, ORAS4 (Balmaseda et al., 2012), which provides monthly SST and SSS values.

The following proof of concept experiment is carried out where the ocean state at OSP, here comprised of SST, SSS, and nitrate, is reconstructed on a monthly time interval. This is comprised of two main steps: (i) determining Φ (the dynamics matrix),

and (ii) applying the Kalman smoother to reconstruct the ocean state (gap filling or imputation). The dynamics matrix is first determined by using Frank Whitney's SST, SSS, and nitrate data. Second, using the estimated Φ, a reconstruction is undertaken with the Kalman smoother wherein the reanalysis SST and SSS (from ORAS4, regular monthly intervals) are used along with the gappy nitrate record (from Frank Whitney). In this manner, the SST, and SSS data are used to impute the values of nitrate. Therefore, the observations to be used in the Kalman smoother (Y_t) are the Frank Whitney nitrate data and the ORAS4 SST and SSS data with; Φ and A_t are to be determined. For simplicity, an assumption of a constant standard deviation of 1 for the W_t matrix and 0.5 for the V_t matrix is used (an approach for the proper estimation of the error standard deviations is taken up in Section 2.3). The reconstructed time series will be X_t over the full analysis period. The procedure is illustrated below.

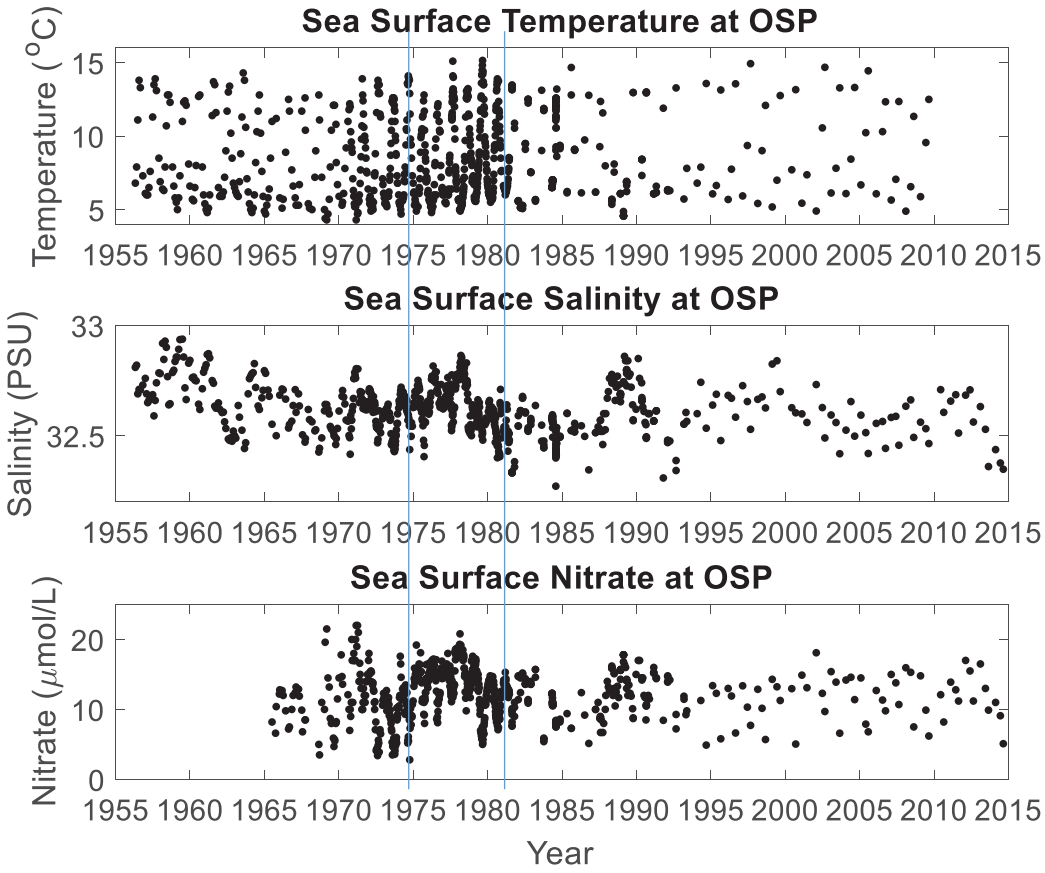

Figure 2.1: Scatter plot of Frank Whitney's raw surface temperature data collected at OSP. Black circles are observations. Panel 1 shows SST data at OSP. Panel 2 shows SSS data at OSP. Panel 3 shows Sea Surface Nitrate data at OSP. The data between the blue, vertical lines (1975 to 1981) are the training data used to determine Φ due to the density of observations for all three variables.

2.2.1: Determining Φ

Suppose one wanted to determine a Φ (the lagged relationship between the chosen variables) (Eq. 1) with a one-month time step using only a single variable, say, temperature. This would be done through the use of a lagged univariate regression. By taking two temperature values one month apart from each other, a linear regression could be used to estimate the relationship between temperature values lagged by one month. This would yield Φ which is in fact the auto-covariance of temperature at lag one. To

obtain an analogous result for multiple variables we perform a multivariate regression. The same concept is used, but now using two, or more, different variables, such as temperature, salinity, and nitrate. The purpose of the regression is to describe how one variable (the response) depends on another (the predictor), lagged by one month; this provides a Φ matrix. More information on this multivariate regression can be found in Appendix A.2.

To estimate the Φ matrix, a section of the time series that has all three variables sampled regularly on a monthly basis is used for training data. For this example, the data from the years 1975 to 1981 are being used (Figure 2.1) as they are from the period of highest density sampling. These data were re-scaled to have zero-mean and unit variance. Carrying out a lagged multivariate regression on these scaled data with a time step of one month yields Φ, a dynamics matrix. The numerical values obtained are:

$$\Phi = \begin{array}{c} \\ T \\ S \\ N \end{array} \begin{array}{c} \begin{array}{ccc} T & S & N \end{array} \\ \begin{bmatrix} 1.0009 & -0.0563 & 0.0565 \\ 0.0384 & 0.9658 & 0.0446 \\ -0.0942 & 0.0885 & 0.7692 \end{bmatrix} \end{array}$$

In this case it is a 3 x3 matrix where the numbers on the diagonal (1.009, 0.9658, and 0.7692) are the auto-regressive coefficients and are similar to the persistence of each state variable, but may be greater than 1 and amplify the previous value, and the off diagonal numbers are the coupling between the different state variables. In all, it indicates by how much each state variable will change according to the previous value of another variable. The coefficients indicate by how much the dependent variables change at the next time step when the independent variable increases by a unit of one. The auto-regressive

coefficients of the variables (diagonal elements) indicates how much influence one variable has on itself from one month to the next. Therefore, temperature amplifies itself from one month to the next, while nitrate (0.7692) does not influence itself as much as temperature. The couplings between the state variables indicate how much influence each variable has on each variable at the next time step; e.g. $T_t = 1.0009T_{t-1} - 0.0563S_{t-1} + 0.0565N_{t-1}$. Generally, the higher the absolute value, the greater the influence with the sign indicating the direction of influence.

2.2.2: Determining A_t

The measurement operator, A_t, also needs to be specified. This matrix determines which observations will be used in the Kalman smoother at each time. According to Eq. 2 it relates the state X_t to the observations Y_t. For the simplified example here, there are two possible cases:

Case 1: measurement for temperature, salinity, and nitrate are all available at time *t*.

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Case 2: measurement for only temperature and salinity are available at time *t*, with nitrate missing.

$$A_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

In order to avoid any times where there are no observed variables, the reanalysis product for sea surface temperature and sea surface salinity at OSP are used since the reanalyses have monthly values for the entire period covered in this project. Note that these temperature and salinity from reanalysis are typically highly consistent with the OSP data with r = 0.99 for the two temperature time series, and r = 0.95 for the salinity time series.

2.2.3: The Errors

The statistics of the model and observation errors need to be determined, specifically their variances or standard deviations. For the simple example shown above, the standard deviation of the model error has been set at 1.0, and the standard deviation observation error has been set at 0.5. Note that these values are completely arbitrary and apply to all the scaled ocean variables equally; further detailed explanations of how to obtain more reasonable error statistics will be provided in Section 2.3. Their interpretation is that when there is a high model error, the reconstruction does not follow the dynamics well and relies more heavily on the observations. Alternatively, when there is a low model error, the reconstruction relies more so on the model than the observations, so the reconstruction is largely based on Φ. Having a high observation error has similar results as having a low model error in that the reconstruction mostly ignores any observations and relies on the model. When there is a low observation error, the reconstruction follows the observations (when there are observations) rather than the model, even when there are potential outliers.

2.2.4: Kalman Smoother Application

Now that all of the input values needed to complete the model have been assigned, the Kalman smoother can be implemented. It is again emphasized that the analysis takes place using scaled, dimensionless variables, and then is converted back into the original units for presentation. Given the data set and the values assigned above, Figure 2.2 shows results of the Kalman smoother for the full analysis period from 1956 to 2017.

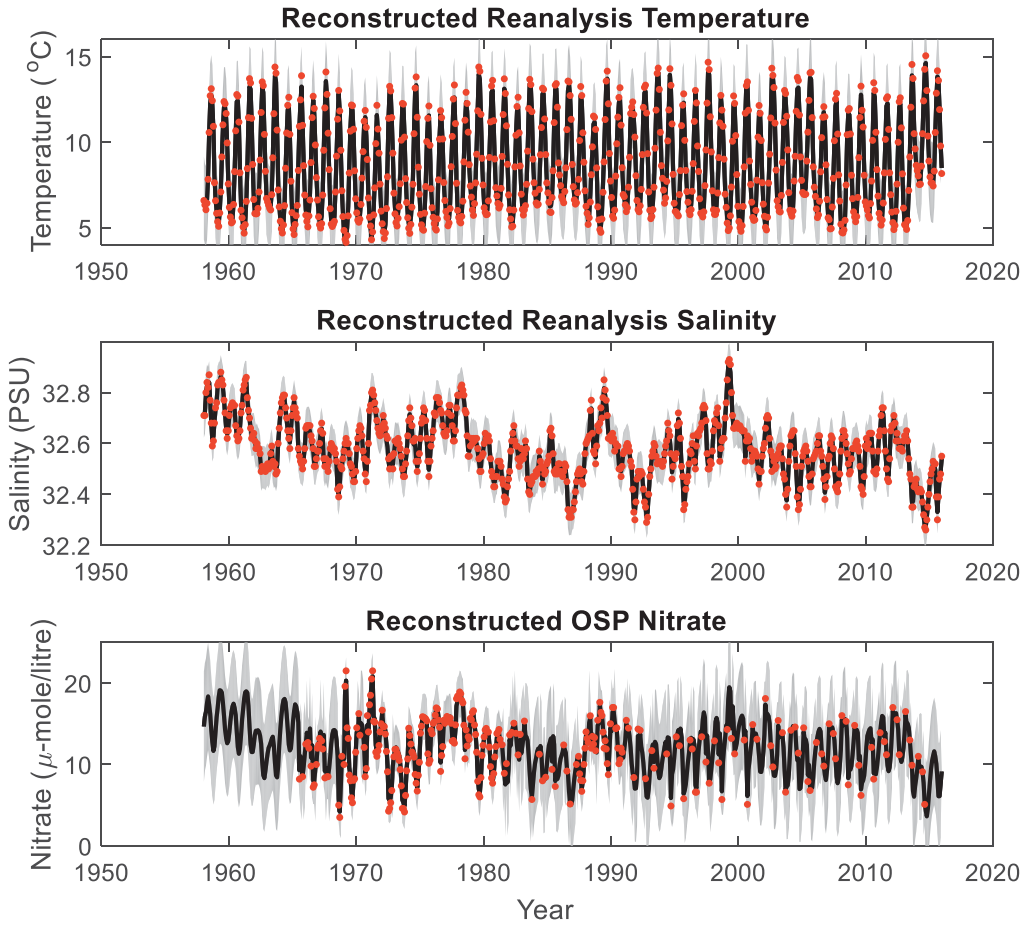

Figure 2.2: The reconstructed time series for temperature (first panel), salinity (second panel), and nitrate (third panel) from 1956 to 1997. The SST and SSS time series are based on the ORSA4 reanalysis data. The black line represents the mean for the reconstructed values, the grey shaded area represents the error bars based on the mean ± 2 x standard deviation. The red dots represent the observations used to create the reconstruction.

The reconstruction results show the same patterns as the original data, but the nitrate time series is now complete with error bars. Note that the error bars get smaller during the time steps where there are data available but are larger in nitrate data voids. As the errors are currently being assumed, the absolute size of the error bars themselves are not accurate but the results likely reflect the relative patterns.

The temperature and salinity reconstructions line up almost exactly with the observations but seeing as there are no gaps in the original reanalyses data, and the observation error is assumed relatively small, that is to be expected. Any discrepancies seen can be accounted by the fact there are small errors associated with both the model and the observations. When there are nitrate observations available, they typically lie within the error bars. The reconstruction, however, often relies more on the model than the observations for the majority of the time series, particularly at the beginning of the time series where there are no nitrate observations and estimates for nitrate are entirely based on the values of temperature and salinity and their relationship with nitrate as dictated by the dynamics matrix. The nitrate values show a dip around 1960, which corresponds with the dip in salinity values in the same time frame. It is difficult to determine how much influence temperature has due to the strong seasonal cycle.

2.3: Full Application of Kalman Smoother to the OSP Data

Now that a simple example has been presented for illustrative purposes, we now apply this method to reconstruct the seven time series (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, oxygen) from OSP. These time series (Figure 2.3) were provided by DFO and cover a maximum time period of May 1956 to June 2017. These

variables were chosen because they have at least two to three samples per year for the majority of the 60 year time series and seemed to be relevant to the goals of this project in that they are known nutrients for phytoplankton (nitrate, silicate, phosphate), proxies for determining phytoplankton mass (chlorophyll), and/or physical variables which can help fill in any gaps in the other variables. As can be seen in Figure 2.3, there are inconsistent sampling rates and some substantial gaps in these time series.

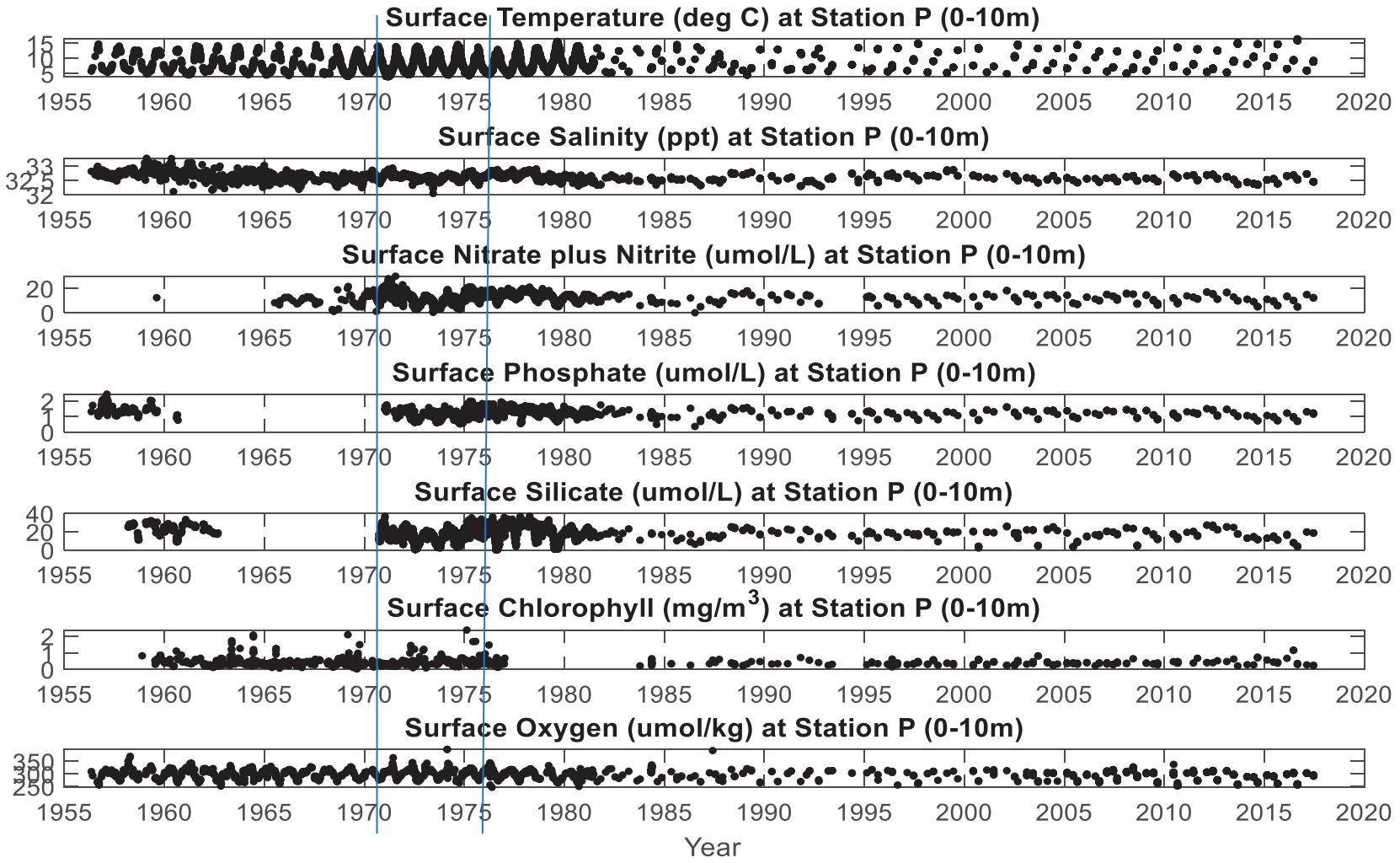

Figure 2.3: Time series of the OSP data averaged over the top 15 m of ocean surface. Black cicrles indicate data points. The data within the vertical blue lines (Sep. 1970 to Mar. 1967) denotes the training data used to determine Φ. Panels A through G represent: temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen, respectively

Temperature and salinity have similar sampling rates with the densest sampling being done pre-1980, and then dropping to a few times a year after that (note that we are not including any of the more recent mooring or Argo float data which would serve to greatly increase the data volume). While temperature has an obvious seasonal cycle it does not appear that salinity has one. Salinity does, however, have a decreasing trend which is well-documented for the sub-Arctic Pacific. Nitrate sampling only really started after 1965, and there are a couple of years after that where it does not appear to have been sampled. Phosphate had some dense sampling pre-1960, but then there are no samples for about a decade. This is similar to silicate, which was sampled from the late 1950s to early 1960 and then again after 1970. Oxygen was a fairly well sampled throughout the full time series. Chlorophyll was densely sampled from 1960 to the mid-1970s, and then started lower density sampling in the mid-1980s. Note that while the chlorophyll samples have units of mg/m^3 , all work with this chlorophyll time series will be done in log chlorophyll.

For carrying out the analysis within the Kalman smoother, we will be making use of scaled, or dimensionless, variables with zero means and unitless variances. The equations used to scale and un-scale each of the variables are:

$$x_s = \frac{x - \bar{x}}{s_x} \qquad (3)$$

$$x = s_x * x_s + \bar{x} \qquad (4)$$

where:

- x_s: the scaled state variable
- x : the original state variable
- $\bar{x}$: the sample mean of the state variable

- s_x: the standard deviation of the state variables.

These time series were scaled (Eq. 3) so the Kalman smoother does not have to estimate a mean, and for numerical efficiency and robustness in determining the dynamics matrix and the error analysis (after the Kalman smoother is run, the variables are un-scaled using Eq.4 to maintain original units for easier interpretation).

2.3.1: Determining Φ

To carry out the statistical reconstruction for the seven time series, the first thing to do, as in section 2.1, is to find a period of high frequency sampling so that a training data set and the dynamics matrix can be obtained. For this set of time series, the training period was set to be September 1970 to March 1976 (between the blue lines in Figure 2.3).

The time series within the training period need to be at a consistent sampling rate (monthly for this project) in order to use the multivariate regression used to determine Φ. For the training period, the raw data is sampled at a high sampling rate, daily to weekly, and not at a consistent monthly rate. Hence, basic linear interpolation was used to downscale the data to obtain a monthly sampling frequency for all of the variables.

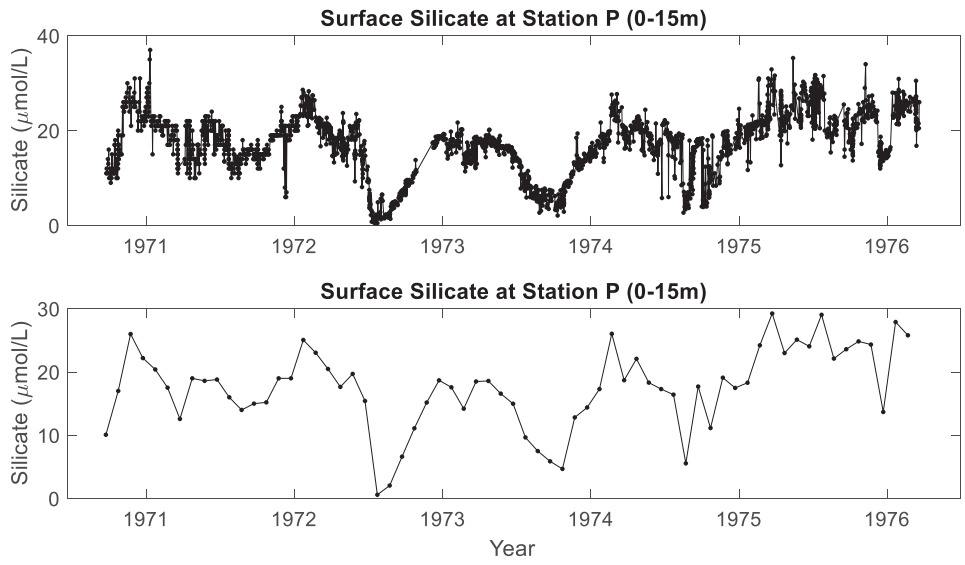

Figure 2.4: Silicate data from training period before (top panel) and after (bottom panel) an interpolation. Black dots denote data points.

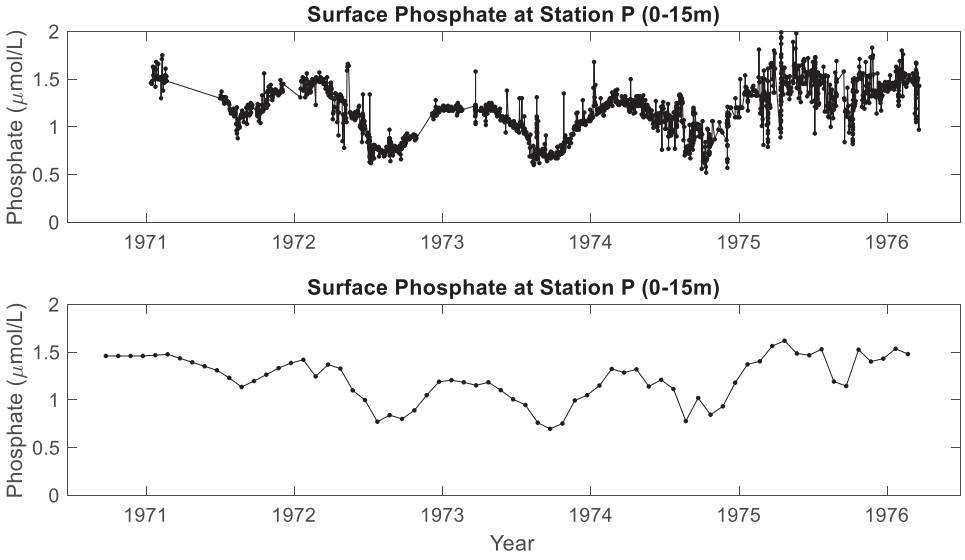

Figure 2.5: Phosphate data from training period before (top panel) and after (bottom panel) an interpolation. Black dot denote data points.

The top panels in Figures 2.4 and 2.5 depict the dense sampling from the raw OSP data, where, for the most part, there are more than one sample per month. The top panel in Figure 2.5 also shows a period in 1971 where there were no samples collected. The

bottom panels for Figure 2.4 and 2.5 show the result of linear interpolation, which puts the time series on a consistent, monthly sampling rate, and provides the correct input data to implement the multivariate regression. This interpolation was successful in maintaining the cycles and trends within the original time series. The multivariate regression is then carried out using the scaled version of the original units for all the variables with the exception for chlorophyll, as mentioned above.

Now that all of the variables have a consistent sampling rate for the training period and are scaled, the multivariate regression resulted in the following $7x7$ matrix for Φ:

$$\Phi = \begin{array}{c} \begin{matrix} \text{T} & \text{S} & \text{N} & \text{P} & \text{Si} & \text{C} & \text{O} \end{matrix} \\ \begin{bmatrix} 1.0669 & 0.0280 & -0.0283 & 0.0302 & -0.0139 & -0.0079 & -0.0812 \\ -0.1124 & 0.7208 & 0.1354 & 0.0585 & -0.0814 & -0.0112 & 0.0159 \\ -0.2249 & 0.1751 & 0.5355 & 0.1710 & -0.0127 & 0.0117 & -0.1409 \\ -0.1555 & -0.0198 & 0.1796 & 0.5663 & 0.1275 & -0.0241 & -0.0923 \\ -0.1785 & 0.0239 & -0.0897 & 0.1246 & 0.7584 & -0.0008 & -0.0837 \\ -0.0983 & -0.0685 & 0.0163 & -0.0731 & 0.0345 & 0.8012 & -0.0541 \\ -0.4695 & 0.1337 & -0.2060 & -0.0075 & 0.0661 & -0.0228 & 0.4874 \end{bmatrix} \end{array} \begin{matrix} T \\ S \\ N \\ P \\ Si \\ C \\ O \end{matrix}$$

As before, the the numbers on the diagonal are the auto-regressive coefficients of each state variable from one month to the next, and the off diagonal numbers are the coupling between the different state variables at a monthly time lag, where the couplings indicate how much influence each variable has on each other variable at the next time step (month). For example, based on this Φ matrix, at each time step, t, scaled salinity is calculated by

$$S_t = -0.1124(T_{t-1}) + 0.7208\,(S_{t-1}) + 0.1354(N_{t-1}) + 0.0585(P_{t-1})$$
$$-0.0814(Si_{t-1}) - 0.0112(C_{t-1}) + 0.0159(O_{t-1}),$$

where the ocean state variables are represented by their first letter (and second for silicate), a convention used throughout the thesis

Differences in the temperature, salinity, and nitrate persistence between this Φ matrix and the simpler one in Section 2.2 are likely due to differences in the raw OSP data used, and the different time period of the training data sets. Here, temperature has the highest persistence at 1.0669 and oxygen has the lowest at 0.4874. All of the temperature couplings are negative, potentially indicating that this variable has negative correlations with all of the other ocean state variables. Chlorophyll does not strongly influence any other variable (no Φ values greater than ±0.1), while temperature has a stronger influence (greater than ±0.1) on four other variables.

Note that the two values for each variable pairing (e.g. salinity/nitrate, nitrate/salinity) are not expected to be the same as one is the influence of N_{t-1} on S_t which would not the same as the influence of S_{t-1} on N_t. As can be seen in the Φ matrix above, this feature results in an asymmetric matrix.

2.3.2: Determining A_t

The set-up of the observation matrix A_t follows the same rationale as in the example of Section 2.1, but now there are seven state variables (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen) instead of three. The temperature and salinity used in the reconstruction are from monthly data from reanalysis and hence observed at all the monthly analysis times. However, the remaining five variables are not all measured at the same times, nor on a regular time grid. As a result, the observation matrix can be anywhere from a 2x7 (case 3) matrix to a 7x7 (case 4) matrix or anything

in between (case 5). That means there are 32 different possible observation matrices for each time step. To illustrate this, some of these possibilities are given below

Case 3: The 2x7 A_t matrix for observed temperature and salinity, with nitrate, phosphate, silicate, chlorophyll, and oxygen missing.

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Case 4: The 7x7 A_t matrix for all state variables; no observations missing.

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Case 5: The 4x7 A_t matrix for observed temperature, salinity, phosphate, and chlorophyll, with nitrate, silicate, and oxygen not observed.

$$A_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

2.3.3: Determining the Model and Observation Errors

Realistic values for error standard deviation (STD) need to be determined so they can provide an accurate reconstruction, and meaningful error bars on the output. There

are two errors types that need to be determined; the model error (W_t) from equation 1, and the observation error (V_t) from Eq. 2.

The model error (W_t) indicates just how much uncertainty there is in the model (Eq. 1) itself, i.e. how accurately it is able to predict the ocean state from one month to the next. The method used to determine the model error is a cross-validation method, and is applied only for our data rich training period. It consists of leaving out a random segment of the time series for each variable, then trying to predict what that segment should look like, and finally comparing the prediction to the left out observations. Low model error would imply the prediction and data segment are close, and high model error the opposite. The procedure for model error prediction was the following. A randomly selected six-month segment was chosen to be left out of the training data as validation data. The period of six months was chosen so the model would have to predict part of the seasonal cycle and not just individual data points. Once the validation data is removed from the training data, Φ is estimated using the multivariate regression method. The variables within the validation period are then estimated. The prediction error is calculated using the standard deviation of the residuals, the residuals being the difference between the left-out observations and their predictions from the regression. This process was repeated 100 times, each with a different randomly selected six-month data segment to be left out. Averaging these 100 experiments yields an estimate for the mean standard deviation which is taken to correspond to the model error (Table 2.1). To determine if 100 repetitions would provide a reasonable model error, the process was also conducted using 10 repetitions and 200 repetitions, but there appeared to be little difference between

the values. As such, the scaled values from the 100 repetition run were used to specify the model error.

Table 2.1: Model error standard deviation results from running 100 repetitions of estimating validation data.

Ocean State Variable	Scaled Model Error
Temperature	0.4783
Salinity	0.4506
Nitrate	0.5143
Phosphate	0.5308
Silicate	0.7202
Chlorophyll	0.9191
Oxygen	0.4776

According to Table 2.1, salinity has the lowest scaled model error at 0.4506 and chlorophyll has the highest at 0.9191. Most of the scaled model errors fall within a range of 0.45 to 0.54 with the exceptions of silicate (0.7202) and chlorophyll. This means that of the seven variables, the dynamics associated with the chlorophyll part of the model is the least well trusted.

The observation error (V_t), measures the uncertainty of the observation data. This error includes instrument or analytical errors, but for these ocean state variables it is dominated by ocean variability. Recall, the analysis is focused on monthly means, but point observations such as those obtained water samples contain variability from a variety of sources that are not present in monthly means. To quantify these observation errors, a single year (1975) within the training data set was chosen. It was assumed that the observation errors calculated during this year represent the observation errors throughout the entire time series. With many observations per month, the within-month variability was computed as a proxy for the observation error. Specifically, the monthly STD of each variable was determined for each of the twelve months, and the mean of these STDs was

used as the observation error for this project (Table 2.2). A couple of other metrics other than standard deviation were also considered to determine which one might provide the best observational error: the range and median (Table 2.3 and Table 2.4 respectively). It was felt that the range did not provide a good characterization of the observation error as it tended to include outliers (extreme highs and lows) and could not provide a consistent error from month to month or year to year. It was also felt that the median did not provide a reasonable error for the seven variables in the given year as it did not account for the spread of the data within each month. As such, the STD method was used for the magnitude of the observation error in this study

In Table 2.2 temperature is seen to have the lowest scaled observation error (0.1566) while chlorophyll has the highest (2.7788) by a large margin as the second highest scaled error is phosphate at 0.7123. This means the chlorophyll observations are less trusted when completing the reconstruction. The range of the scaled observation errors (Table 2.2) is greater than the range of the scaled model error (Table 2.1), but most of the variables have a lower observation error, than model error. Oxygen has the lowest scaled observation error for the range and median observation errors (Tables 2.3 and 2.4) and chlorophyll has the highest error when dealing with the range, but phosphate has the highest error when dealing with the median. The STD method of determining error appears to provide scaled observation errors which have values between the range and median methods. It needs to be noted that the number of samples used in these calculations varied greatly among the variables. Temperature, salinity, nitrate, phosphate, and silicate had more than 400 samples each in 1975, while chlorophyll had 31 samples and oxygen had 42 samples.

Table 2.2: Results for observation error (V_t) using standard deviation.

Ocean State Variable	Scaled Observation Error
Temperature	0.1566
Salinity	0.2313
Nitrate	0.3920
Phosphate	0.7123
Silicate	0.4691
Chlorophyll	2.7788
Oxygen	0.2291

Table 2.3: Results for observation error (V_t) using range.

Ocean State Variable	Scaled Observation Error
Temperature	0.6646
Salinity	1.2756
Nitrate	1.8288
Phosphate	3.0166
Silicate	1.9173
Chlorophyll	5.7925
Oxygen	0.5282

Table 2.4: Results for observation error (V_t) using the median.

Ocean State Variable	Scaled Observation Error
Temperature	0.2878
Salinity	0.4814
Nitrate	0.4914
Phosphate	0.6031
Silicate	0.5065
Chlorophyll	0.2915
Oxygen	0.1460

Now that all the components of the Kalman smoother equations have been collected and/or determined, the reconstruction can be completed.

2.4: Results of the Kalman Smoother

Using the information provided above, the Kalman Smoother was used to reconstruct the time series for the period 1956 to 2017 for all seven variables (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen) being studied in this project. The results are shown in Figure 2.6. Within these reconstructed time series, the temperature and salinity reconstructions follow the observations almost exactly with very little error associated with the reconstruction. At the beginning of the nitrate reconstruction (from 1956 to 1965), where there is only one observation, the series looks similar to the corresponding period of the salinity reconstruction, after which the observations fall within the error bars, but the error are generally larger than the temperature and salinity reconstruction error bars. The period of time without any observations in the phosphate does not particularly resemble the temperature or salinity time series, though there does still appear to be a seasonal cycle. When there are observations, the observations mostly lie within the error bars. Silicate's reconstruction is similar to phosphate's, though the error associated with a lack of observations in the silicate reconstruction is larger than phosphate's, and that is due to the fact that the silicate reconstruction has a lower observation error and higher model error than phosphate's reconstruction. The chlorophyll reconstruction somewhat follows the observations, and the period with no observations does not obviously resemble any other reconstruction, which is understandable given the high model and observation errors chlorophyll has. Finally, the oxygen observations mostly fall within the error bars.

To partially validate this method of statistical reconstruction, the temperature time series from OSP which was used in the Φ calculation (Figure 2.3) but was not included in

the observations used to create the reconstruction was plotted on top of the reconstruction. (Figure 2.8). For the most part, the observations and reconstruction line up, but it appears that the reconstruction undershoots the upper extremes and undershoots the lower extremes. This is likely due to any differences between the reanalysis time series and the OSP time series.

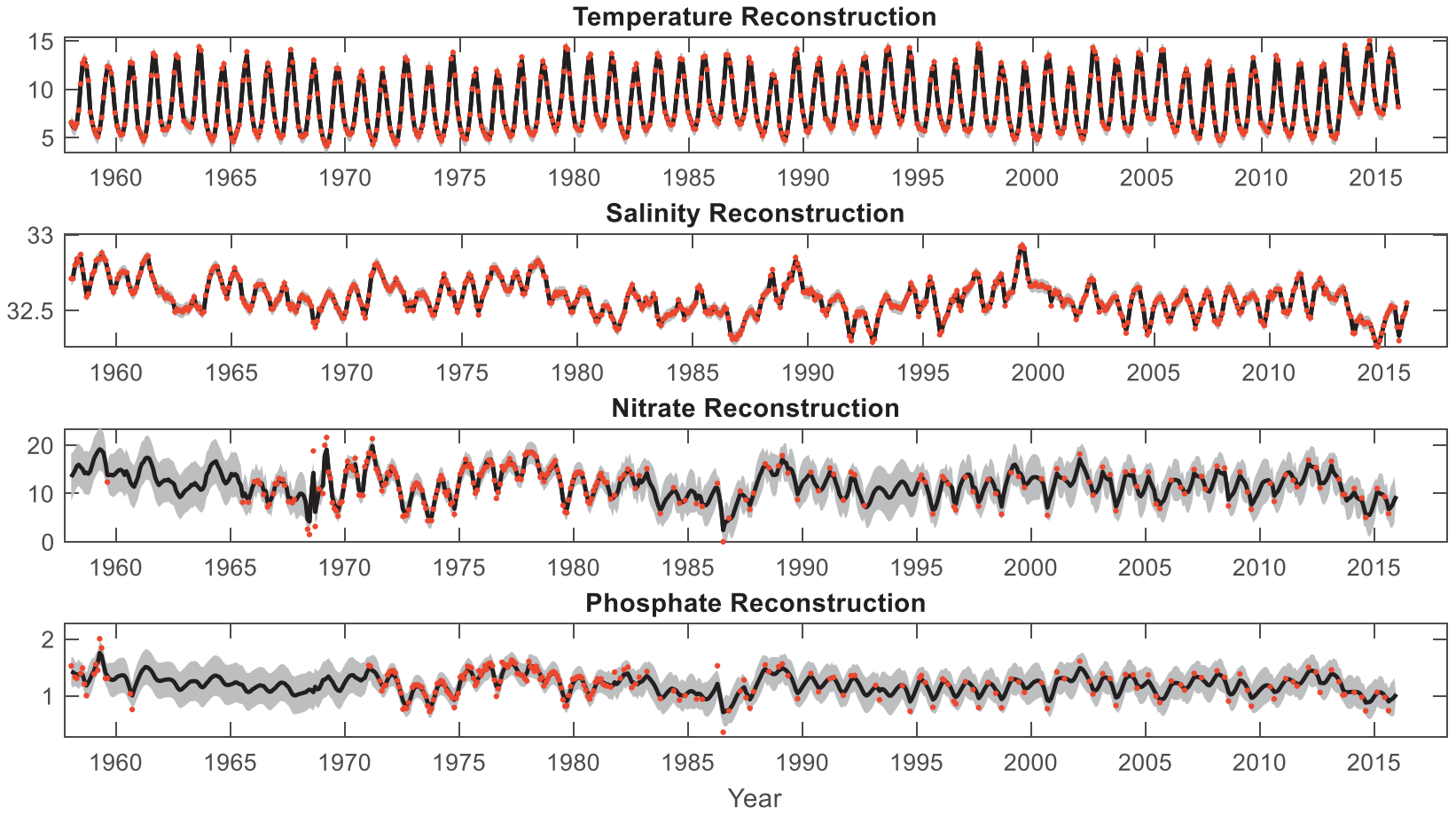

Figure 2.6: Reconstructed OSP temperature, salinity, nitrate, and phosphate using the results from the Kalman smoother from 1956 to 2017. The black line represents the mean for the reconstructed values, the grey shaded area represents the error bars based on the mean ± 2 x standard deviation. The red dots represent the DFO observations used to create the reconstruction.

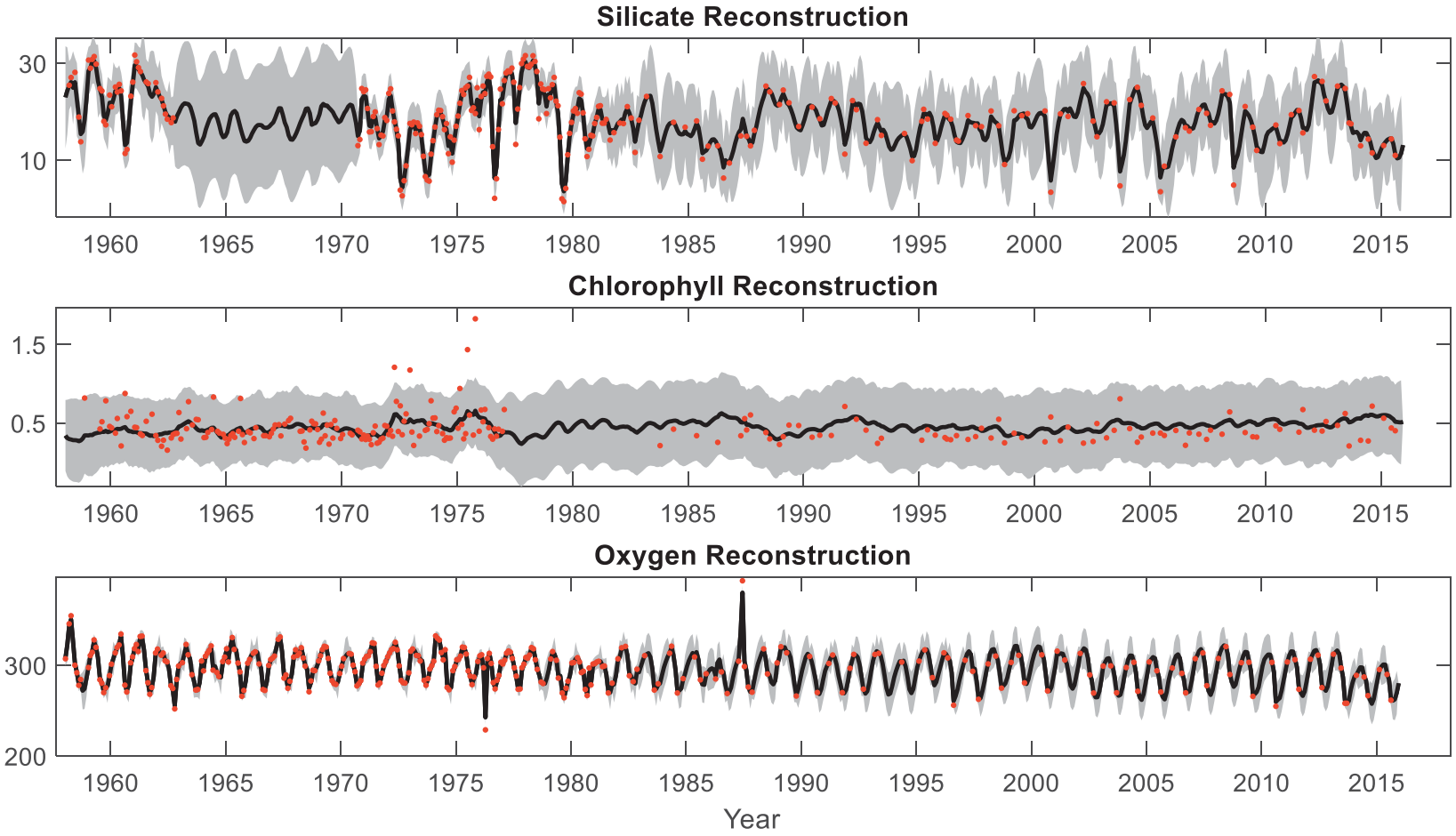

Figure 2.7: Reconstructed OSP silicate, chlorophyll, and oxygen using the results from the Kalman smoother from 1956 to 2017. The black line represents the mean for the reconstructed values, the grey shaded area represents the error bars based on the mean ± 2 x standard deviation. The red dots represent the DFO observations used to create the reconstruction.

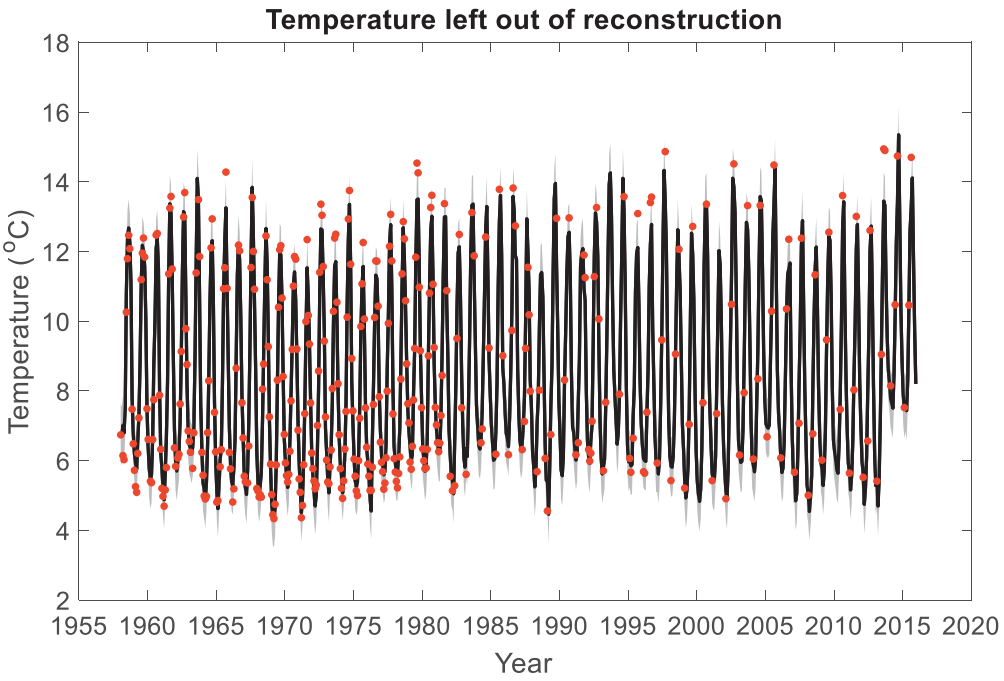

Figure 2.8: An overlay of observational temperature data aquired from the DFO (red dots) and the results of the temperature statistical reconstruction from the Kalman Smoother using the ORSA4 reanalysis data (black line). Shaded grey regions denote the confidence interval for the statistical reconstrucion.

The scaled seasonal cycle was examined by looking at the average monthly values for the seven variables (Figure 2.9). The seasonal cycle is most visible in temperature and oxygen. Temperature has its highest values in August/September and Oxygen has its lowest values in September/October, most likely resulting from negative correlation between temperature and oxygen (warmer water is less capable of holding dissolved oxygen than colder water). The weaker seasonal cycles for nitrate, phosphate, silicate, and chlorophyll are to be expected at OSP as it is located within an HNLC region (Pitchford and Brindley, 1999) and does not normally experience seasonally predictable phytoplankton blooms.

To get a better understanding as to whether there are cycles longer than the annual seasonal cycle in the time series, the seasonal cycle (and any other cycle shorter than ENSO's cycle) was removed using a low pass filter (a filter that removes frequencies

higher than a selected frequency, 2 years in this case and using a zero-phase Butterworth filter). The resulting sub-seasonal time series (Figure 2.10) displays any variability on time scales longer than 2 years, which is more appropriate for relating to the longer climate cycles that are of interest. From a visual inspection temperature and oxygen have a 3-5 year cycle, silicate and nitrate have a 5-10 year cycle, while salinity has a longer cycle that looks to be at approximately decadal. Phosphate and chlorophyll don't appear to have any strong cycle based on visual inspection alone.

Looking at the first half of the time series, there appears to be a drop in temperature during the late-1960s to early-1970s, which has been documented before, as shown by Jones et al. (2016) who note that "the northern hemisphere record shows gradual cooling from the mid-1940s through to the mid-1970s". There also appears to be a decline in salinity during this period. Aside from the temperature and salinity time series, there is nothing that stands out from any of the first half of the time series apart from their 3-10 year cycles. During the second half of the time series there is not much worth noting except that the temperature and silicate scales are higher, and the salinity scale is lower than in the first half of the time series. Most of the variables appear to have a 2-3 year cycle in the time series, but a visual inspection cannot tell much more than that. In Chapter 3, we will use a more quantitative spectral analysis to investigate these cycles further.

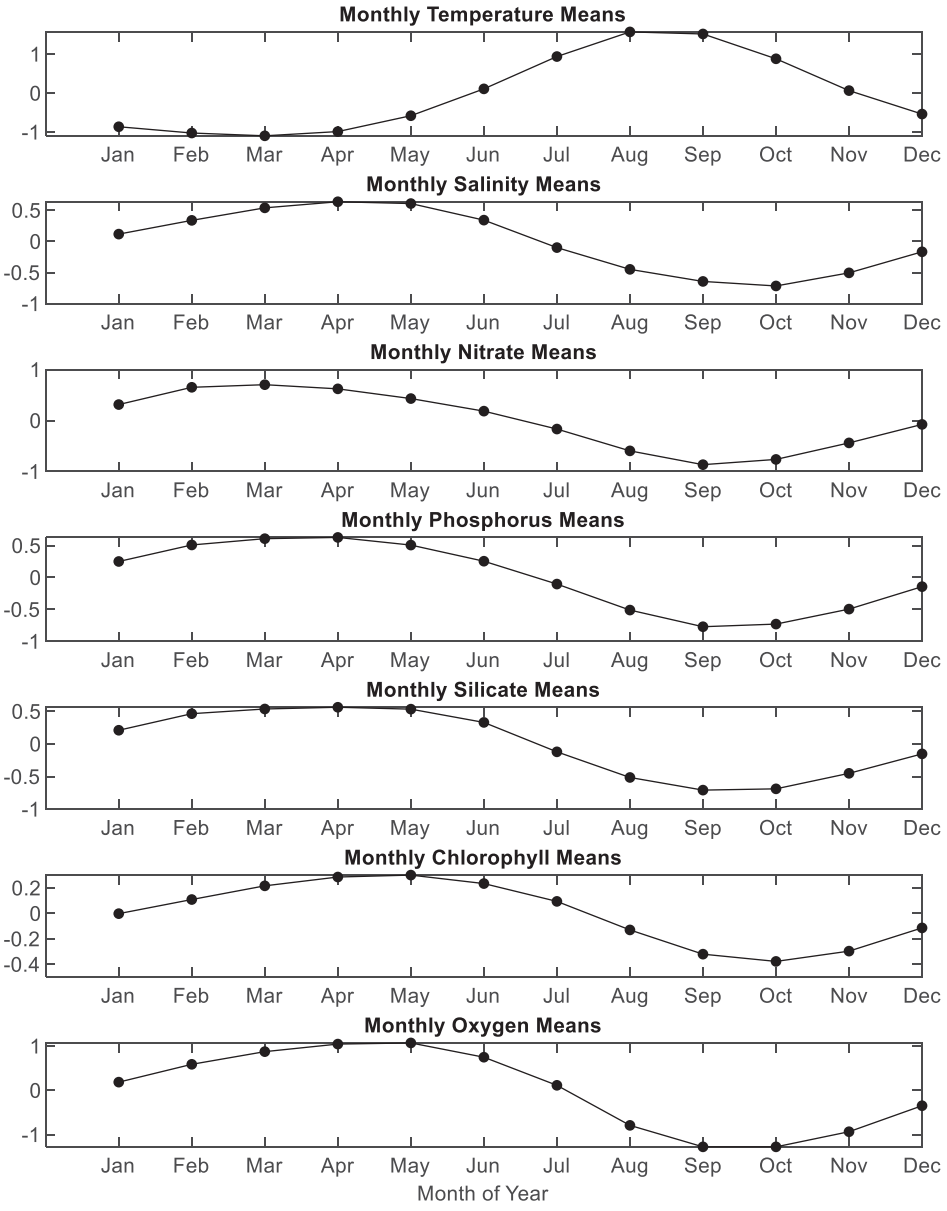

Figure 2.9: Scaled, monthly averages throughout the calendar year for each of the reconstructed seven variables in the study (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen).

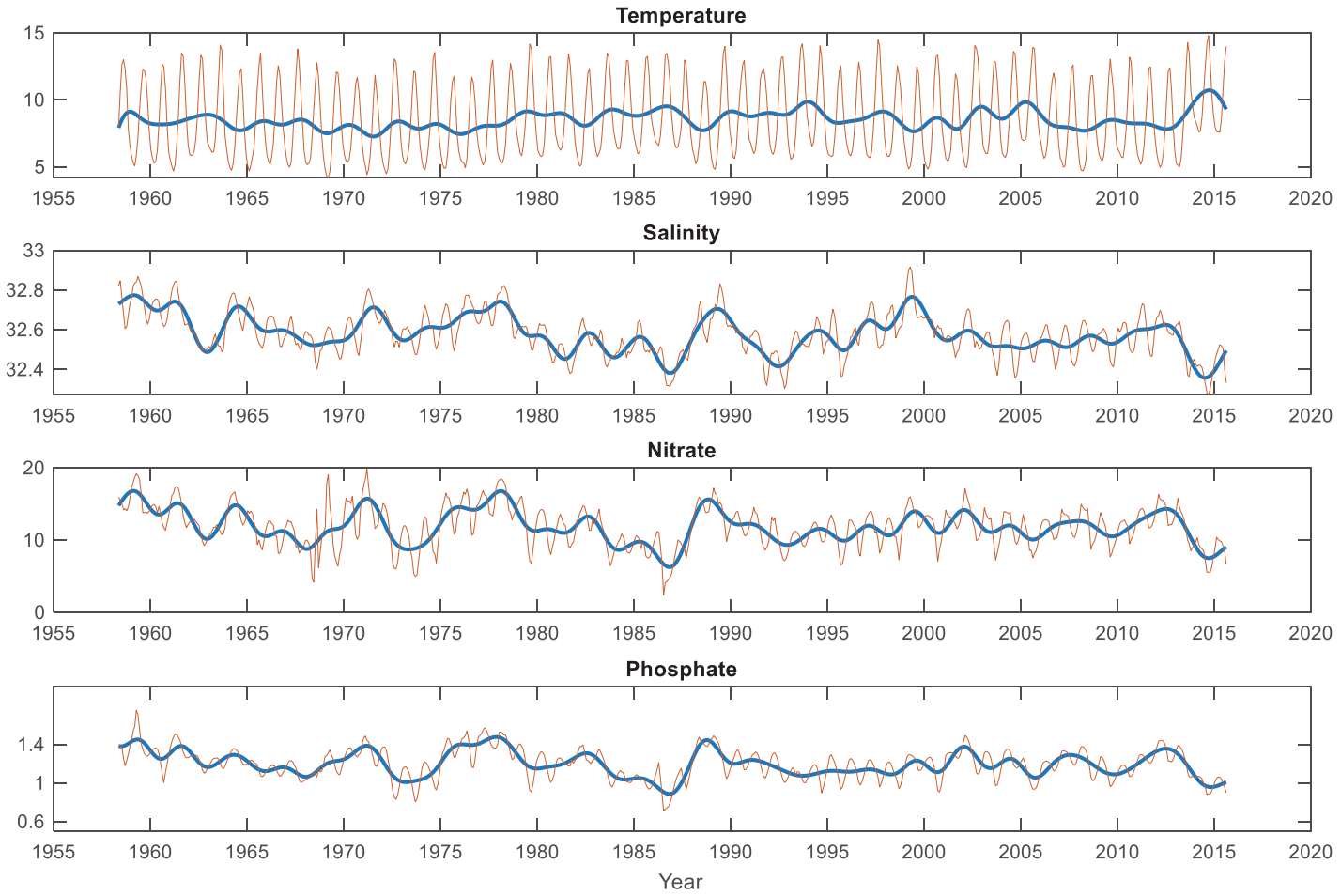

Figure 2.10: Reconstructed (red line) and smoothed (blue line) ocean state variables (temperature, salinity, nitrate, phosphate). These are based on application of a low pass filter which isolates everything with a cycle of 2 years or less within the reconstruction time series.

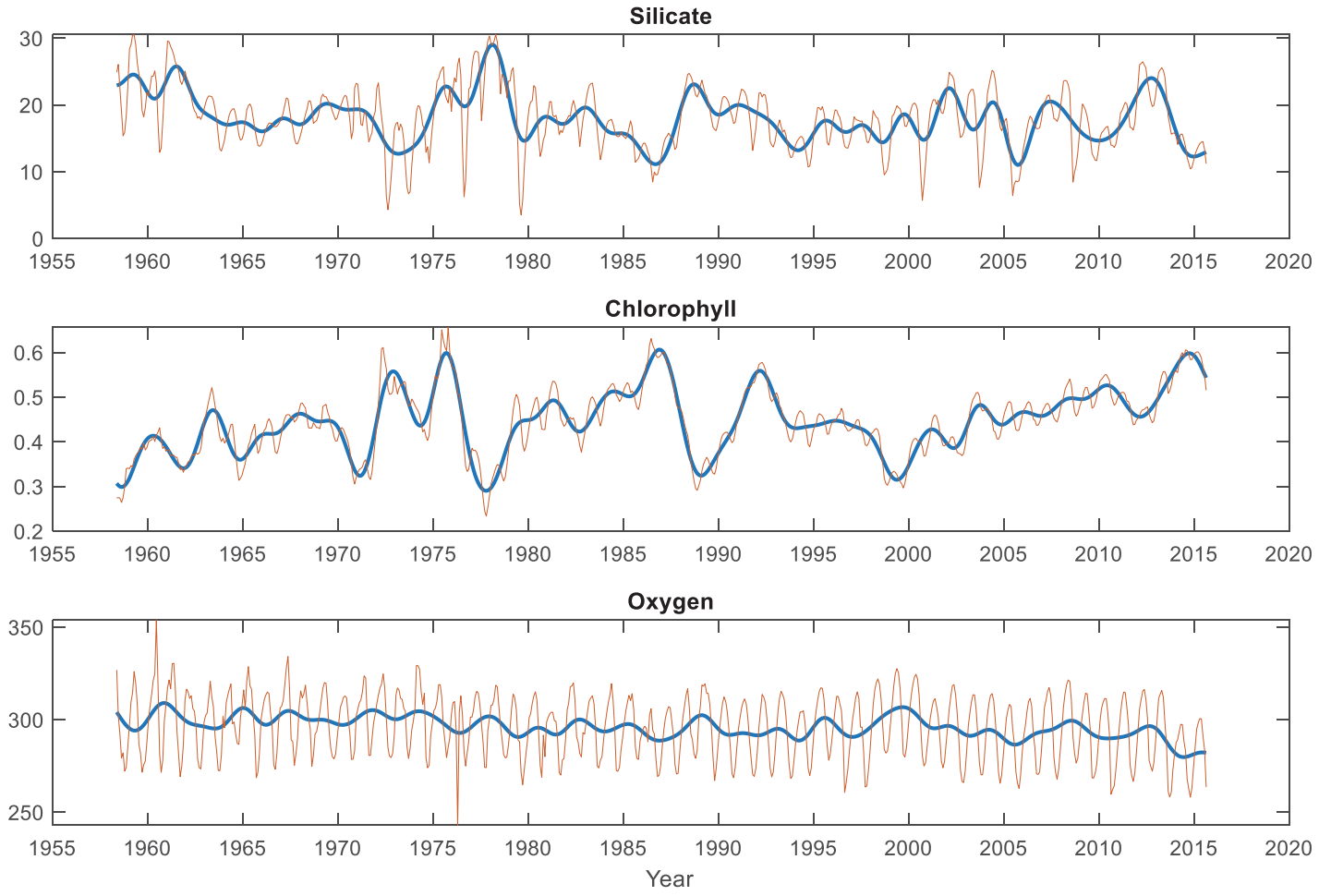

Figure 2.11: Reconstructed (red line) and smoothed (blue line) ocean state variables (silicate, chlorophyll, and oxygen). These are based on application of a low pass filter which isolates everything with a cycle of 2 years or less within the reconstruction time series.

Looking at how the ocean state variables relate to each other may help in the analysis of the ocean variable/climate mode comparisons in that they may aid in explaining some results that do not follow what has been recorded in literature or have not been well documented in literature. When the sub-seasonal time series are correlated against each other (Table 2.5), the variables show significant correlations (p-values < 0.05) across the board (Figures with pairwise scatter plots are in Appendix C (Figures C.1-C.7)). Note that the p-values were computed by transforming the correlation to create a t-statistic having N-2 degrees of freedom, where N is the number of data points in each time series.

Temperature has a negative correlation with all variables except chlorophyll, which follows the Φ matrix. All other variables follow this pattern; where temperature and chlorophyll are anti-correlated to the other variables. The strongest correlation value for the variables is the nitrate/phosphate correlation with a value of 0.95.

Table 2.5: Variable vs variable correlations from the low pass filter series from Figure 2.11. The r column depicts the correlation between the variables, and the p-value column depicts the p-values for the correlations at 95% CI.

Variable Pairing	r	p-value
Temperature/Salinity	-0.57	<0.01
Temperature/Nitrate	-0.54	<0.01
Temperature/Phosphate	-0.55	<0.01
Temperature/Silicate	-0.45	<0.01
Temperature/Chlorophyll	0.36	<0.01
Temperature/Oxygen	-0.70	<0.01
Salinity/Temperature	-0.57	<0.01
Salinity/Nitrate	0.83	<0.01
Salinity/Phosphate	0.68	<0.01
Salinity/Silicate	0.56	<0.01
Salinity/Chlorophyll	-0.77	<0.01
Salinity/Oxygen	0.64	<0.01

Variable Pairing	r	p-value
Nitrate/Temperature	-0.54	<0.01
Nitrate/Salinity	0.83	<0.01
Nitrate/Phosphate	0.95	<0.01
Nitrate/Silicate	0.82	<0.01
Nitrate/Chlorophyll	-0.71	<0.01
Nitrate/Oxygen	0.41	<0.01
Phosphate/Temperature	-0.55	<0.01
Phosphate/Salinity	0.68	<0.01
Phosphate/Nitrate	0.95	<0.01
Phosphate/Silicate	0.90	<0.01
Phosphate/Chlorophyll	-0.63	<0.01
Phosphate/Oxygen	0.37	<0.01
Silicate/Temperature	-0.45	<0.01
Silicate/Salinity	0.56	<0.01
Silicate/Nitrate	0.82	<0.01
Silicate/Phosphate	0.90	<0.01
Silicate/Chlorophyll	-0.54	<0.01
Silicate/Oxygen	0.38	<0.01
Chlorophyll/Temperature	0.36	<0.01
Chlorophyll/Salinity	-0.77	<0.01
Chlorophyll/Nitrate	-0.71	<0.01
Chlorophyll/Phosphate	-0.63	<0.01
Chlorophyll/Silicate	-0.54	<0.01
Chlorophyll/Oxygen	-0.58	<0.01
Oxygen/Temperature	-0.70	<0.01
Oxygen/Salinity	0.64	<0.01
Oxygen/Nitrate	0.41	<0.01
Oxygen/Phosphate	0.37	<0.01
Oxygen/Silicate	0.38	<0.01
Oxygen/Chlorophyll	-0.58	<0.01

Chapter 3 – Comparing Reconstructed Ocean Variables to Climate Indices

3.1: Introducing the Climate Indices

In this chapter, the climate indices used in this project will be introduced, including how they are computed and what their typical cycle lengths are. They will then be compared and contrasted to the reconstructed OSP ocean time series. Specifically, the important climate mode frequencies will be identified so the ocean state variables at OSP can be compared with the climate indices. The SOI, PDO, and MEI data were provided by the NOAA ESRL Physical Sciences Laboratory, Boulder, Colorado, USA, from their website at http://psl.noaa.gov/, and the NPGO data were provided by (Di Lorenzo, 2019).

3.1.1: Southern Oscillation Index

The Southern Oscillation Index (SOI) (NOAA, 2020) is an index that is based on the differences in standardized sea level pressure between Tahiti and Darwin, Australia, where the negative phase of SOI is representative of lower air pressure at Tahiti and relative to that at Darwin, and vice versa for a positive phase (Power & Kociuba, 2010). The SOI corresponds well with changes in sea surface temperature across the eastern Tropical Pacific, so prolonged periods of positive or negative phase coincide with abnormally cold or warm water (respectively) across the eastern Tropical Pacific. These prolonged periods are referred to as El Niño and La Nina, and the oscillation between these phases is known as the El Niño--Southern Oscillation (ENSO). El Niño and La Nina typically occur every 2 to 7 years, which is interannual variability rather than decadal variability, which is the focus on this thesis, but as McPhaden et al. (2006) state,

"ENSO is unique among climate phenomena in its strength, predictability, and global influence," so it would be remiss to exclude it from this study. The effects of ENSO on ocean temperature and ecosystems have been well studied over the years, with a small sample of these studies being: Di Lorenzo et al. (2010), Schwing et al. (2002), Miller et al., (2004), and Houk et al. (2020).

The monthly SOI data in Figure 3.1 (as well as all of the climate data) were smoothed using the same 2-year low pass filter to remove the high-frequency variability in the data. This practice helps in identifying the dominant cycle(s) in the data, which appears to be around 5 years in Figure 3.1, and aids in determining any overall trends in the data. Smoothing climate data in this manner is common practice to determine longer scale (decadal, centennial) variability. For example, all variability with periods less than or equal to 8 years was removed in a study of climate variability in Australia (Power et al 1999), and a 10-year moving average was used to smooth the data for a global multidecadal climate variability (McCabe & Palecki 2005). As can be seen in Figure 3.1, smoothing the data eliminates the high-frequency variability in the data while maintaining the slower, mulit-year variability. A cut off of 2 years was chosen for the low pass filter, as opposed to the longer 8 and 10 years mentioned above because ENSO is one of the climate variations that is of interest for this study.

Along with changes in SST, ENSO events are characterized by global shifts in winds, temperature, and precipitation patterns (Fasullo et al., 2018). During El Niño years, the impact on the marine ecosystem spans the majority of the Pacific Ocean (including OSP), and the resulting decline in primary productivity affects the mortality, fertility, and geographic distribution of marine mammals and fish species, including fish

species that are economically and socially important (McPhaden et al., 2006). As year to year variability in carbon concentrations in the atmosphere are also dominated by ENSO events (McPhaden et al., 2006), SOI is a medium to long scale climate index which is of interest in how it affects biogeochemical variables in the sea surface by itself and in combination with PDO and NPGO (discussed below).

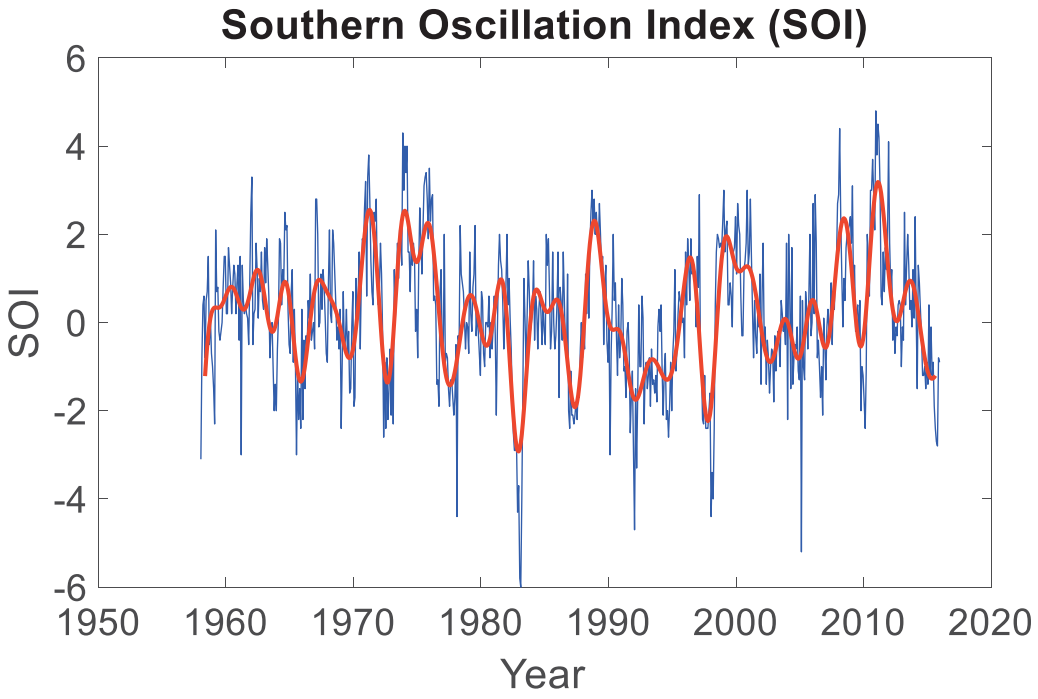

Figure 3.1: Time series of the monthly (blue line) and smoothed with a 2-year low-pass filter (red line) data for the SOI index.

3.1.2: Pacific Decadal Oscillation

The Pacific Decadal Oscillation (PDO) (Mantua, 2018) is a cyclical pattern of ocean-atmosphere climate variability which is centred over the northern, mid-latitude Pacific basin. It was first acknowledged as being a climate oscillation in 1997 with a paper by Mantua et al. (1997), looking at the impacts it has on salmon production. The PDO is calculated by taking the spatial average of the monthly SST of the Pacific Ocean

north of 20°N. To account for global warming, the global average anomaly is subtracted from this value (Mantua & Hare, 2002). The PDO's decadal cycle produces an El-Niño-like spatial pattern of Pacific climate variability (though more meridionally extended, and with an opposite pole in the northwest Pacific), which can be recovered from sea surface temperature anomalies for the north Pacific (Schneider & Cornuelle, 2005). Due to this, it is widely used as an index for decadal variability for the Pacific climate (Schneider & Cornuelle, 2005). The decadal cycle can be seen in Figure 3.2, along with some shorter cycles. It also shows that PDO has been in a predominantly negative phase since the mid-1990s.

Studies have shown that there have just been two full PDO cycles in the last century; cool (negative) periods from 1890-1924 and 1947-1976, and warm (positive) periods from 1925-1946 and 1977-1995 (Mantua & Hare, 2002). In western North America, PDO produces similar, but weaker, climatic conditions than El Niño; decreased winter precipitation, snowpack, and stream flow in the north west, and higher precipitation in the south west during a positive phase (Macdonald & Case, 2005). The climatic conditions are the opposite during a negative PDO phase. It was noted that the regime shift in the 1970's led to an increase in most Alaskan salmon populations and a decrease in the west coast salmon populations as a result of an overall warming of the North East Pacific (NEP) (Hare and Mantua 2000). Fluctuations in the strength in both the positive and negative phase occur in ENSO time bands (Macdonald & Case, 2005). It has also been found that for the past 200 years there has been a 50 to 70-year periodicity in PDO (Macdonald & Case, 2005).

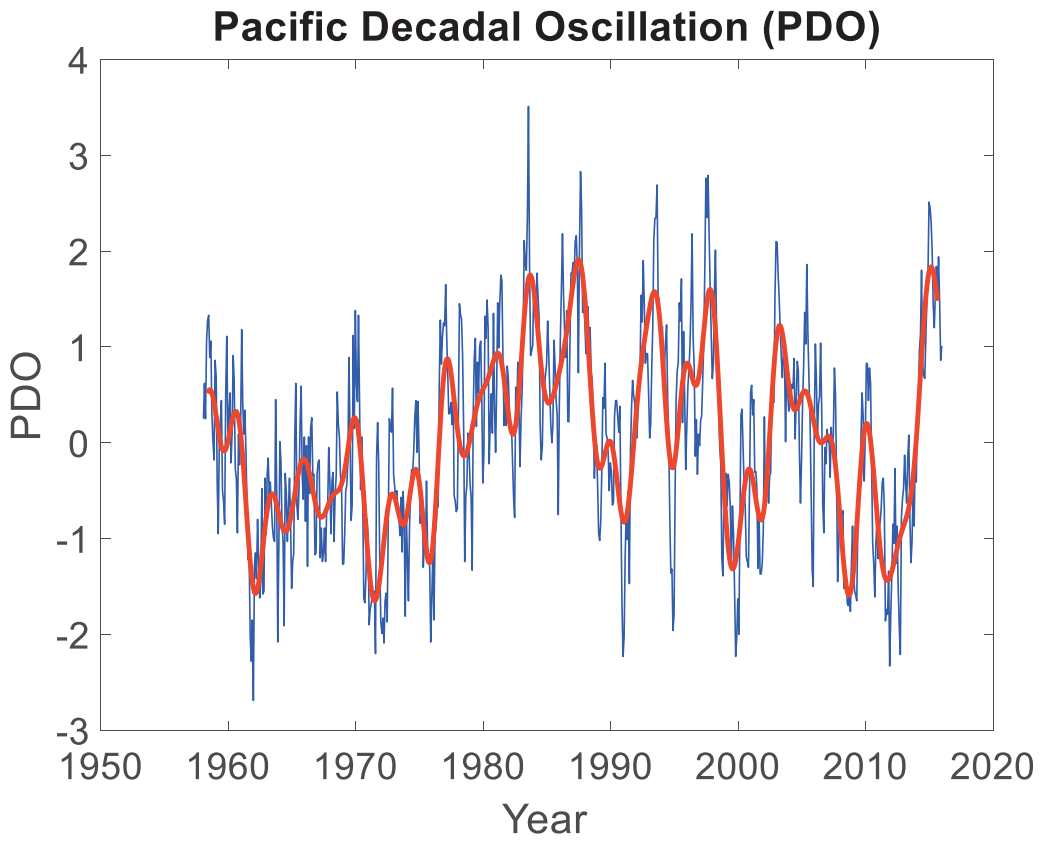

Figure 3.2: Time series of the monthly (blue line) and smoothed with a 2-year low-pass filter (red line) data for the PDO index.

3.1.3: North Pacific Gyre Oscillation

Another large-scale climate index is the North Pacific Gyre Oscillation (NPGO) (Di Lorenzo, 2019). First identified in 2008 by Di Lorenzo et al., NPGO has a decadal variability that is characterised by sea surface temperature anomalies (SSTa), but is determined by its second most dominant feature, sea surface height anomalies (SSHa) (Di Lorenzo et al., 2008, Li Yi et al., 2015). It is defined as the time series of the second empirical orthogonal function (EOF) of the SSHa from 25°N to 65°N, and 180°W to 110°W (Li Yi et al, 2015). The monthly averages for NPGO were computed from Di Lorenzo et al. (2008)'s spatial pattern of SSHa from 1950 to 2004, and after that they were computed from satellite SSHa. The discovery of NPGO provided an explanation for

fluctuations in salinity, nutrients, and chlorophyll which were not able to be attributed to other climate variability such as PDO and ENSO (Di Lorenzo et al., 2008).

Even though NPGO was discovered so recently, there have still been studies done on its relationship with the ocean. The NPGO is the oceanic depiction of the North Pacific Oscillation (NPO) (Linkin & Nigam, 2008), and is forced by the atmosphere (Chhak et al., 2009), indicating the relationship between the atmosphere and the ocean. There has also been a study that links multi-year heat waves to the dynamics of both PDO and NPGO (Joh &, Di Lorenzo, 2017). Given that NPGO displays a distinct decadal variability with a significant spectral peak at 18 years (Figure 3.3) (Li Yi et al., 2015), and it can result in SSTa that resembles El Niño-like patterns (Di Lorenzo et al., 2008), NPGO is thus a low-frequency climate index that can significantly influence sea surface nutrients and chlorophyll, and therefore microorganisms like phytoplankton. There appear to be a couple of other minor cycles as well in Figure 3.3, but the dominant cycle is the 18-year cycle, and there does not appear to be much of a trend for the last 60 years.

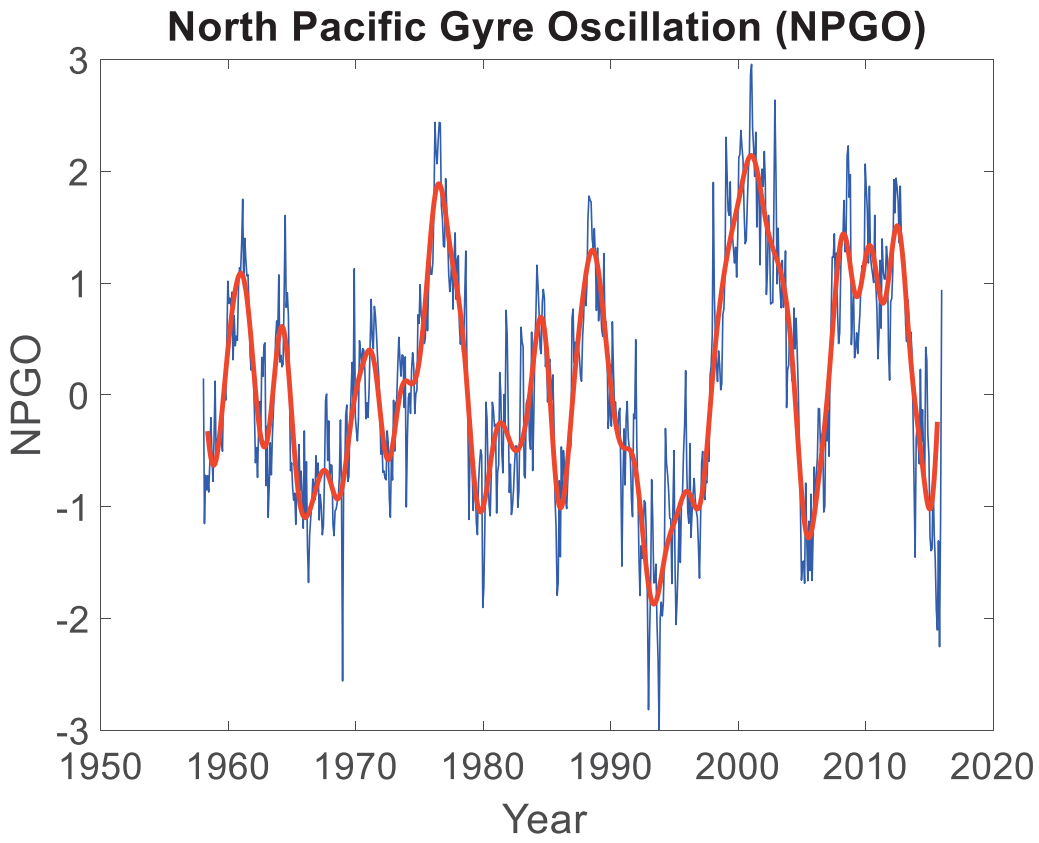

Figure 3.3: Time series of the monthly (blue line) and smoothed with a 2-year low-pass filter (red line) data for the NPGO index.

3.1.4: Multivariate El Niño Southern Oscillation Index

The multivariate El Niño Southern Oscillation Index (MEI) (Wolter, 2018) is another method of measuring ENSO events in the Tropical Pacific. The standard SOI is a useful indicator of ENSO events, but fails to capture the ocean-atmosphere interactions (Mazzarella et al., 2013). The MEI is based on the principal component of six variables over the tropical Pacific. Those six variables are: sea level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and the cloudiness of the sky (Mazzarella et al., 2013). Given the differences in the computational method to obtain MEI and SOI, and that the resulting indices are

noticeably different from each other (Figure 3.1 and Figure 3.4), we include both MEI and SOI in this study. While the monthly values for MEI (Figure 3.4, blue line) are less variable than the monthly values for SOI, the filtered values (Figure 3.4, red line) show a similar pattern. There is also a noticeable upward trend that is not seen in the SOI data, likely due to the role of absolute temperature rise.

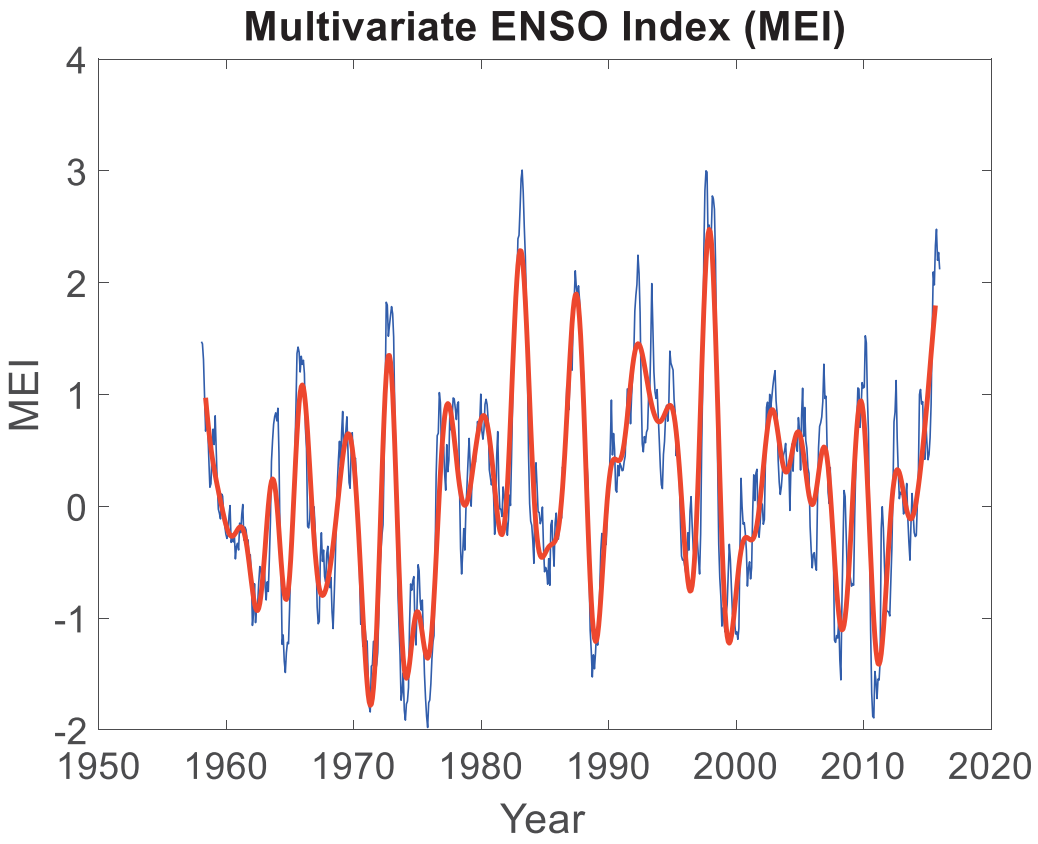

Figure 3.4: Time series of the monthly (blue line) and smoothed with a 2-year low-pass filter (red line) data for the MEI index.

3.2: Determining the Frequencies of Interest

The dominant frequencies for the climate modes cycle need to be determined so the ocean state variables can be analysed at these frequencies as well. These frequencies will be isolated using spectral analyses of the climate modes (Figures 3.5-3.8) and then

frequencies have the highest spectral power will be determined. The power spectrum was calculated using Welch's method (Shumway & Stoffer, 2017), (Figures 3.5b-3.8b) and shows the strength of the periodic (recurring) signals at all of the frequencies resolvable with the climate data. It should be noted that the exact value of the power does not matter as much as its relative strength due to the variation in the magnitudes of the time series. The effects of the filtering can be also seen in Figures 3.5b-3.8b where there are no peaks in the power spectrum beyond a frequency of 0.5 (a two-year cycle). Note that if the power spectrum were done on the original monthly data, a peak would be expected at a frequency of 1.0 due to the seasonal cycle. To account for the relatively long-term cycles these climate modes operate at, only the powers of the lowest frequencies (from 0 to 0.6) were considered (Figures 3.5c- 3.8c), where for example, a frequency of 0.1 has a 10-year cycle. The upper frequency limit was chosen based on the fact a low pass filter with a frequency cutoff of two years was used.

Since this study is interested in how the climate modes affect the biogeochemical variables, the two frequencies which contain the two highest powers for each of the climate indices have been selected and are highlighted in Table 3.1. These are the largest values for the frequencies of interest within the range of 0-0.6 that are resolvable given the time series length. Further spectral analyses, particularly cross-spectral analyses with the ocean state variables being studied in this project, will also focus on these frequencies of interest as stronger signals at these frequencies could indicate a correlation between the climate modes and ocean state variables.

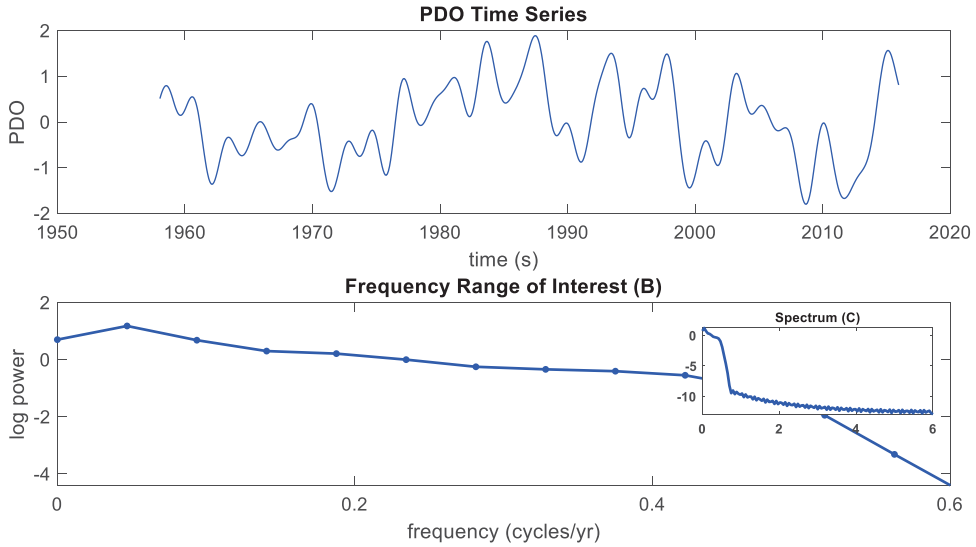

Figure 3.5: The time series and power spectrum of the PDO index. Panel A shows the filtered PDO time series from 1958 to 2017. Panel B shows the power spectrum as a log power per frequency for all frequencies. Panel C shows the power spectrum as a log power per frequency for the range of possible climate mode frequencies.

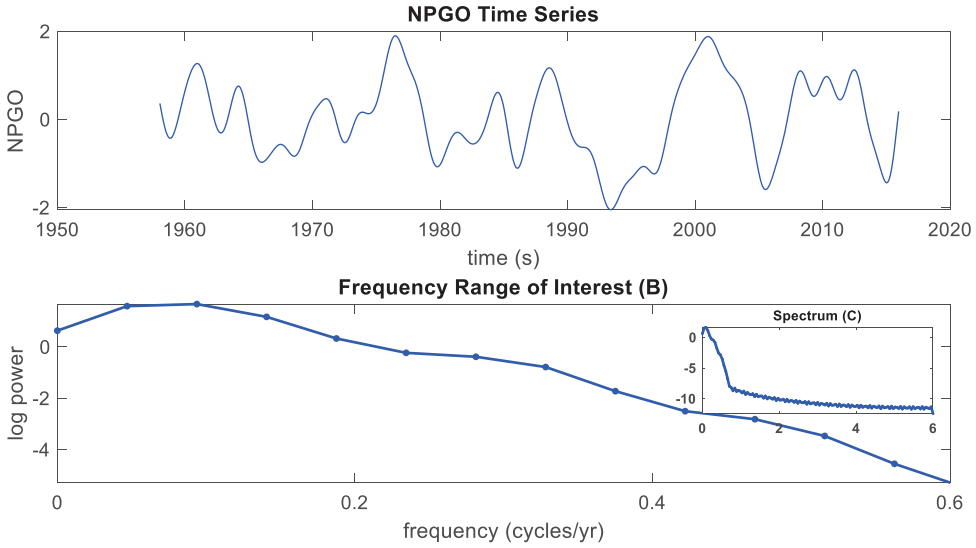

Figure 3.6: The time series and power spectrum of the NPGO index. Panel A shows the smoothed NPGO time series from 1958 to 2017. Panel B shows the power spectrum as a log power per frequency for all frequencies. Panel C shows the power spectrum as a log power per frequency for the range of possible climate mode frequencies.

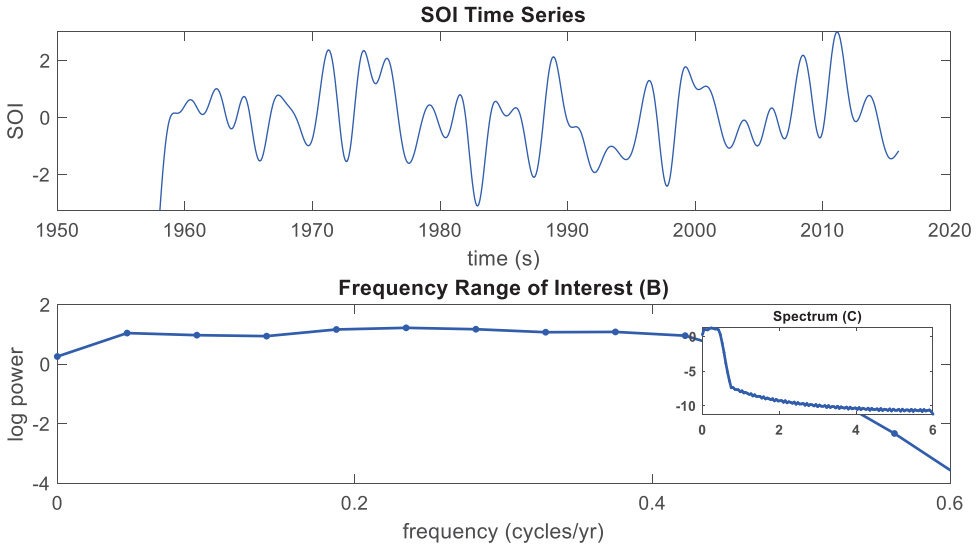

Figure 3.7: The time series and power spectrum of the SOI index. Panel A shows the smoothed SOI time series from 1958 to 2017. Panel B shows the power spectrum as a log power per frequency for all frequencies. Panel C shows the power spectrum as a log power per frequency for the range of possible climate mode frequencies.

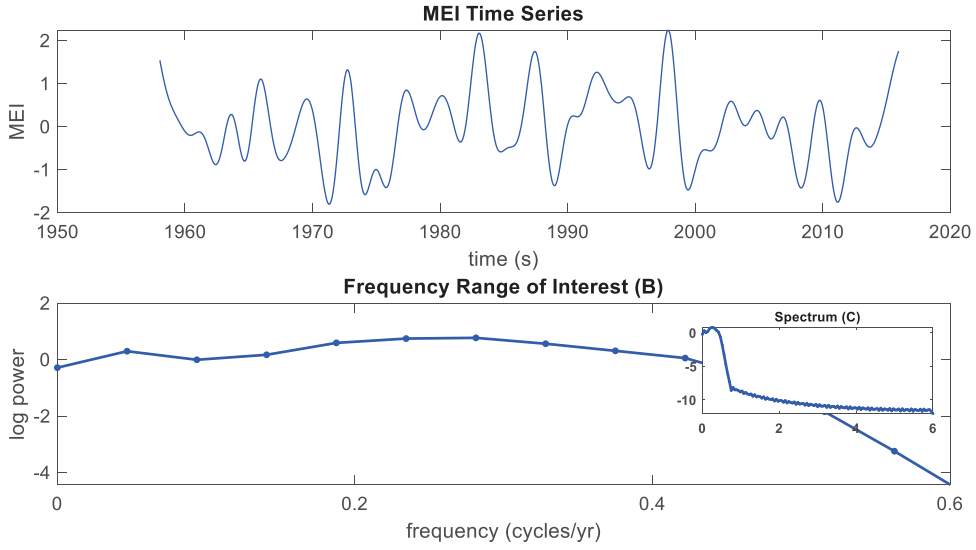

Figure 3.8: The time series and power spectrum of the MEI index. Panel A shows the smoothed MEI time series from 1958 to 2017. Panel B shows the power spectrum as a log power per frequency for all frequencies. Panel C shows the power spectrum as a log power per frequency for the range of possible climate mode frequencies.

Table 3.1: The frequencies, power spectral density, and corresponding period of associated cycles for the climate indices. The highlighted rows indicate the frequencies (and their associated cycles) with the two highest powers for each climate index.

Climate index	Frequency	Period (years)	Power
PDO	0.04688	21.33	1.182
	0.09375	10.66	0.684
	0.1406	7.11	0.303
	0.1875	5.33	0.214
NPGO	0.04688	21.33	1.579
	0.09375	10.66	1.659
	0.1406	7.11	1.163
	0.1875	5.33	0.320
SOI	0.04688	21.33	1.045
	0.09275	10.66	0.975
	0.1406	7.11	0.942
	0.1875	5.33	1.166
MEI	0.04688	21.33	0.294
	0.09375	10.66	-0.008
	0.1406	7.11	0.168
	0.1875	5.33	0.590
	0.2344	4.27	0.742
	0.2812	3.55	0.768

According to Figure 3.5 and Table3.1, PDO has a dominant cycle of roughly 20 years (frequency of 0.04688) and a secondary cycle of 10 years (frequency of 0.1875). The NPGO (Figure 3.6) has dominant cycles of 10 years (frequency of 0.09375) and 20 years. The SOI (Figure 3.7) has a dominant cycle of 5 years and a secondary cycle of 20 years, and MEI (Figure 3.8) has a dominant cycle of about 4 years. The secondary cycles in PDO and SOI are likely reflections of each others' cycles as the two climate modes are known to be related but with different periods. Compared to the other indices, MEI does not have any relatively strong, longer cycles (10+ years), and NPGO does not have any strong shorter cycles. That being said, the frequency of 0.1406 (seven-year cycle) is not a

dominant frequency for any of the indices (Table 3.1), suggesting that climate variations between 5 and 10 year periods are relatively weak.

3.3: Comparing Climate Modes with the Ocean Variables

To compare and contrast the climate modes to the ocean state variables, correlations, cross-correlations, and cross-spectral analyses were completed to identify the strength of the relationships as well as any time lags that may occur in the signals. The climate modes and ocean variables were first correlated against each other (i.e. at the same time steps hence no time lags). Using a p-value of 0.05, Table 3.2 displays the results of the significant correlations between the variables (all correlation table scan be found in Appendix B). In terms of the climate indices, chlorophyll has the weakest correlations with all four climate indices compared to the other ocean variables. Temperature appears to have the strongest correlations for all four climate modes compared to the other variables. Salinity, nitrate, phosphate, silicate, and chlorophyll all have correlations ranging from weak to moderate with the correlations with PDO being the strongest across the board.

Table 3.2: The correlations between the ocean state variables (temperature, salinity, nitrate, phosphate, silicate, chlorophyll, and oxygen). These are represented respectively in the table by their first letter (and second for silicate). All values are significant at a 95% CI.

	T	**S**	**N**	**P**	**Si**	**C**	**O**
PDO	0.62	-0.45	-0.43	-0.37	-0.31	0.25	-0.48
NPGO	-0.47	0.39	0.47	0.45	0.37	-0.19	0.23
SOI	-0.40	0.29	0.24	0.18	0.08	-0.09	0.26
MEI	0.45	-0.34	-0.30	-0.25	-0.18	0.18	-0.39

When looking at the time series plots of the climate indices together with the ocean state variables (Figures 3.9-3.15), some of these correlations can be seen with just a visual inspection. Temperature (Figure 3.9) does appear to be correlated with PDO except for a brief period during the early-1960s. It does, however, appear to have a negative correlation with NPGO and SOI throughout the entire time series, and a positive correlation with MEI, though not as strong as the correlation with PDO. These observations concur with the values in Table 3.2 in that temperature has a weak correlation with PDO, a strong correlation with NPGO and SOI, and a moderate correlation with MEI.

In terms of salinity (Figure 3.10), there are periods where it appears correlated with PDO (1956 to 1970), but most of the time it appears to be negatively correlated. When comparing salinity to the NPGO time series, where the two time series look to be correlated for most of the period. The case is the same for the salinity and SOI time series. For the salinity and MEI time series, however, there does appear to be a negative correlation between the two as seen in Figure 3.10, and that is corroborated by the corresponding low value in Table 3.2.

Nitrate (Figure 3.11) appears to be negatively correlated with PDO throughout the entire time series. There appears to be a positive correlation between nitrate and NPGO for the whole time series as well. SOI, however, appears to have both positive and negative correlations with nitrate within the time series (e.g. positive in the early 1960s, and negative in the early 1980s), likely indicating that there is not a strong overall correlation between the two, or that the assumption that Φ does not change over time is incorrect There also seems to be a bit of a negative correlation between nitrate and MEI

(seen from 1990 to 2000). The values in Table 3.2 show that there is a moderately strong, negative correlation between nitrate and PDO; that there is a moderately strong, positive correlation between nitrate and NPGO; a moderately weak, positive correlation with SOI; and that there is a moderately weak, negative, correlation between nitrate and MEI.

When looking at the phosphate/climate mode time series (Figure 3.12), the PDO time series appears to have a negative correlation. For NPGO and SOI there appears to be positive correlations between the variables, though NPGO appears to have a stronger correlation than SOI. MEI appears to have negative correlation with phosphate, but weaker than the correlation between phosphate and PDO. Table 3.2 corroborates these observations.

Silicate (Figure 3.13) appears to have similar correlations to the phosphate/climate mode time series. With the silicate/PDO and silicate/MEI time series appearing to have negative correlations. The NPGO series appears to have a negative correlation between the variables. The SOI correlation is difficult to determine visually, which corresponds to the weaker correlation value in Table 3.2.

When it comes to chlorophyll (Figure 3.14), there is a negative correlation with PDO in the mid to late-1970s, but a positive correlation for the rest of the time series. The time series of chlorophyll and NPGO are generally doing the opposite thing hence they are anti-correlated, but not strongly. The SOI also appears to be anti-correlated with chlorophyll, but not as strongly as NPGO. When it comes to the MEI time series, some parts look like they are correlated, but as other parts look like they are anti-correlated there is likely a weak correlation between the two variables.

Oxygen (Figure 3.15) does appear to have negative correlation with PDO. With a few exceptions (early-1970s, mid-1990s) there is a moderately weak, positive correlation between oxygen and NPGO. The oxygen/SOI correlation seems to be slightly stronger than the oxygen/NPGO correlation. MEI is similar to PDO in that there is a moderately strong, negative correlation with oxygen. The values in Table 3.2 match these observations.

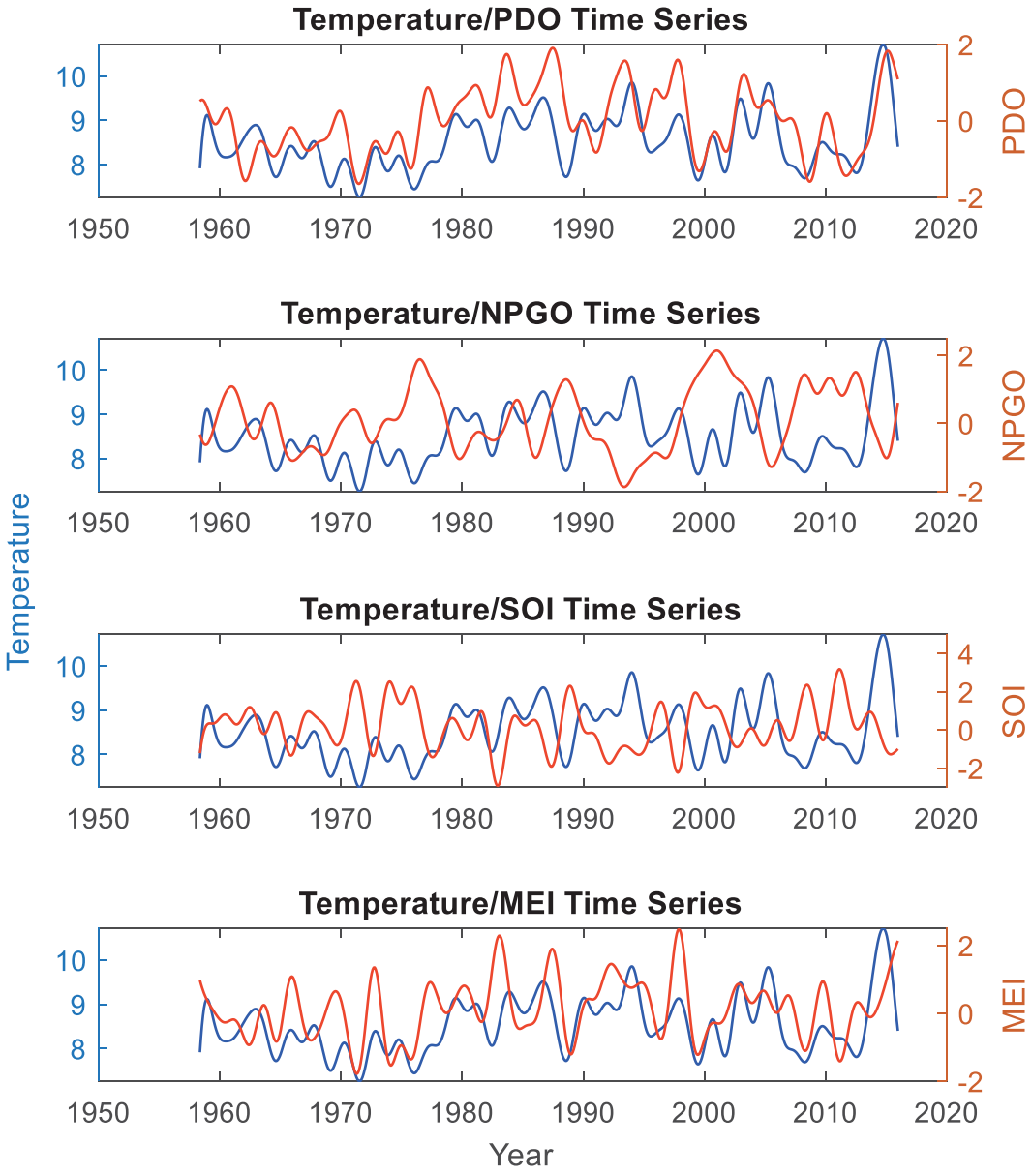

Figure 3.9: Overplots of temperature and the climate modes time series. Temperature is depicted in blue while the climate modes are in red. Units are as previously established.

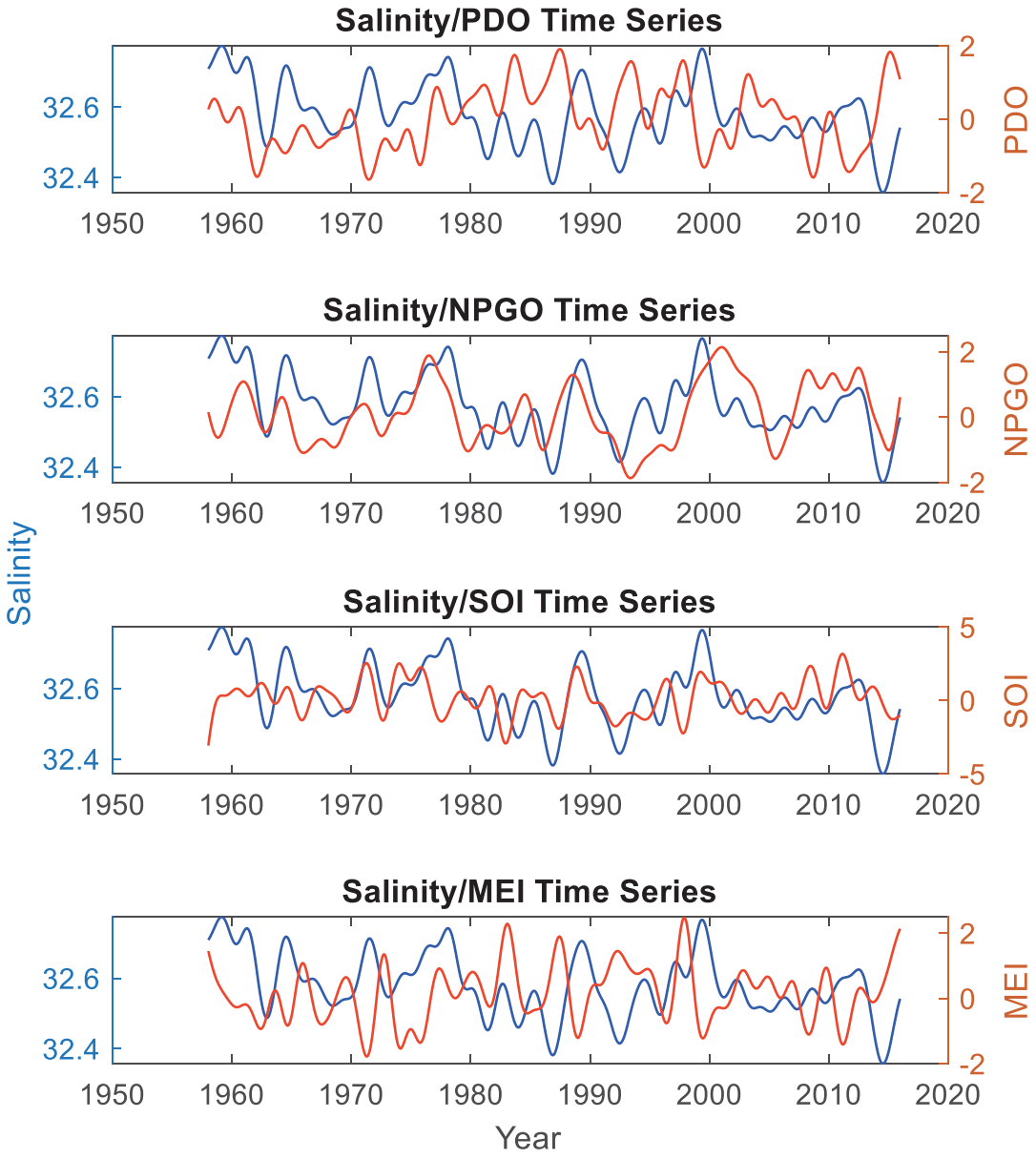

Figure 3.10: Overplots of salinity and the climate modes time series. Salinity is depicted in blue while the climate modes are in red. Units are as previously established.

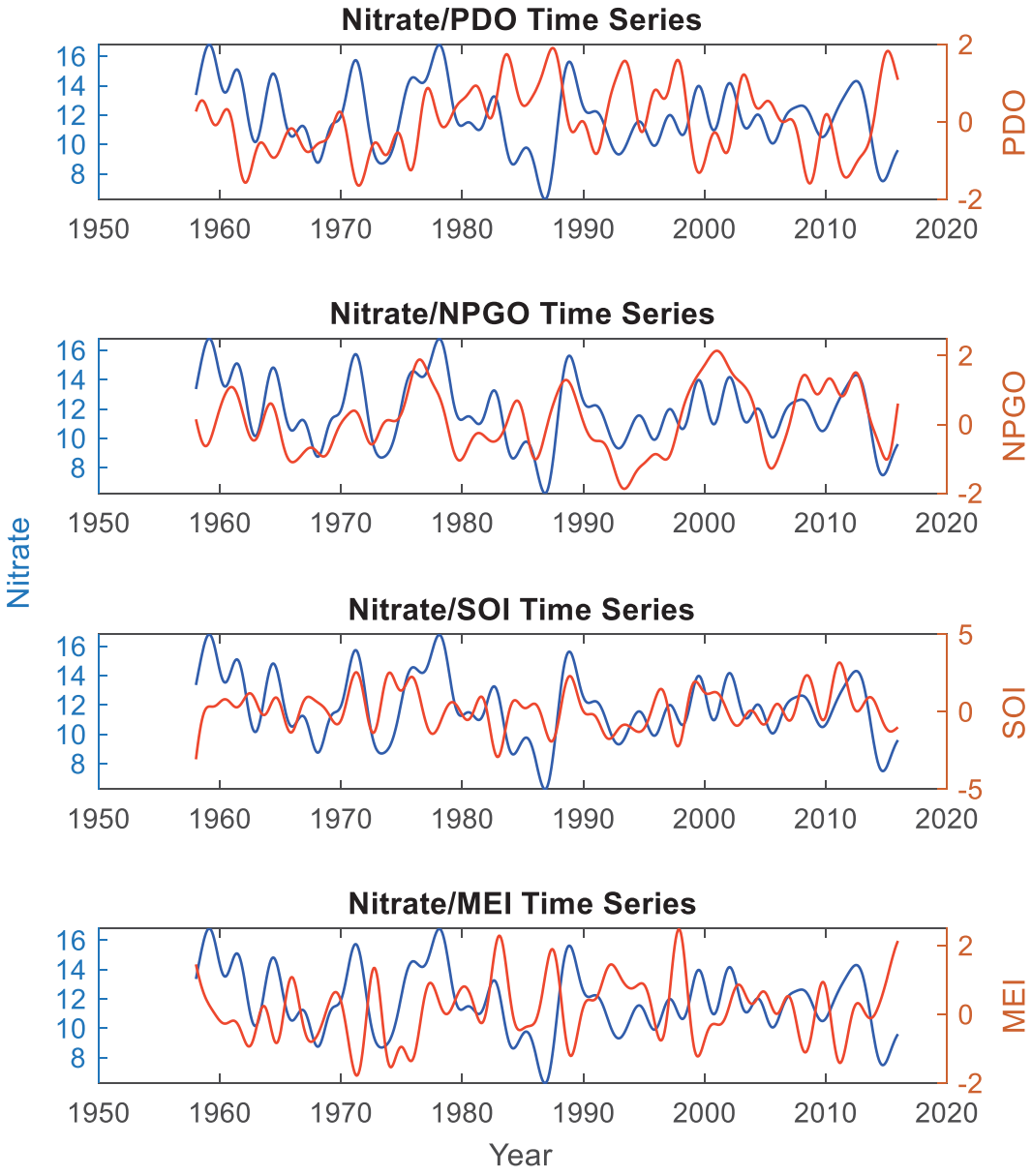

Figure 3.11: Overplots of nitrate and the climate modes time series. Nitrate is depicted in blue while the climate modes are in red. Units are as previously established.

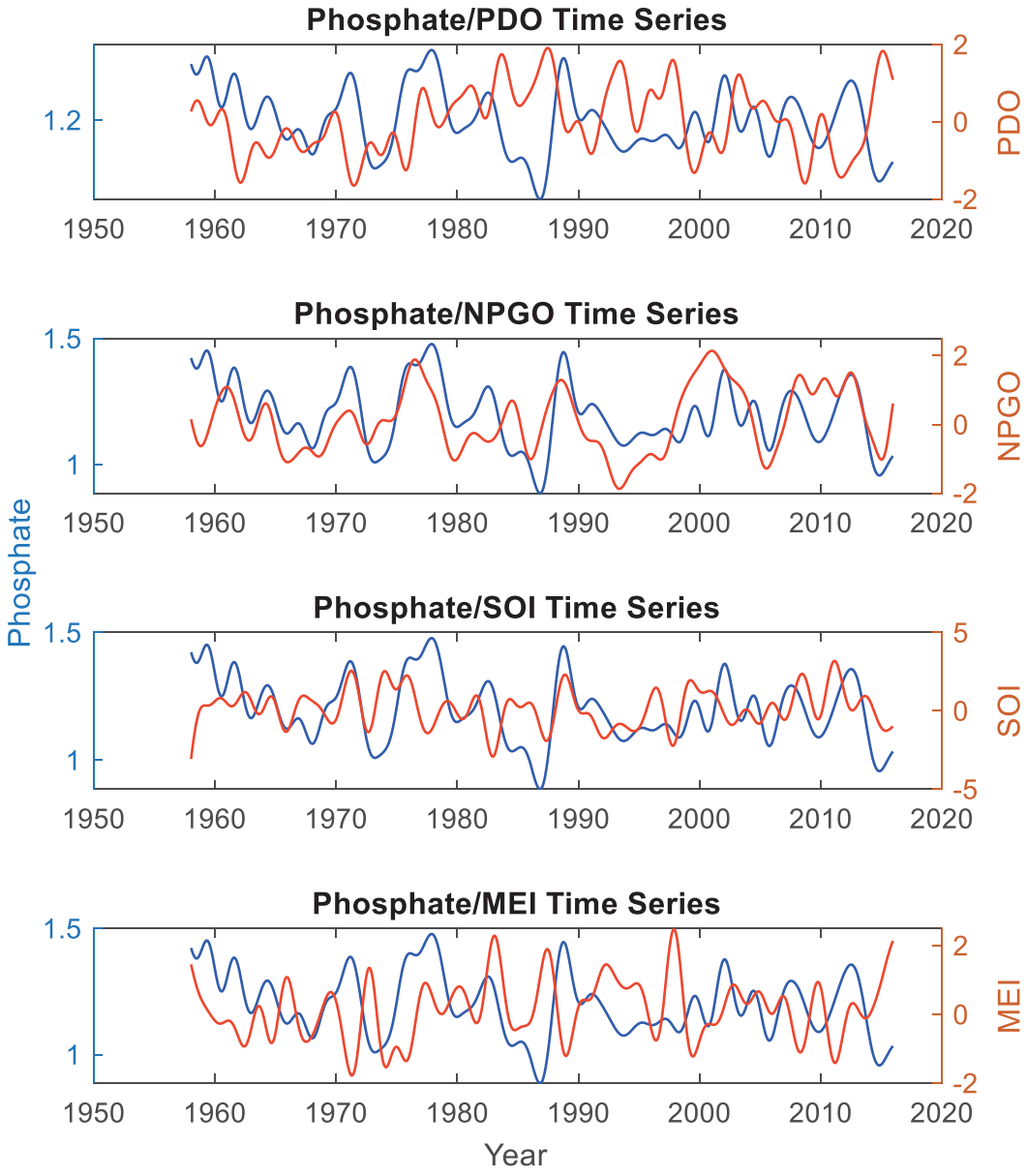

Figure 3.12: Overplots of phosphate and the climate modes time series. Phosphate is depicted in blue while the climate modes are in red. Units are as previously established.

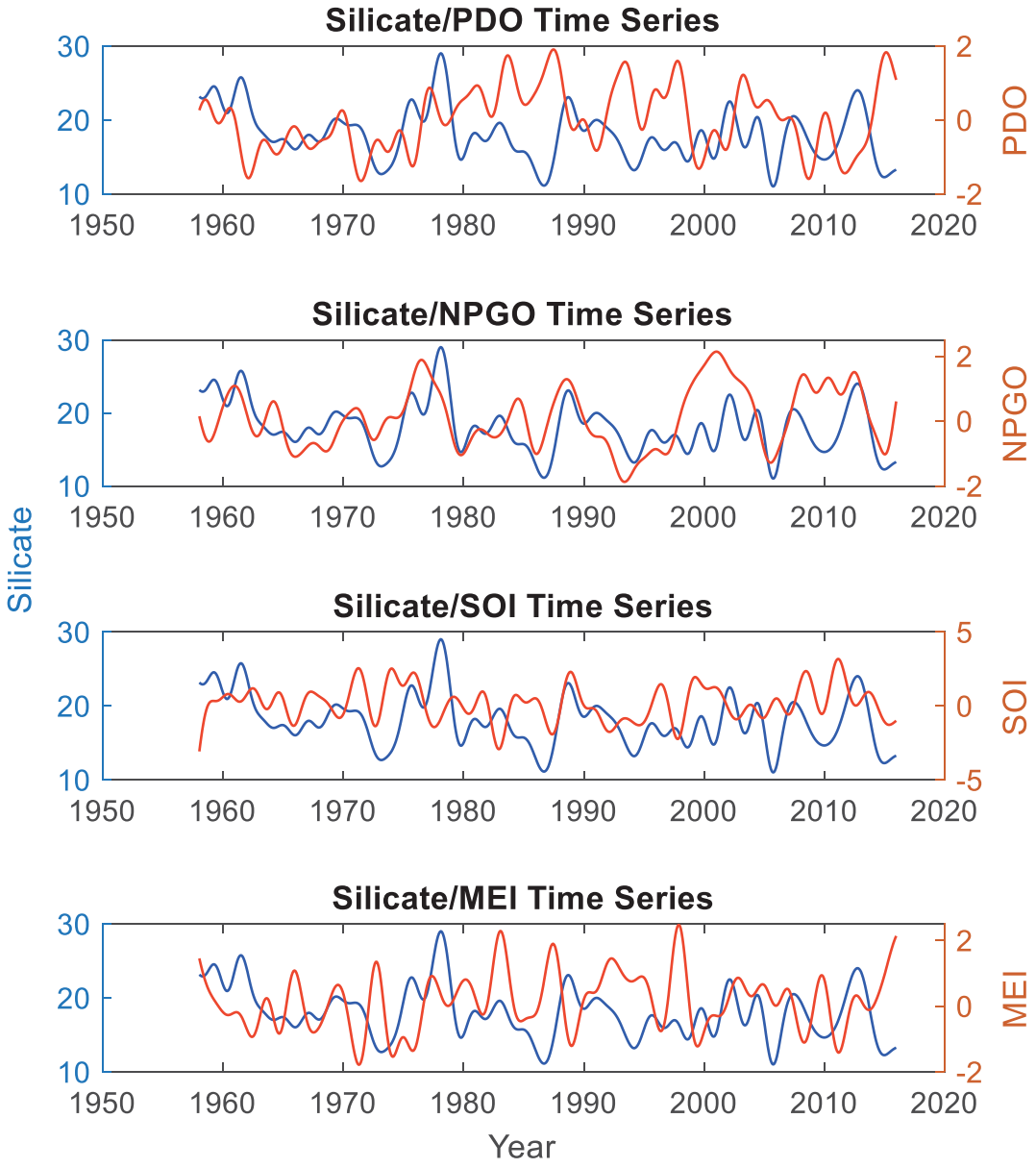

Figure 3.13: Overplots of silicate and the climate modes time series. Silicate is depicted in blue while the climate modes are in red. Units are as previously established.

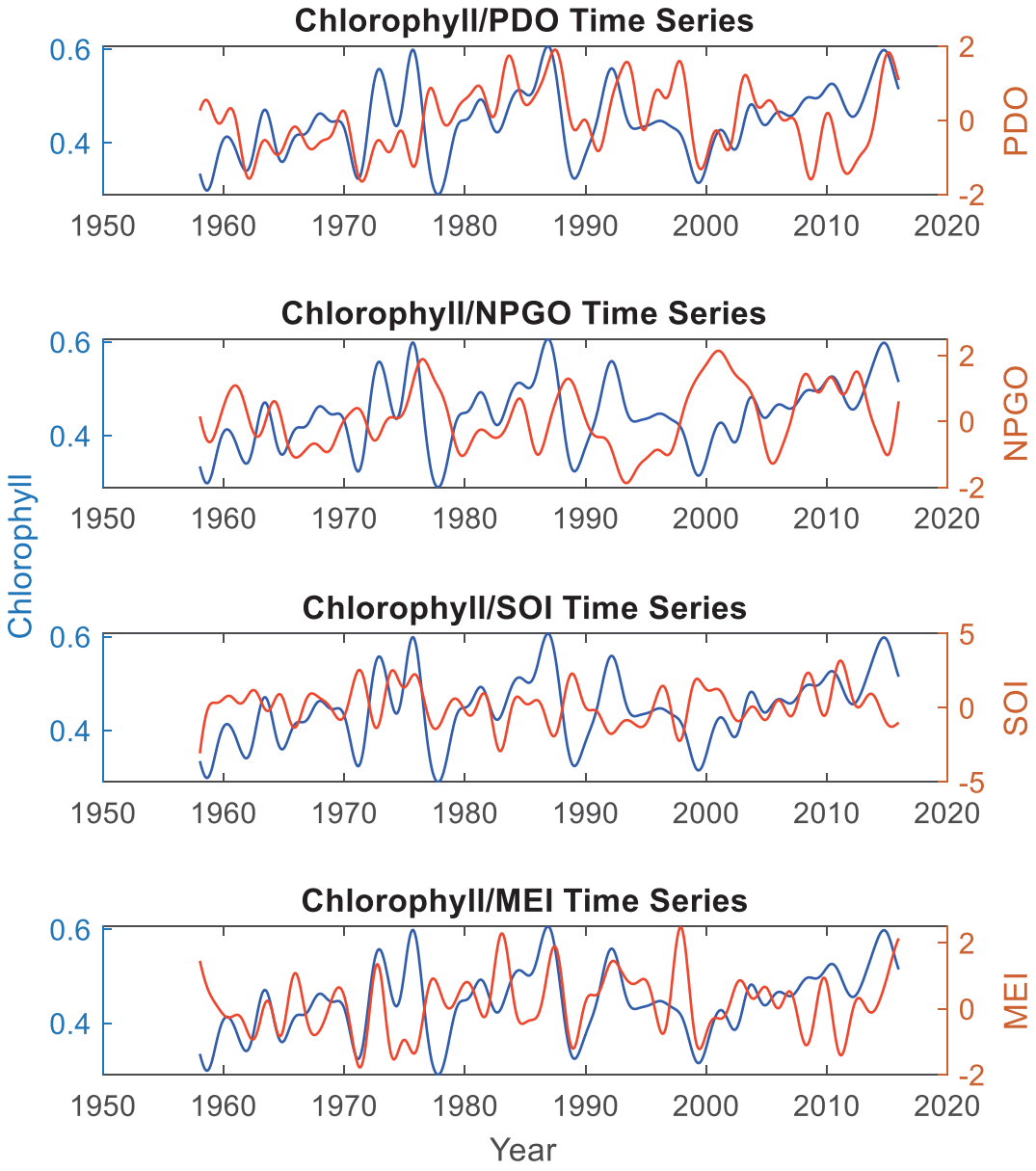

Figure 3.14: Overplots of chlorophyll and the climate modes time series. Chlorophyll is depicted in blue while the climate modes are in red. Units are as previously established.

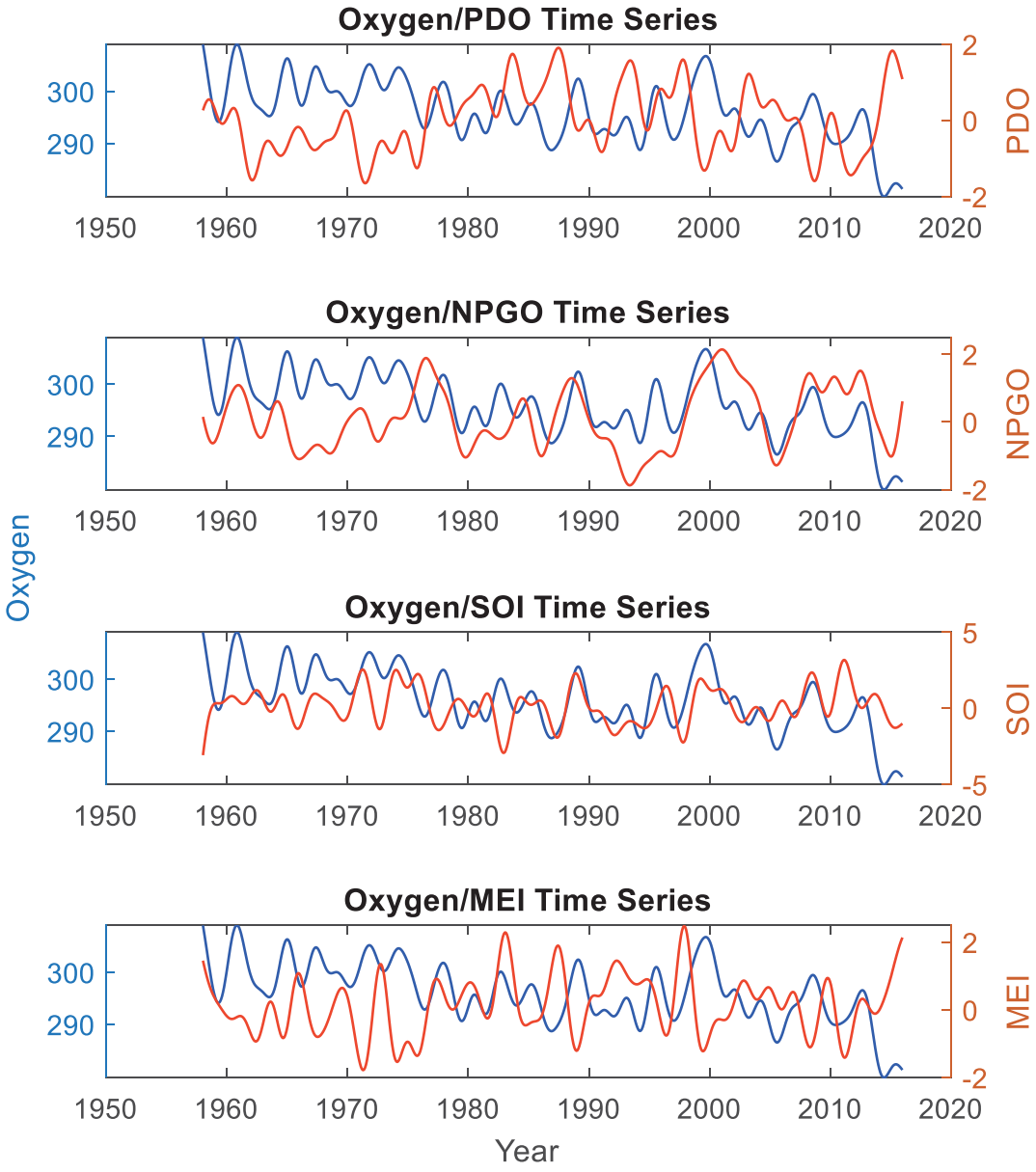

Figure 3.15: Overplots of oxygen and the climate modes time series. Oxygen is depicted in blue while the climate modes are in red. Units are as previously established.

3.4 Determining Lagged Correlations Between Ocean State Variables and Climate Indices

3.4.1 Lagged Correlations

The next step was to determine the amount of lag (or lack thereof) between the ocean variables and climate modes. Cross-correlations were completed to determine the time lag in which the correlation between the two variables was the strongest, and cross-spectral analyses were completed to determine the strength of the signal relationships at the frequencies of interest, as well as the lag associated with those relationships.

The cross-correlation analyses yield a number which indicates the time lag (in months) of the strongest correlation between the two variables. Since we postulate that the climate variables correspond to the forcing, and the ocean state corresponds to the response, we need to define the lag/lead convention. A negative lag number indicates that the first variable of the pair (always an ocean state variable) is ahead of the second variable (i.e. leads), while a positive number indicates that the first variable is the lagging variable (response). We thus anticipate positive or zero lags. Due to the interconnectedness of the climate system, a negative lag (ocean leads climate) may also be possible, but for the purposes of this study we restricted the positive outcomes to climate leads the ocean. A lag period of 63 months was set to be the limit of influence for this study as the auto-correlations of the climate modes (shown in Figure 3.16) show that the longest amount of time necessary for decorrelation amongst the climate modes is for PDO, which takes 63 months before it becomes effectively zero (using a 90% confidence interval calculated using the square root of the inverse error function for each value (Giles, 2011)). Therefore, 63 months may be thought of as the decorrelation or 'memory' timescale for the system. Using a statistical significance cut-off was attempted, but due to

the large sample size the lags would have to more than 600 months for there to be a statistical non-significance, which is unrealistic for this situation. This cut-off point means that, for this study, any strong correlations (or signals for the cross-spectral analyses) that have a lag greater than 63 months are likely not the result of one of the climate modes or any direct forcing between the variables.

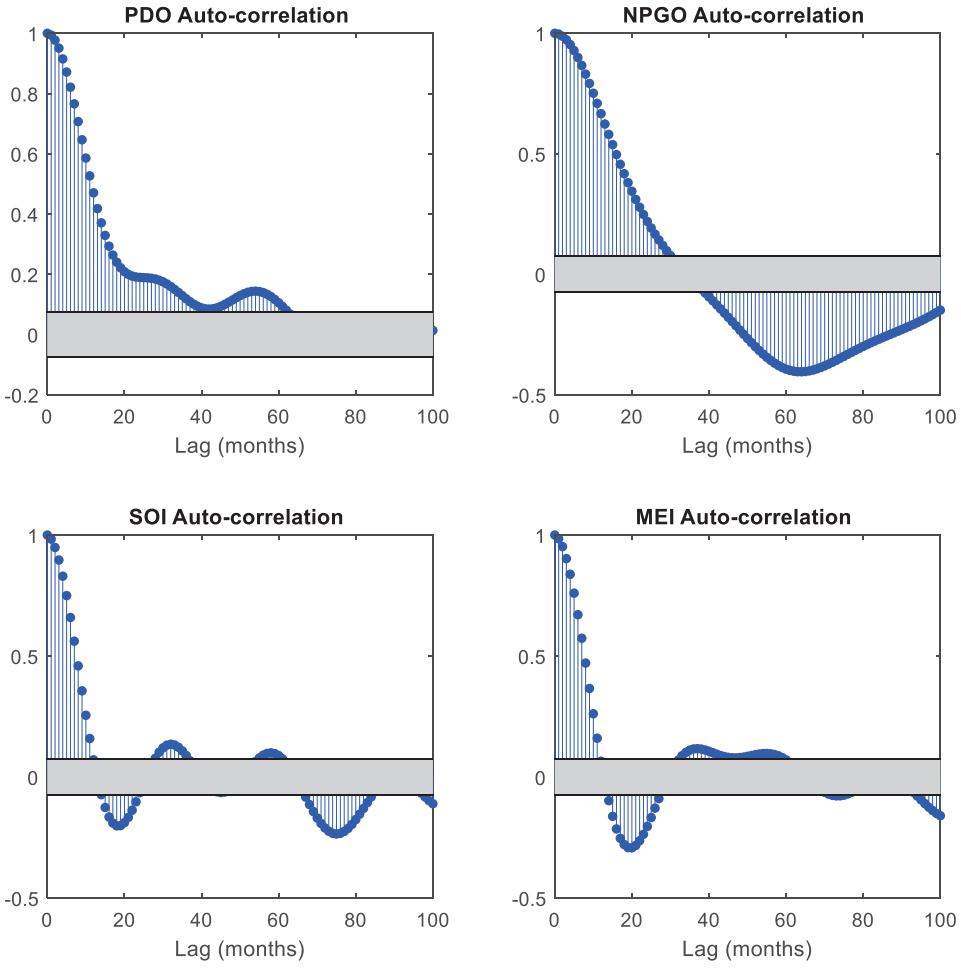

*Figure 3.16: The sample auto-correlations for the filtered climate modes. The blue dots indicate the auto-correlation while the blue lines help visualise the corresponding lag. The shaded grey area indicates the 95% CI in which auto-correlation values are not significant. The confidence bounds are based on an asymptotic normal distribution of 0.5*log(1+R)(1-R), with an approximate variance equal to 1/(N-3) where R is the correlation.*

Hence, when looking at the cross-correlation and cross-spectral relationships between the ocean variables and the climate indices, if the ocean variable is the leading variable in both cross-correlation and cross-spectral analyses, that relationship will not be considered important since this study is looking at how changes in the climate modes can influence changes in the surface ocean, and not the other way around.

Table 3.3: The strongest correlations and their respective lag from the results of cross-correlations for the lag values that are within 18-month cut off and are positive. The ocean state variables are represented by their first letter (and second for silicate)

Variable Pairing	Cross-Correlations (absolute)	Lag (months)
T/NPGO	0.58	0
T/SOI	0.43	2
S/NPGO	0.54	3
S/SOI	0.32	2
S/MEI	0.31	0
N/NPGO	0.58	4
N/MEI	0.27	0
P/NPGO	0.58	5
P/SOI	0.24	31
P/MEI	0.21	0
Si/NPGO	0.54	6
O/NPGO	0.46	2

Given that the relationships between the ocean state variables and the climate modes are the subject of this study, these are the relationships that need to be focused on. Temperature has a moderately strong correlation with NPGO and SOI while being in phase (NPGO) or lagging by 2 months (SOI) (Table 3.3). Salinity has correlations with similar strengths for NPGO and SOI as well, but with slightly longer lags (i.e. salinity takes longer to respond), and has a slightly weaker correlation with MEI, which it is in phase with. Nitrate has a moderately strong correlation with NPGO, with a lag of 4

month, and a correlation of about half that with MEI, which is in phase. Phosphate also has moderately high correlations with NPGO and a lag (response time) of 5 months, but has moderately weak correlations with SOI and MEI, with the respective lags being 31 and 0 months. Silicate has a moderately strong correlation with NPGO with a lag of 6 months. Chlorophyll does not have any relationships where there is a positive (or zero) lag not exceeding 63 months. Oxygen has a moderately strong correlation with NPGO, and takes 2 months to respond.

Based on these values, no ocean state variable responds to all four of the climate modes, but NPGO is most likely to be the mode that drives the ocean response. The corresponding lag values for NPGO indicate that there can be 0 to 6 months before the ocean state variable responds to the climate mode. The SOI forces three of the variables with lags of 2 and 31 months. The MEI only forces three variables, which are in phase, and tends to have weakest correlations. The PDO does not appear to lead any of the ocean state variables.

3.4.2 Spectral Lag

Cross-spectral analyses were also conducted to determine the lagged relations between the variables as a function of frequency, and so, in some sense, is a generalization of the cross-correlation. The figures below (Figures 3.17-3.29) depict the cross-spectral analyses of the ocean state/climate mode variable pairs which had a positive cross-correlation lag that was less than 63 months, as well as the temperature/PDO pairing since that is a pairing with a known relationship where the

climate mode leads the ocean variable (all other variable pair figures are in Appendix C (Figures C.8-C.35)). The coherency spectrum indicates how strongly the two variables are related at the same frequencies, and given that the frequencies (cycles/year) of interest are 0.04688, 0.09375, 0.1875 for all climate modes, as well as 0.2344, and 0.2812 for the MEI, these will be focused on in of this study. The coherencies do not indicate much other than whether or not the two time series share cycles at similar frequencies, but it should be noted that the expectation that PDO influences temperature (mentioned in section 3.2.1) is not depicted in their coherency spectrum (Figure 3.29). The phase spectrum determines how out of phase the two time series are within the frequency range of interest (i.e. the time lag at a given frequency) where a negative phase indicates the climate mode drives the ocean state variable.

When looking at the cross-spectral analyses for the highlighted pairings in Table 3.3, (Figures 3.17-3.29), there are a few things to note. In Figure 3.17 there is a strong coherency between phosphate/MEI (panel C) in the 4- to 5-year cycle frequency, and the phase (panel B) jumps from one extreme to the other from one frequency to the next (phase wrapping). This is indicative of the two variables being anti-correlated. This is also the case for nitrate/MEI and salinity MEI, (Figures 3.18, 3.19). The variables which are in phase but anti-correlated are adjusted by 180° to obtain more realistic lags. There is more coherency for the phosphate/SOI pairing (Figure 3.20) at the 10 to 20-year cycle frequencies than at the 5-year cycle, but according to Table 3.4 phosphate also responds to SOI at the 5-year cycle. The same can be said for salinity/SOI (Figure 3.21). Figure 3.22 (temperature/SOI) has phase wrapping, and stronger coherencies at the longer cycles. Oxygen/NPGO, nitrate/NPGO, and salinity/NPGO (Figures 3.23, 3.26, 3.27) do

not have any phase wrapping, and their phases are all negative, meaning that the ocean state variables respond to NPGO. Silicate/NPGO, phosphate/NPGO, and temperature/NPGO (Figures 3.24, 3.25, 3.28) do show phase wrapping, so they are anti-correlated The temperature/PDO pairing (Figure 3.29), which did have a strong correlation, but did not have temperature responding to PDO in the cross-correlation results, also does not have a negative phase, indicating that temperature does not respond to PDO.

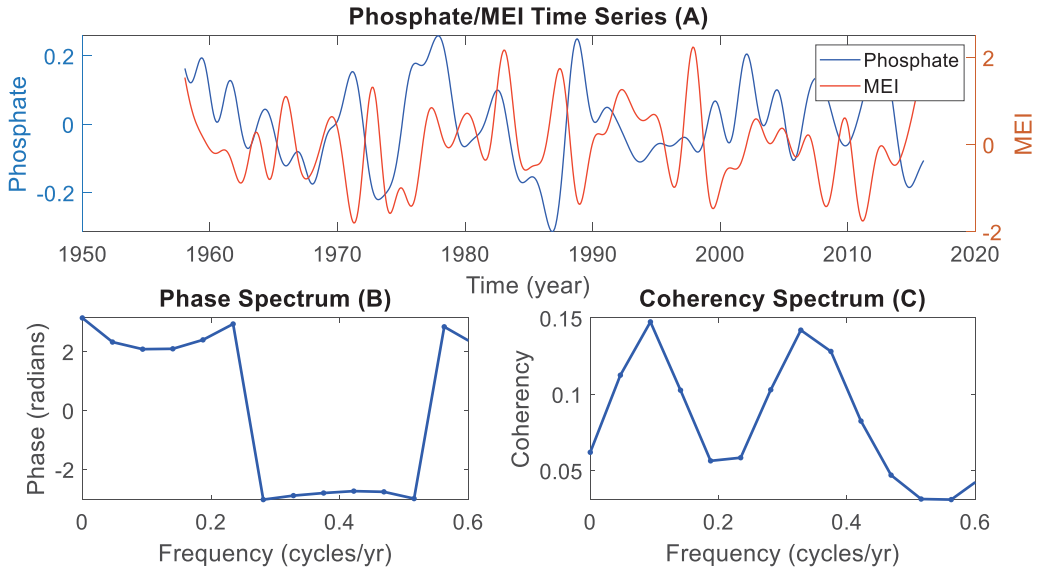

Figure 3.17: An overplot of the scaled and smoothed silicate and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

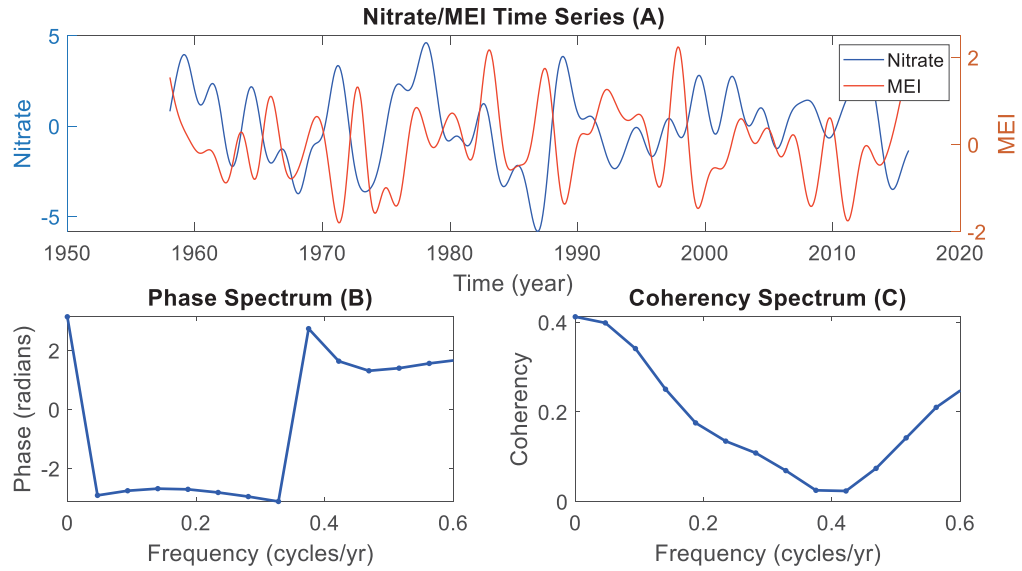

Figure 3.18: An overplot of the scaled and smoothed nitrate and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

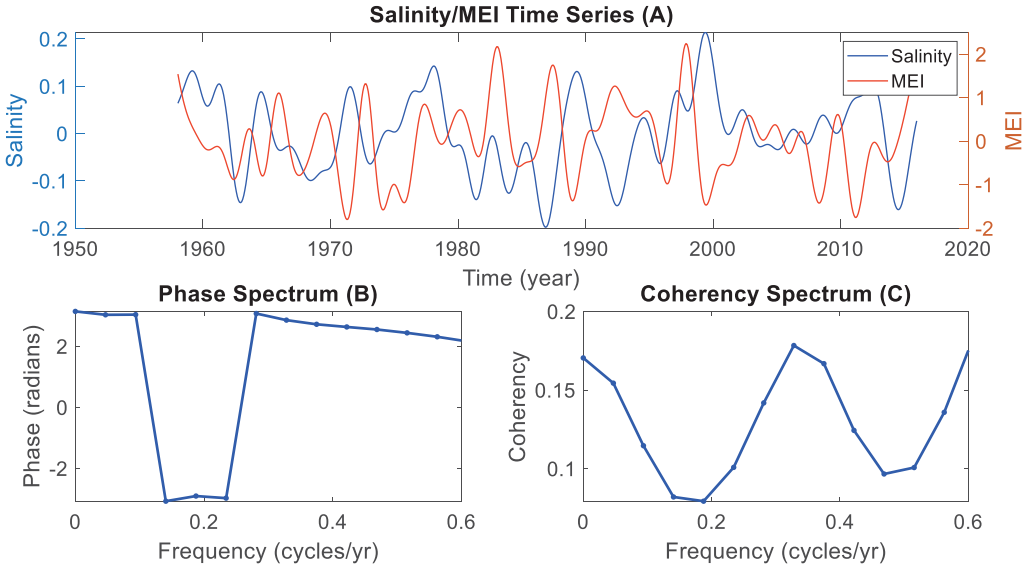

Figure 3.19: An overplot of the scaled and smoothed salinity and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

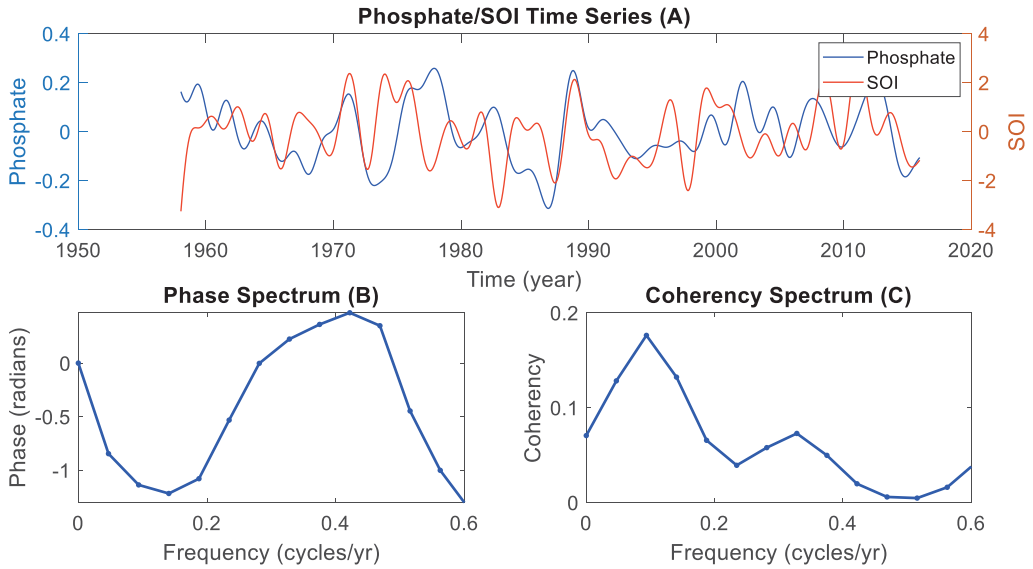

Figure 3.20: An overplot of the scaled and smoothed phosphate and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

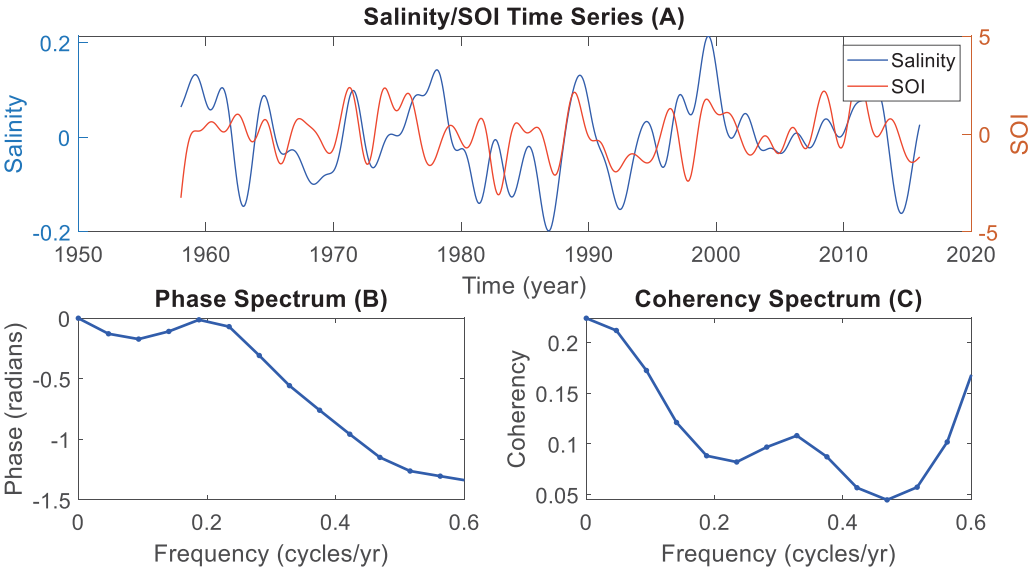

Figure 3.21: An overplot of the scaled and smoothed salinity and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

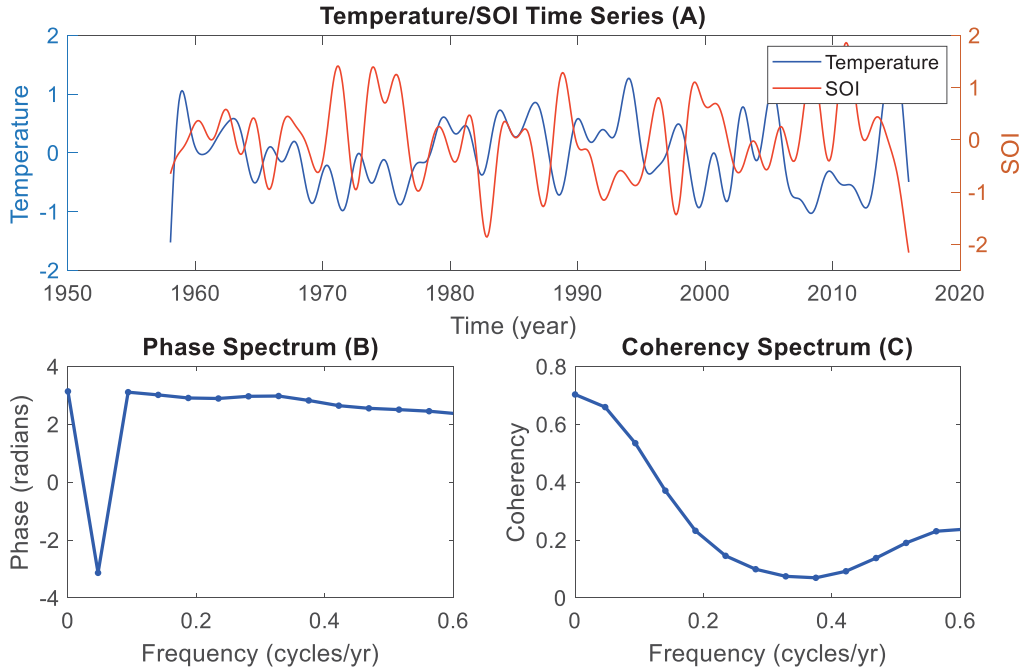

Figure 3.22: An overplot of the scaled and smoothed temperature and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

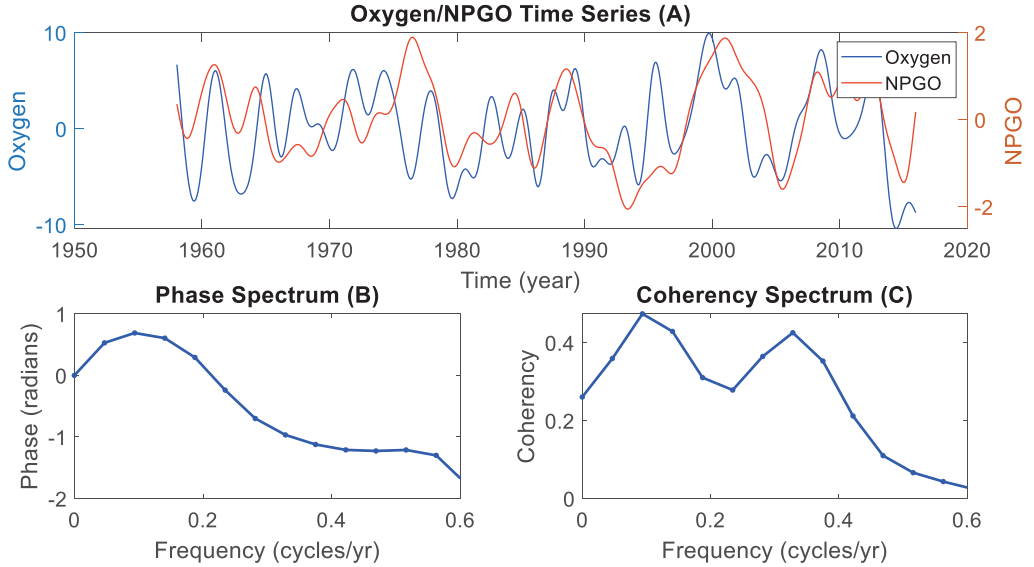

Figure 3.23: An overplot of the scaled and smoothed oxygen and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

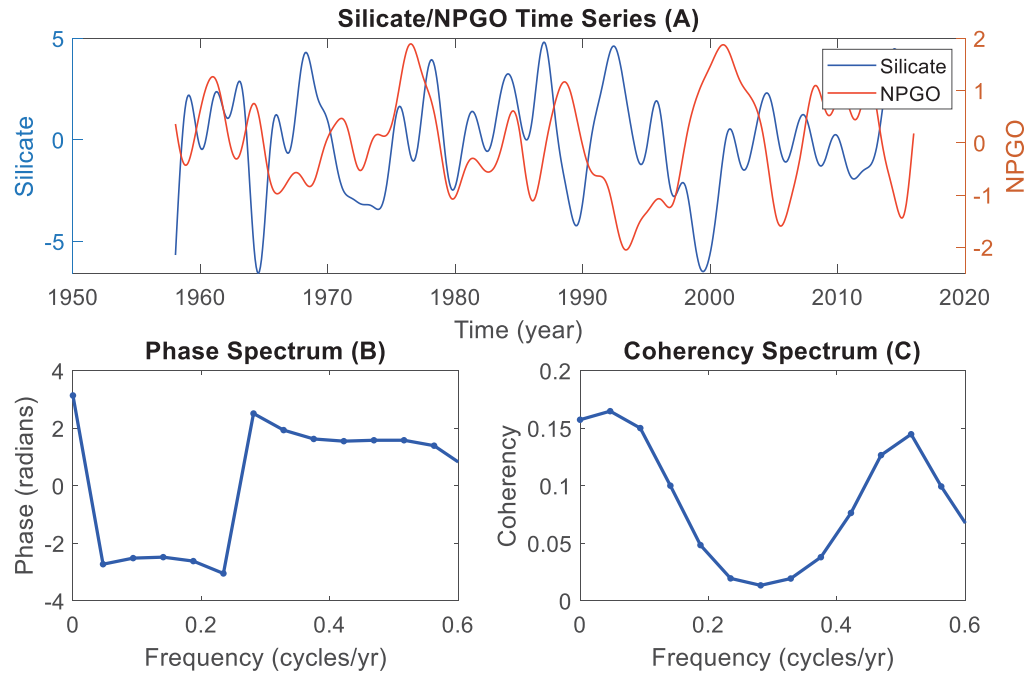

Figure 3.24: An overplot of the scaled and smoothed silicate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

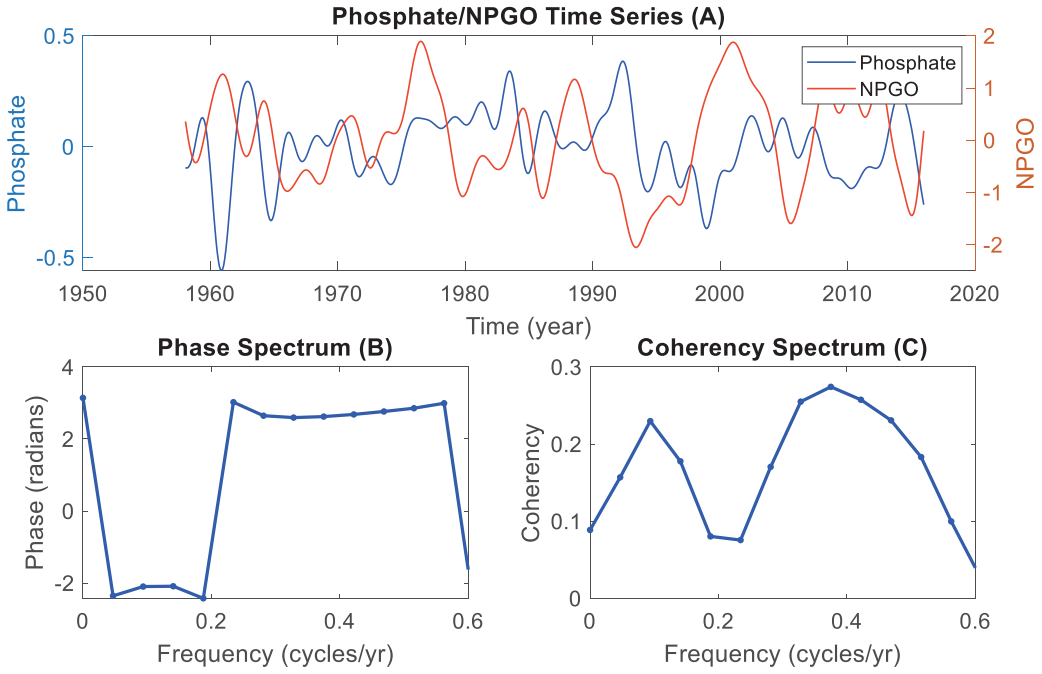

Figure 3.25: An overplot of the scaled and smoothed phosphate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

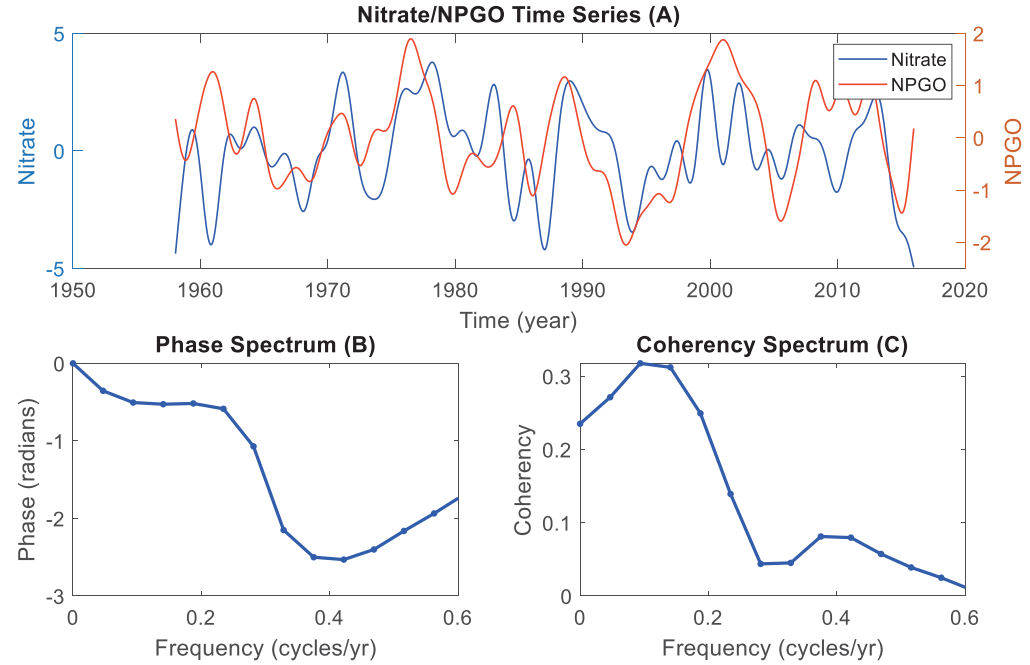

Figure 3.26: An overplot of the scaled and smoothed nitrate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

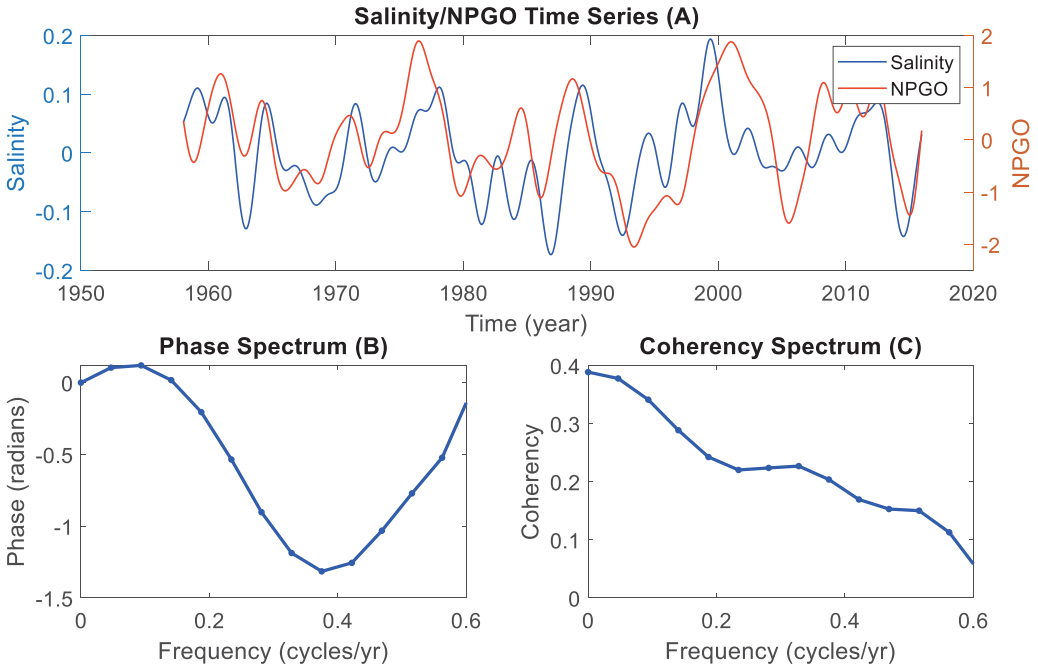

Figure 3.27: An overplot of the scaled and smoothed salinity and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

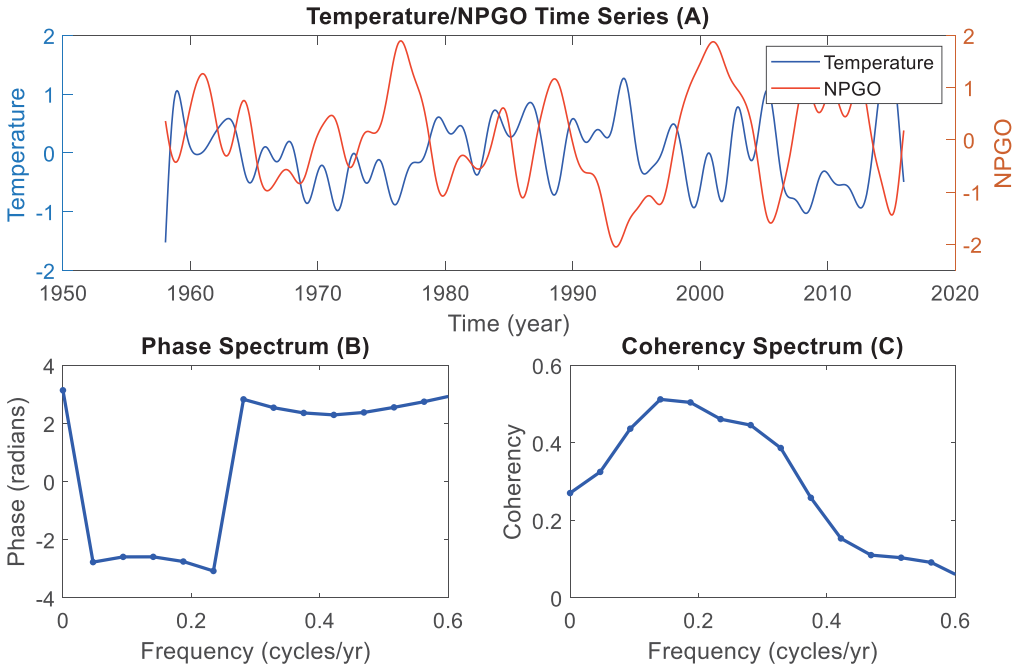

Figure 3.28:An overplot of the scaled and smoothed temperature and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

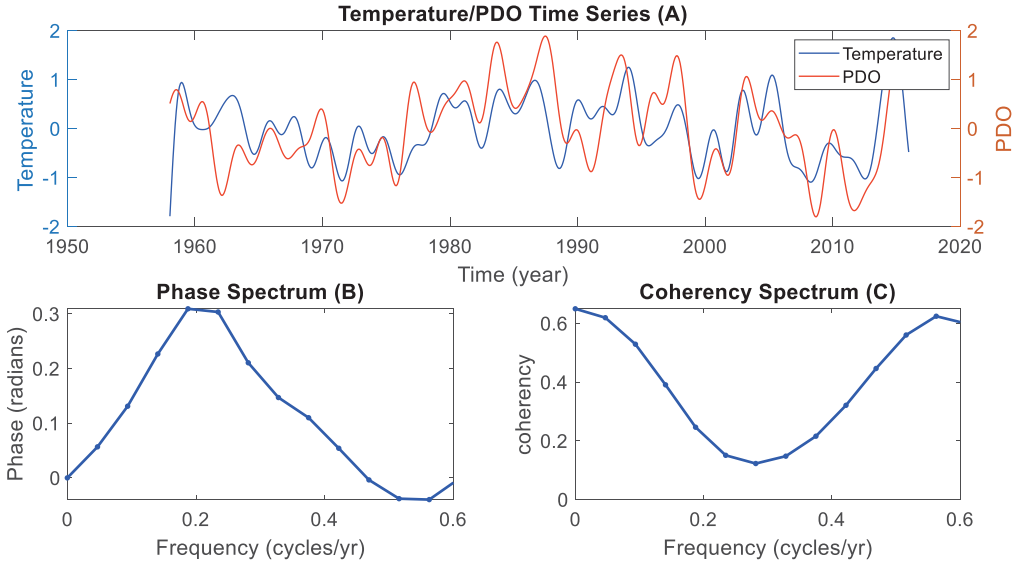

Figure 3.29: An overplot of the scaled and smoothed temperature and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

With this spectral analysis, having a negative phase means that the climate mode is the leading factor in the relationship. That being said, and keeping the same 63-month cut off that was used in the correlation analyses, the highlighted values in Table 3.4 are the values that meet all of the criteria for the frequencies of interest (or are nearly in phase (lag<1 month)).

To turn the phase spectrum into a lag of months, the following equation was used:

$$Lag = \frac{(\phi * \frac{180}{\pi})}{360} * \frac{1}{\omega} * 12 \qquad (5)$$

where:

- ϕ is the corresponding phase value at frequency ω in radians
- ω is the frequency of interest in cycles per year.

The cross-spectral analyses show that fifteen of the pairings meet the criteria for at least three of the frequencies of interest (Table 3.4). Of these fifteen pairs, ten have the same result as the cross-correlation lag analyses, and the oxygen/NPGO and temperature/NPGO cross-correlation result is not seen in the cross-spectral analysis. The five pairs that have a negative lag that is no more than 63 months are: chlorophyll/NPGO, silicate/SOI, nitrate/SOI, silicate/MEI, and temperature/MEI. Given that the silicate/PDO cross-spectral analysis meets the criteria in the 10- and 20-year cycle frequencies, which are the two strongest frequencies for PDO, the it would be safe to say that PDO leads silicate. The highlighted cross-spectral analysis pairings which are not in Table 3.3 might differ because these (cross-spectral) lags might not occur at the highest cross-correlation,

which is how the lags in Table 3.3 were calculated. This could also be the case for chlorophyll/MEI, where only the 10-year cycle meets the criteria even though NPGO's dominant cycles are 3 and 4 years.

Out of NPGO, SOI, and MEI, NPGO has, on average, the shortest lags, and it had the strongest cross-correlation, which indicates that NPGO has a more significant impact on the ocean state variables than the other climate modes. There is no obvious difference between SOI and MEI (which makes sense as they represent the same climate mode), and PDO has the weakest impact on the ocean state variables. The differences in the lags between the cross-correlation analysis and cross-spectral analysis could be accounted for by the fact that the cross-correlation analyses are based on the point of strongest correlation between the two variables, and the cross-spectral analyses are based on specific frequencies of the variable's cycles.

The lack of negative lags within the 63-month period in the cross-spectral results when there were lags in the cross-correlation results could be due to extra "noise" in the time series that was not filtered out with the low pass filtered used. It could also be due to the climate modes having a range of important frequencies (e.g. ENSO has a 3 to 5-year cycle) and the analysis had difficulty comparing those frequencies to the ocean state frequencies.

Table 3.4: The resulting lag values in months from the cross-spectral analyses and Eq. 5. Ocean state variables are depicted by their first letter (and second for silicate). Negative values (first variable lags the second) that fall within the 18 month criteria are highlighted. '–' denotes the pairings that do not have dominant cycles in the frequencies indicated.

	Period 21.33 years	**Period 10.66 years**	**Period 5.33 years**	**Period 4.27 years**	**Period 3.55 years**
T/PDO	2.30	2.67	3.19	-	-
S/PDO	8.12	8.99	9.02	-	-
N/PDO	0.09	2.66	7.41	-	-
P/PDO	-6.04	61.13	5.36	-	-
Si/PDO	-11.89	-8.23	3.85	-	-
C/PDO	38.23	-42.45	10.87	-	-
O/PDO	7.65	56.89	5.46	-	-
T/NPGO	14.09	10.43	3.50	-	-
S/NPGO	-3.57	-1.90	-2.25	-	-
N/NPGO	-9.50	-7.30	-4.13	-	-
P/NPGO	-14.06	-9.78	-4.26	-	-
Si/NPGO	-16.22	-11.08	-5.19	-	-
C/NPGO	-7.99	-7.07	-6.10	-	-
O/NPGO	19.81	13.01	1.18	-	-
T/SOI	0.48	-0.23	-1.45	-	-
S/SOI	-5.25	-3.52	-0.13	-	-
N/SOI	-24.91	-17.87	-7.15	-	-
P/SOI	-34.30	-23.12	-11.11	-	-
Si/SOI	-40.93	-25.83	-13.48	-	-
C/SOI	16.32	11.39	-21.71	-	-
O/SOI	6.12	5.77	4.02	-	-
T/MEI	-0.60	-0.97	-1.06	-0.22	0.72
S/MEI	-4.54	-2.16	2.39	1.35	-0.47
N/MEI	-25.27	-17.24	-4.34	-0.62	0.32
P/MEI	-33.46	-21.65	-7.79	-1.90	0.66
Si/MEI	-40.63	-24.31	-8.82	-0.04	3.77
C/MEI	9.75	-0.95	10.41	6.73	3.95
O/MEI	8.43	8.11	5.24	1.73	1.92

Chapter 4 –Discussion

Long-term ocean monitoring time series can be very useful and informative, but also difficult to deal with. Their use in determining how climate modes and the ocean surface interact with each other can be helpful for increasing our knowledge of the ocean and how it is influenced by climate modes. The issue with these long-term time series is that they are often incomplete, having variable sampling rates and gaps which change with the variable under consideration. Therefore standard time series analyses cannot be used on them, and it is difficult to compare them to climate mode time series. Filling in the gaps with the Kalman smoother algorithm allowed the ocean state variables to be on a regular time grid, and relationships established with the climate modes to determine if climate modes influence the ocean state variables. It was found that while all the climate modes displayed some amount of influence on the ocean state, NPGO had the greatest amount of influence.

This thesis analysed seven ocean variables in the 60-year time series from Ocean Station Papa (OSP) in the sub-Arctic Pacific. These data had inconsistent sampling rates and gaps in the data. To deal with this issue, a novel reconstruction approach was applied to the OSP time series. This was based on a Kalman smoother solution for a state space model. It allowed for the infilling or imputation of gaps in the record by making use of the relationships between the ocean state variables, in essence borrowing information from one state variable to find out about another. A key part of this procedure was establishing the relationships between the ocean state variables which was done using lagged multivariate regression. Furthermore, to obtain a monthly record, the temperature and salinity reanalysis product, ORAS4, was used within the statistical reconstruction.

This created a complete time series which could then be analysed using standard time series analyses such as cross-correlation and cross-spectral analyses. As the reconstructions account for both model and observation error, the reconstructed time series do not match up exactly with the observations, but provide reasonable, complete, estimated time series with appropriate error bars. For further analysis, the reconstructed time series were low-pass filtered to get rid of the seasonal cycle and other short-term cycles since the goal here was to compare climate modes.

There are a few assumptions that have been made in this study, as well as some known issues. When looking at the Kalman Smoother, it was assumed that the dynamics matrix Φ, which embodies the time lagged relationships between the ocean variables, does not change over time. Since there were only limited time periods in which all seven of the ocean state variables were sampled at a high enough frequency, it could not be determined if a Φ calculated over a different sampling period would produce the same results as the Φ which was calculated for the training period of 1970 to 1976. Should there be a data set where multiple Φ's could be calculated over different sampling periods, it could be determined if Φ does change over time, and if that would imply significant change to the results. Relatedly, it is therefore also assumed that the sampling period used to calculate Φ was not influenced by any strong climatic event, like the ones being studied. If Φ was calculated using a sampling period during a strong El Niño, for example, it would influence the results, and may explain why some results were not as expected. Another issue with the Kalman Smoother is that the model error was calculated using its ability to predict the observational data, which means that there is some

observational error incorporated into the model error. It is currently unknown if there is a way to get around this issue.

This method of statistical reconstruction can homogenise irregular, cyclical data sets, and the dynamics matrix used in the reconstruction (Φ) can also provide some information on the mechanisms of the system. Since Φ is asymmetrical, the values within the matrix can be used to examine the physical and biogeochemical mechanisms within the system, which may be explored in future studies. As well, future studies may be able to incorporate the uncertainty estimates included in the Kalman smoother. The time series analyses used in this study do not typically have methods of incorporating uncertainty values, but it may be possible to do so by using weighted values based on the inverse of error in the analyses (i.e. values from the Kalman smoother with high uncertainty estimates have less weight than the values with low uncertainty estimates). That being said, should this method of statistical reconstruction be used for non-time series analyses, the uncertainty estimates from the Kalman smoother should be able to be used.

Some of the limitations of this method of statistical reconstruction include: there needs to be at least one period of time where all the variables have been sampled at a large enough rate to establish a reliable Φ; and when there are limited observations, the reconstruction may be dominated by the variability and trends seen in the predictor variables (temperature and salinity for this study). The results of the second limitation may mitigated by de-seasonalising the data before it goes through the Kalman smoother, and may negate the need for a low-pass filter and any sources of error that may be associated with it. To do this, the first thing to do would be to put the raw data on a monthly grid so the distribution of the data would be the same as it was in Section 2.3.

Then would be to remove the monthly means from the raw data to obtain the anomalies. Next, take the training data period, interpolate it, and use multivariate regression to create a new Φ. The observation errors should be the same as the seasonalised observation errors since they are based on the standard deviation of the raw data, but the model error should be recalculated with the new Φ. The procedure from there on in would be the exact same as described in Section 2.3. De-seasonalising this process may remove the dominating predictor cycles from the other variable reconstructions, but it may also increase the model error as the Kalman smoother is ideal for reconstruction cycles and it may struggle to capture the lower-frequency cycles. While it is currently unknown how the seasonalised/de-seasonalised results compare with each other, the best method to use in future projects may vary depending on the project.

Application of this method of statistical reconstruction could potentially be used on any ocean time series. It could aid in projects aimed at studying the seasonal and interannual variability of ocean variables by providing detailed information that may otherwise be unavailable (Henson, 2014), whether due to a lack of resources, or one (or more) variables being measured less frequently than others. Further study of the Kalman smoother on ocean time series such as HOT and BATS, where there are more frequent, regular sampling of ocean variables could provide information for refining this method, and determine the extent of its capabilities.

Four climate modes, PDO, NPGO, SOI, and MEI, were analysed together with the seven low-pass filtered ocean state variables. Correlations, cross-correlations, and cross-spectral analyses were completed to determine if and how the climate modes influenced the ocean state variables. When looking at the lag of the strongest correlation

in the cross-correlation analyses, NPGO had the greatest number of positive lags which were 63 months or less (six out of seven) and the corresponding correlations ranged from 0.46 to 0.58. SOI and MEI had three correlations each, ranging from 0.21 to 0.43, and PDO did not have any non-zero lagged correlations which met the criteria.

The cross-spectral analyses resulted in each of the four climate modes leading at least one ocean state variable with a lag that met the criteria. According to this analysis, PDO leads nitrate, phosphate, silicate, and chlorophyll with lags ranging from 0.09 to 42.45 month; NPGO leads salinity, nitrate, phosphate, silicate, and chlorophyll with lags ranging from 2.25 to 16.22 months; SOI leads temperature, salinity, nitrate, phosphate, and silicate with lags ranging from 0.13 to 40.93 months; and MEI temperature, salinity, nitrate, phosphate, and silicate with lags ranging from 0.22 to 40.63 months). There were some discrepancies between the cross-correlation and cross-spectral analyses that could be due to spurious relationships, that the dominant climate mode frequencies were not isolated in the time series, or that the 'dominant' cycle in the climate modes is often within a range of frequencies and does not provide accurate results when only looking at two or three of those frequencies together as is done in the cross-correlation.

Based on these results, it appears that NPGO has the greatest influence among the seven ocean state variables in this study. Many of the previous studies done on NPGO and ocean surface biogeochemical variables have been done by Di Lorenzo and others, which is not surprising given that Di Lorenzo was the one who initially defined the NPGO, and it was initially defined with data that included data collected at OSP. In these studies, it was found that there is a correlation of 0.4 between the NPGO index and SSS anomalies at OSP (Figure 4.1) (Di Lorenzo et al., 2009). It needs to be noted, though, that

these observations were SSS anomalies made from using a depth of 150m while this study is looking at the mean SSS of the top 15m. Studies in ocean carbon often avoid the surface ocean because the signals are less clear in the surface due to all the physical, biological, and chemical processes taking place there. The correlation between NPGO and SSS anomalies is attributed to variations in the low frequencies, not the seasonal cycle (Di Lorenzo et al., 2009).

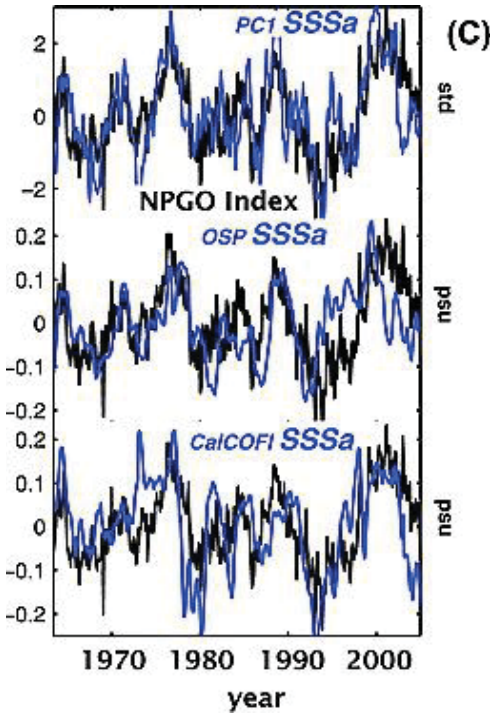

Figure 4.1: Figure 2 from Di Lorenzo et al., 2008. Figure description: Timeseries of NPGO index (black) compared to PC1 of SSSa (R = 0.67, 99%), observed SSSa at ocean Station Papa (OSP) (R = 0.40, 96%) at the offshore end of Line P [Crawford et al., 2007], and observed SSSa from CalCOFI program (R = 0.56, 99%). The PC1 of SSSa is normalized by its standard deviation, units are in standard deviations (std).

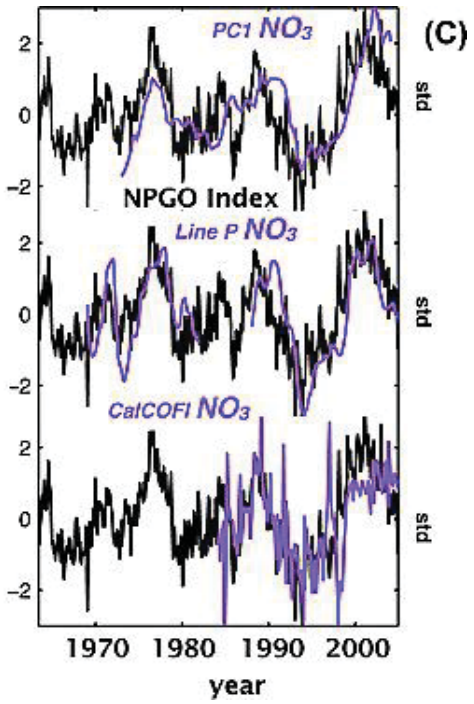

Figure 4.2: Figure 3 from Di Lorenzo et al., 2008. Figure Desciprtion: Timeseries of NPGO index (black) compared to PC1 of NO3 (R=0.65, 99%), observed mix layer NO3 at Line-P [Pena and Varela, 2007] (R=0.68, 99%), and observed NO3 from CalCOFI program (R=0.51, 95%). All time series are normalized by their standard deviations, units are in standard deviations (std).

Similar to the SSS, nitrate levels at OSP were also correlated to the NPGO index at Line P when looking at nitrate values 150m (Figure 4.2). Given that it has been found that low-frequencies variations in surface nitrate levels are strongly, positively correlated with phosphate, silicate, salinity, and negatively correlated with oxygen (Wong et al., 2007, Whitney et al., 2007), correlations between nitrate and NPGO were to be expected. That expectation can be seen in the completed analyses where NPGO had a strong correlation with nitrate and leads with a lag within the 63 month criteria.

The PDO appears to have the least amount of influence among the ocean state variables. Previous studies suggest that PDO changes in SST are asymmetrical between winter and summer in the NEP (Gregg & Conkright, 2002), and PDO has an east/west gradient in temperature (Overland et al., 1999) (Figure 4.3) this could mean that there would not be a strong correlation between PDO and temperature at OSP even though

PDO index is based on SST anomalies. Figure 4.3 shows that OSP (C) is located in a region close to the null, and only experiences a small degree of SST variation during PDO events. Some correlation could be expected between chlorophyll and PDO as Gregg & Conkright (2002) found some evidence that chlorophyll decreased when the PDO entered a positive regime. In terms of SSS, it was found that there was a NNW/SSE character (Overland, et al., 1999), which is similar to the interannual variability in precipitation, and given that PDO's climatic effects are most visible in the North American/North Pacific region (Hare & Mantua, 2000). The results show that PDO only leads nitrate, phosphate, silicate, and chlorophyll during the 10-20 year cycles, and has no leading cross-correlations.

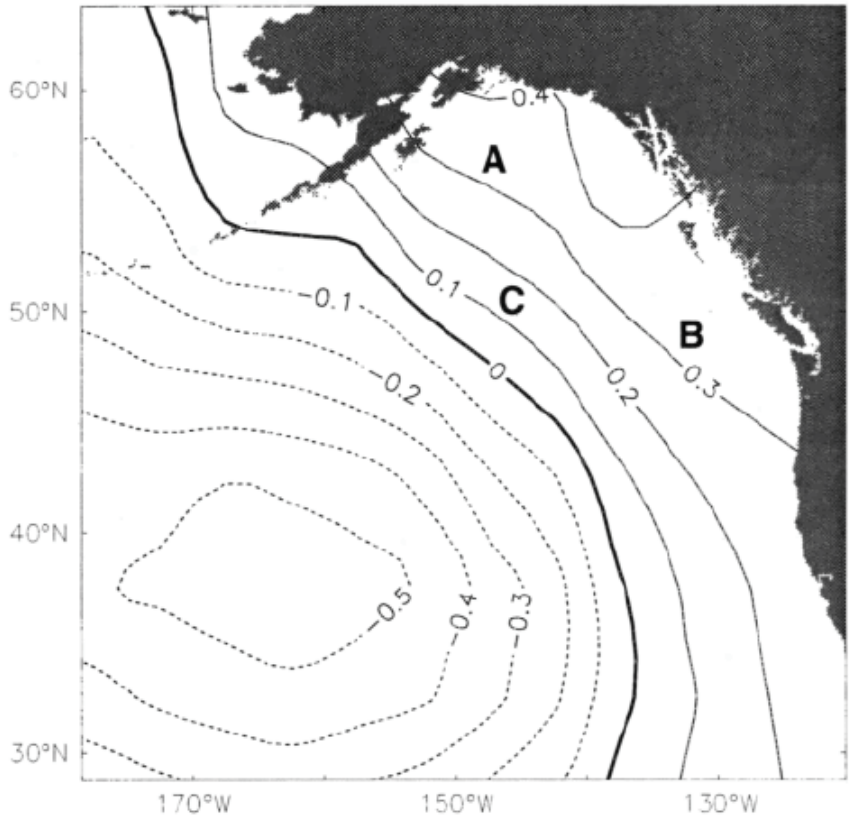

Figure 4.3: Figure 1 from Overland et al., 1999. Figure description: Spatial pattern of SST regressed upon the October-March PDO index, plotted in °C per standard deviation of the PDO index. Letters denote three station locations: A = GAK11 (57.1°N, 148°W), B = Papa line station 7 (49.1°N, 132.4°W), and C = OWS P (50°N, l45°W).

The SOI appeared to have a slightly stronger influence on the ocean state variables than MEI given that the represent the same climatic event. The SOI was expected to be negatively correlated with SST as a strong, negative SOI is indicative of El Niño, which results in warm waters in the sub-Arctic Pacific (Niebauer, 1988). Since the MEI is calculated using a greater number of variables than SOI, it should also be negatively correlated to SST, but its correlation would not be as strong. Nitrate and Silicate levels are known to drop during El Niño events in the northern Pacific Ocean (Whitney & Welch, 2002), so SOI and MEI should also be negatively correlated with these variables. The SOI had significant correlations with temperature, salinity, and phosphate while MEI had significant correlations with salinity, nitrate, and phosphate. Both SOI and MEI have a leading lag with temperature, salinity, nitrate, phosphate, and silicate with respect to the cross-spectral analyses.

Given that when the NPGO is positive the changes in the wind create favourable up-welling conditions in the Alaskan Gyre, it makes sense that NPGO has the greatest influence on most of the ocean state variables at OSP (Di Lorenzo et al., 2008). The positive NPGO index coincides with an increase in nutrients in the surface waters at OSP due to nutrients being brought up to the surface waters from the nutrient-rich waters below. This also provides an explanation to the lags observed between NPGO and the ocean state variables as the nutrient-rich water from the Alaskan Gyre up-welling needs time to be transported to OSP from the centre of the Gyre via the Alaska current. The lack of influence from the PDO may be the result of OSP being located near the PDO node (Figure 4.3), where the effects of the PDO are not as strongly observed.

The SOI and MEI should not be expected to have as strong of a influence on the ocean state variables at OSP as NPGO given that both indices are recorded in the tropical Pacific, whereas NPGO (and PDO) is recorded in the North Pacific, closer to where OSP is located. The lags seen in the SOI with the ocean state variables may be partly explained by the air temperature/ocean temperature gradient (i.e. how long it takes for SST to reflect changes in air temperature) (Lau & Nath, 2001). This would also help explain the lack of lags seen in some of the MEI/ocean state variable relationships because one of the parameters MEI is calculated by is SST, which would negate the air temperature/ocean temperature lag.

It needs to be noted, however, that all of the climate modes are intertwined and feed/force each other in some capacity. As the climate modes will not be separated and independent from each other in this study, the results will not always allow for a direct relationship between the biogeochemical variables and the climate modes. Even if a climate mode were as isolated as it could be, there are still connections between the climate, atmosphere, and ocean that are unknown, so there would still be extra influences not accounted for in the study which could result in a mis-interpretation of the results.

Should this work be continued in the future, there a few topics that have been identified and could be addressed. One of the original goals of this project under which this work was funded was to help determine how climatic events affect fish recruitment, so if there are any fish recruitment data that can be added to the reconstruction, it should be possible to produce a time series similar to those in this project and relate fish recruitment to climate modes. As well, incorporating and analysing the recent moorings into the data sources used may allow some of the assumptions surrounding Φ to be tested.

Even if all of the variables are not available from the more recent mooring data, creating a Φ using the variables available from the mooring data and then using those available variables to create a Φ during the sampling period used in this project should determine if Φ does change over time, or if a climatic event significantly influenced Φ. As well, the comparison between different data sets of the same variable could be tested with a cross-validation exercise, similar to the one used to determine the model errors. While this project used monthly data for the observations used in the reconstruction and to create the Φ matrix (the data were there and freely available), using two to four samples per year should also provide enough information to obtain a reasonable statistical reconstruction, and may negate the need to filter out the annual cycle, though the annual anomalies could be used as well. It should also be noted that if there is a poor Φ value for a pair of variables and there is low model error for those variables, the reconstruction may provide results that have visibly different trends or cycles during periods with no observations when compared to periods with observations. Finally, this work is not limited to OSP. This method of reconstruction could be used on any ocean station in the open ocean, and maybe even on some coastal data sources if a few modifications are made. As long as there is a correlation among the variables selected, a statistical reconstruction, like the Kalman Smoother, could provide reliable time series for data sets that are incomplete.

Bibliography

Aumont, O., and L. Bopp. "Globalizing Results from Ocean in Situ Iron Fertilization Studies." *Global Biogeochemical Cycles*, vol. 20, no. 2, (2006), doi:10.1029/2005gb002591.

Balmesda, M. A., K. Mogensen, and A. T. Weaver. "Evaluation of the ECMWF ocean reanalysis system ORAS4." *Quarterly Journal of the Royal Meteorological Society*, vol. 139, no. 674, (2013), pp. 1132-1161.

Bingham, F.M., and R. Lukas. "Seasonal cycles of temperature, salinity, and dissolved oxygen observed in the Hawaii Ocean Time-series." *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 43, no. 2-3, (1996), pp. 199-213.

Bond, N. A., Cronin, M. F., Freeland, H., &Mantua, N. (2015). "Causes and impacts of the 2014 warm anomaly in the NE Pacific." *Geophysical Research Letters*, vol. 42, no. 9, (2015), pp. 3414-3420.

Cao, L, and K. Caldeira. "Can Ocean Iron Fertilization Mitigate Ocean Acidification?" *Climatic Change*, vol. 99, no. 1-2, (2010), pp. 303–311., doi:10.1007/s10584-010-9799-4.

Cavole, L.M., A.M. Demko, R.E. Diner, A. Giddings, I. Koester, C.M.L.S. Pagniello, M. Paulsen, A. Ramirez-Valdez, S.M. Schwenck, N.K. Yen, M.E. Zill, and P.J.S. Franks. "Warm-Water Anomaly in the Northeast Pacific: Winners, Losers, and the Future." *Oceanography*, vol. 29, no. 2, (2016), pp. 273-285.

Chhak, K.C., E. Di Lorenzo, N. Schneider, P.F. Cummins. "Forcing of Low-Frequency Ocean Variability in the Northeast Pacific." *Journal of Climate*, vol. 22, no. 5, (2009), pp. 1255-1276.

D'Asaro, E.A., J. Thomson, A.Y. Shcherbina, R.R. Harcourt, M.F. Cronin, M.A. Hemer, and B. Box-Kemper. "Quantifying upper ocean turbulence driven by surface winds." *Geophysical Research Letters*, vol. 41, no. 1, (2013).

Di Lorenzo, E. "NPGO index monthly averages from Jan-1950 to Jul-2019." (2019). http://www.o3d.org/npgo/npgo.php

Di Lorenzo, E., Flechter, J., Schneider, N., Bracco, A., Miller, A. J., Franks, P. J. S., Bograd, S. J., Moore, A. M., Thomas, A. C., Crawford, W., and Peña, A. "Nutrient and salinity decadal variations in the central and eastern North Pacific." *Geophysical Research Letters*, vol. 36, no. 14, (2009).

Di Lorenzo, E., N. Schneider, K. M. Cobb, P. J. S. Franks, K. Chhak, A. J. Miller, J. C. McWilliams, S. J. Bograd, H. Arango, E. Curchiste, T. M. Powell, P. Rivière. "North Pacific Gyre Oscillation links ocean climate and ecosystem change." *Geophysical Research Letters*, vol. 35, no. 8, (2008).

Di Lorenzo, E., K.M. Cobb, J.C. Furtado, N. Schneider, B.T. Anderson, A. Bracco, M.A. Alexander, and D.J. Vimont. "Central Pacific El Nino and decadal climate change in the North Pacific Ocean." *Nature Geoscience*, vol. 3, (2010), pp. 762-765.

Diogou, N., D.M. Palacios, J.A. Nystuen, E. Papathanassiou, S. Katsanevakis, and H. Klinck. "Sperm whale (*Physeter macrocephalus*) acoustic ecology at Ocean Station PAPA in the Gulf of Alaska – Part 2: Oceanographic drivers of interannual variability." *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 150, (2019).

Fasullo, J. T., B. L. Otto-Bliesner, and S. Stevenson. "ENSO's Changing Influence on Temperature, Precipitation, and Wildfire in a Warming Climate". *Geophysical Research Letters*, vol. 45, no. 17, (2018), pp. 9216-9225., https://doi.org/10.1029/2018GL079022

Fassbender, A. J., C. L. Sabine, and M. F. Cronin. "Net Community Production and Calcification from 7 Years of NOAA Station Papa Mooring Measurements." *Global Biogeochemical Cycles*, vol. 30, no. 2, (2016), pp. 250–267., doi:10.1002/2015gb005205.

Felis, T., A. Suzuki, H. Kuhnert, M. Dima, G. Lohmann, and H. Kawahata. "Subtropical Coral Reveals Abrupt Early-Twentieth-Century Freshening in the Western North Pacific Ocean." *Geology*, vol. 37, no. 6, (2009), pp. 527–530., doi:10.1130/g25581a.1.

Freeland, H. "A short history of Ocean Station Papa and Line P". *Progress in Oceanography*, vol. 75 no. 2, (2007), pp. 120-125.

Freeland, H. “Evidence of Change in the Winter Mixed Layer in the Northeast Pacific Ocean: a Problem Revisited.” *Atmosphere-Ocean*, vol. 51, no. 1, (2013), pp. 126-133.

Freeland, H., K. Denman, C. S. Wong, F. Whitney, and R. Jacques. “Evidence of change in the winter mixed layer in the Northeast Pacific Ocean”. *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 44 no. 12, (1997), pp. 2117-2129.

Frost, B. W. “Grazing Control of Phytoplankton Stock in the Open Subarctic Pacific Ocean: a Model Assessing the Role of Mesozooplankton, Particularly the Large Calanoid Copepods Neocalanus Spp.” *Marine Ecology Progress Series*, vol. 39, no. 1, (1987), pp. 49–68., doi:10.3354/meps039049.

Garrett, J. “Station PAPA.” *The Canadian Encyclopedia*, Historic Canada, (2006), www.thecanadianencyclopedia.ca/en/article/station-papa/

Giles, M.B. “Approximating the erfinv function.” *GPU Computing Gems*, vol. 2, (2011).

Gould J., M. Kienast, M. Dowd, E. Schefub. “An open-ocean assessment of alkenone deltaD as a paleo-salinity proxy”. *Geochimica et Cosmochimica Acta*, vol. 246, (2019), pp. 246:478-497.

Gregg, W. W., and M. E. Conkright. “Decadal changes in global ocean chlorophyll.” *Geophysical Research Letters*, vol. 29, no. 15, (2002), pp. 20-1-20-4.

Hare, S. R., and N. J. Mantua. “Empirical evidence for the North Pacific regime shifts in 1997 and 1989.” *Progress in Oceanography,* vol. 47, no. 2-4, (2000), pp. 103-145.

Haug, G. H., D. M. Sigman, R. Tiedemann, T. F. Pedersen, and M. Sarnthein. “Onset of Permanent Stratification in the Subarctic Pacific Ocean.” *Nature News*, vol. 401, (1999).

Henson, S.A. “Slow science: the value of long ocean biogeochemistry records.” *Philos Trans A Math Phys Eng Sci*, vol. 372, no. 2025, (2014).

Houk, P., A. Yalon, S. Maxin, C. Starsinic, A. McInnis, M. Gouezo, Y. Golbuu, and R. van Woesik. “Predicting coral-reef futures from El Nino and Pacific Decadal Oscillation events.” *Scientific Reports*, vol 10, no. 7735, (2020).

Joh, Y., and E. Di Lorenzo. “Increasing Coupling Between NPGO and PDO Leads to Prolonged Marine Heatwaves in the Northeast Pacific.” *JGR Biogeosciences*, vol. 44, no. 22, (2017), pp. 11,663-11,671, doi: 10.1002/2017GL075930.

Johnson, R.A., and D.W. Wichern. *Applied Multivariate Statistical Analysis*. 6th ed., Prentice Hall, 2007.

Jones, P.D., D. E. Parker, T. J. Osborn, and K. R. Briffa. “Global and hemispheric temperature anomalies—land and marine instrumental records.” *Trends: A Compendium of Data on Global Change*, (2016). doi: 10.3334/CDIAC/cli.002

Kaiser, Michael J., M. J. Attrill, P. J. leB Williams, S. Jennings, D. N. Thomas, D. K. A. Barnes. *Marine Ecology: Processes, Systems, and Impacts.* 1st Ed., Oxford University Press, 2005.

Koch, F., M.A. Marcoval, C. Panzeca, K.W. Bruland, and S.A. Sañudo-Wilhelmy. “The Effect of Vitamin B12on Phytoplankton Growth and Community Structure in the Gulf of Alaska.” *Limnology and Oceanography*, vol. 56, no. 3, (2011), pp. 1023–1034., doi:10.4319/lo.2011.56.3.1023.

Krause, J.W., M.W. Lomas, and D.M. Nelson. ”Biogenic silica at the Bermuda Atlantic Time-series Study site in the Sargasso Sea: Temporal changes and their inferred controls based on a 15-year record.” *Global Biogeochemical Cycles*, vol. 23, no. 3, (2009).

Lau, N-C. and M.J. Nath. “Impact of ENSO on SST Variability in the North Pacific and North Atlantic: Seasonal Dependence and Role of Extratropical Sea-Air Coupling.” *Journal of Climate*, vol. 14, no. 13, (2001), pp. 2846-2866.

Li Yi, Daling, L. Zhang, and L. Wu. "On the mechanisms of decadal variability of the North Pacific Gyre Oscillation over the 20th century." *JGR Oceans*, vol. 120, no. 9, (2015).

Linkin, M.E., and S. Nigam. "The North Pacific Oscillation – West Pacific Teleconnection Pattern: Mature – Phase Structure and Winter Impacts." *Journal of Climate*, vol. 29, no. 9, (2008).

MacDonald, G.M., and R.A. Case. "Variations in the Pacific Decadal Oscillation over the past millennium." *Geophysical Research Letters*, vol. 32, no. 8, (2005), doi: 10.1029/2005GL022478.

Mackas, D. L., R. Goldblatt, and A.G. Lewis. INterdecadal variation in developmental timing of Neocalanus plumchrus populations at Ocean Station P in the subarctic North Pacific. *Canadian Journal of Fisheries and Aquatic Sciences,* vol. 55, no. 8, (1998), pp. 1878-1893.

Maldonado, M.T, and N.M. Price. "Utilization of Iron Bound to Strong Organic Ligands by Plankton Communities in the Subarctic Pacific Ocean." *Deep Sea Research Part III*, vol. 46, no. 11-12, (1999).

Malmstrom, R.R., A. Coe, G.C. Kettler, A.C. Martiny, J. Frias-Lopez, E.R. Zinser, and S.W. Chisholm. "Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans." *The ISME Journal*, vol. 4, (2010), pp. 1252-1264.

Mantua, N.J. "PDO Index." (2018). http://research.jisao.washington.edu/pdo/PDO.latest

Mantua, N.J., and S.R. Hare. "The Pacific Decadal Oscillation." *Journal of Oceanography*, vol. 58, no. 1, (2002), pp. 35-44.

Mantua, N.J., S.R. Hare, Y. Zhang, J.M. Wallace, and R.C. Francis. "A Pacific interdecadal climate oscillation with impacts on salmon production." *Bulletin of the American Meteorological Society,* vol. 78, (1997), pp. 1069-1079.

Mazzarella, A., A. Giuliacci, and N. Scafetta. "Quantifying the Multivariate ENSO Index (MEI) coupling to CO_2 concentration and to the length of day variations." *Theoretical and Applied Climatology*, vol. 111, (2013), pp. 601-607.

McCabe, G.J., and M.A. Palecki. “Multidecadal Climate Variability of Global Lands and Oceans.” *International Journal of Climatology,* vol. 26, (2006), pp. 849-865., doi: 10.1002/joc.1289

McPhaden, M.J., S.E. Zebiak, M.H. Glantz. ”ENSO as an Integrating Concept in Earth Science.” *Science*, vol. 314, no. 5806, (2006).

Miller, A.J., F. Chai, S. Chiba, J.R. Moisan, and D.J. Neilson. “Decadal-Scale Climate and Ecosystem Interactions in the North Pacific.” *Journal of Oceanography,* vol. 60, (2004), pp. 163-188.

Newman, M., G.P. Compo, and M.A. Alexander. “ENSO-Forced Variability of the Pacific Decadal Oscillation.” *Journal of Climate*, vol. 16, no. 23, (2003).

Niebauer, H.J. “Effects of El Nino-Southern Oscillation and North Pacific weather patterns on interannual variability in the subarctic Bering Sea.” *Journal of Geophysical Research: Oceans*, vol. 93, no. C5, (1988).

NOAA/NCEP Climate Prediction Center (CPC) SOI monitoring data, (2020).

Overland, J.E., S. Salo, and J.M. Adams. “Salinity signature of the Pacific Decadal Oscillation.” *Geophysical Research Letters*, vol. 29, no. 9, (1999), pp. 1337-1340.

Peña, M.A., and D.E. Varela. “Seasonal and interannual variability in phytoplankton and nutrient dynamics along Line P in the NE subarctic Pacific.” *Progress in Oceanography*, vol. 75, no. 2, (2007), pp. 200-222.

Pitchford, J.W., and J. Brindley. “Iron limitation, grazing pressure and oceanic high nutrient-low chlorophyll (HNLC) regions.” *Journal of Plankton Research,* vol. 21, no. 3, (1999), pp. 525-547.

Power, S., and G. Kociuba. “The impact of global warming on the Southern Oscillation Index”. *Climate Dynamics*, vol. 37, (2010), pp. 1745-1754.

Power, S., F. Tseitkin, V. Mehta, B. Lavery, S. Torok, and N. Holbrook. “Decadal climate variability in Australia during the twentieth century.” *International Journal of Climatology*, vol. 19, no. 2, (1999), pp. 169-184.

Radenac, M-H., J. Jouanno, C. Tchamabi, M. Awo, B. Bourlès, S. Arnault, and O. Aumont. “Physical drivers of the nitrate seasonal variability in the Atlantic cold tongue.” *Biogeosciences*, vol. 17, no. 2, (2019). 10.5194/bg-2019-338.

Rohli, R.V., A.J. Vega. *Climatology* (fourth ed.). (2018). Jones & Bartlett Learning.

Salter, I., A.E.S. Kemp, C.M. Moore, R.S. Lampitt, G.A. Wolff, and J. Holtvoeth. “Diatom Resting Spore Ecology Drives Enhanced Carbon Export from a Naturally Iron-Fertilized Bloom in the Southern Ocean.” *Global Biogeochemical Cycles*, vol. 26, no. 1, (2012), doi:10.1029/2010gb003977.

Santoso, A., M.J. Mcphaden, and W. Cai. “The Defining Characteristics of ENSO Extremes and the Strong 2015/2016 El Nino.” *Reviews of Geophysics*, vol. 55, no. 4, (2017).

Särkkä, S. “Bayesian Filtering and Smoothing.” *Cambridge University Press*, 2013.

Schneider, N., and B.D. Cornuelle. “The Forcing of the Pacific Decadal Oscillation*.” *Journal of Climate*, vol. 18, no. 21, (2005), pp. 4355–4373., doi:10.1175/jcli3527.1.

Schwing, F.B., T. Murphree, L. deWitt, P.M. Green. “The evolution of oceanic and atmospheric anomalies in the northeast Pacific during the El Nino and La Nina events of 1995-2001.” *Progress in Oceanography*, vol. 54, no. 1-4, (2002), pp. 459-491.

Shumway, R.H., and D.S. Stoffer. *Time Series Analysis and Its Applications: with R Examples*. 4th ed., Springer, 2017.

Sieracki, M.E., P.G. Verity, and D. K. Stoecker. “Plankton community response to sequential silicate and nitrate depletion during the 1989 North Atlantic spring bloom.” *Deep Sea Research Part II: Topical Studies in Oceanography,* vol. 40, no. 1-2, (1993), pp. 213-225.

Smith, C. “Overview of Current Ocean Reanalyses.” *Reanalyses.org | Advancing Reanalysis,* (2010), https://reanalyses.org/ocean/overview-current-reanalyses.

Storto, A., and S. Masina. “C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis.” *Earth System Science Data*, vol. 8, (2016), pp.679-696.

Stramma, L., S. Schmidtko, S.J. Bograd, T. Ono, T. Ross, D. Sasano, and F.A. Whitney. “Trends and decadal oscillations of oxygen and nutrients at 50 to 300m depth in the equatorial and North Pacific.” *Biogeosciences,* vol. 17, no. 3, (2020), pp.813-831.

Thomson, R.E., and S. Tabata. “Steric height trends at Ocean Station Papa in the Northeast Pacific Ocean.” *Marine Geodesy*, vol. 11, no. 2-3, (1987), pp. 103-113.

Tabata, S. “Trends and Long-Term Variability of Ocean Properties at Ocean Station P in the Northeast Pacific Ocean.” *Aspects of Climate Variability in the Pacific and the Western Americas*, vol. 55, (1989).

Whitney, F.A. “Anomalous winter winds decrease 2014 transition zone productivity in the NE Pacific.” *Geophysical Research Letters*, vol. 42, no. 2, (2015).

Whitney, F.A., D.W. Crawford, and T. Yoshimura. “The Uptake and Export of Silicon and Nitrogen in HNLC Waters of the NE Pacific Ocean.” *Deep Sea Research Part II: Topical Studies in Oceanography,* vol. 52, no. 7-8, (2005), pp. 1055–1067., doi:10.1016/j.dsr2.2005.02.006.

Whitney, F.A., H.J. Freeland, M. Robert. “Persistently declining oxygen levels in the interior waters of the eastern subarctic Pacific.” *Progress in Oceanography*, vol. 75, no. 2, (2007), pp. 179-199.

Whitney, F.A., C.S. Wong, and P.W. Boyd. “Interannual variability in nitrate supply to surface waters of the Northeast Pacific Ocean”. *Marine Ecology Progress Series,* 170, (1998), pp. 15-23.

Wolter, K. “Multivariate ENSO index (MEI).” *Physical Science Laboratory, NOAA*. (2018). https://www.psl.noaa.gov/enso/mei.old/mei.html.

Wong, C. S. “Carbon Dioxide—A Global Environmental Problem into the Future.” *Marine Pollution Bulletin*, vol. 9, no. 10, (1978), pp. 257–264., doi:10.1016/0025-326x(78)90607-0.

Wong, C. S.., F.A. Whitney, D.W. Crawford, K. Iseki, R. Matear, W.K. Johnson, J.S. Page, and D. Timothy. “Seasonal and Interannual Variability in Particle Fluxes of Carbon, Nitrogen and Silicon from Time Series of Sediment Traps at Ocean Station P, 1982–1993: Relationship to Changes in Subarctic Primary Productivity.” *Deep Sea Research Part II: Topical Studies in Oceanography,* vol. 46, no. 11-12, (1999), pp. 2735–2760., doi:10.1016/s0967-0645(99)00082-x.

Wong, C. S., J.R. Christian, S.-K.E. Wong, J. Page, L. Xie, and S. Johannessen. “Carbon Dioxide in Surface Seawater of the Eastern North Pacific Ocean (Line P), 1973–2005.” *Deep Sea Research Part I: Oceanographic Research Papers*, vol. 57, no. 5, (2010), pp. 687–695., doi:10.1016/j.dsr.2010.02.003.

Wong, C.S., L. Xie, and W.W. Hsieh. “Variations in nutrients, carbon and other hydrographic parameters related to the 1967/1977 and 1988/1989 regime shifts in the sub-arctic Northeast Pacific.” *Progress in Oceanography*, vol. 75, no. 2, (2007), pp. 326-342.

Appendices

Appendix A: Algorithms and Multivariate Regression

A.1: Kalman Smoother Model and Algorithm

Model:

$$X_t = \Phi X_{t-1} + W_t$$
$$Y_t = AX_t + V_t$$

Algorithm:

For t = 1 to T

(i) Prediction Step:

$\hat{X}_{t|t-1} = \Phi \hat{X}_{t-1|t-1}$ Mean of prediction probability density function (pdf) at time t

$M_t = \Phi P_{t-1} \Phi' + \mathrm{Q}$ Variance of predictive pdf at time t

(ii) Observation update:

$P_t = (M_t^{-1} + A' R^{-1} A)^{-1}$ Variance of filter pdf

$K_t = P_t A' R^{-1}$ Kalman gain matrix

$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t(Y_t - A\hat{X}_{t|t-1})$ Filter mean at time t

End (for t)

For more details see Särkkä (2013).

A.2: Multivariate Regression

A more detailed representation of the multivariate regression model used in this thesis is the following.

Let Y be the $n * p$ response matrix, X be an $x * (q - 1)$ matrix such that all entries of the first column are 1's, and q predictors. Let β be an $(q - 1) * p$ matrix of fixed parameters. Let E be an $n * p$ matrix such that E~N (0, Σ).

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ y_{31} & y_{32} & \cdots & y_{3p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1q} \\ 1 & x_{21} & x_{22} & \cdots & x_{2q} \\ 1 & x_{31} & x_{32} & \cdots & x_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{q1} & \beta_{q2} & \cdots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1p} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2p} \\ \epsilon_{31} & \epsilon_{32} & \cdots & \epsilon_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{n1} & \epsilon_{n2} & \cdots & \epsilon_{np} \end{pmatrix}$$

Where:

- Y: response matrix (columns are the ocean state variables at time t)
- X: predictor matrix (columns are the ocean state variables at time t-1)
- β: the quantity estimated by the regression, i.e. the matrix Φ
- E: residual errors.

This regression equation is then solved using standard least squares procedure. See Johnson & Wichern (2007) for more details.

Appendix B: Tables

Table B.1: The complete list of ocean state variable/ climate mode correlations. The r column depicts the correlation between the variables, and the p-value column depicts the p-values for the correlations. Highlighted pairings have p-values less than 0.05, and hence are statistically significant.

Variable Pairing	r	p-value
T/PDO	0.62	<0.001
T/NPGO	-0.47	<0.001
T/SOI	-0.40	<0.001
TMEI	0.45	<0.001
S/PDO	-0.45	<0.01
S/NPGO	0.39	<0.001
S/SOI	0.29	<0.001
S/MEI	-0.34	<0.001
N/PDO	-0.43	0.09
N/NPGO	0.47	<0.01
N/SOI	0.24	<0.01
N/MEI	-0.30	<0.01
P/PDO	-0.39	<0.01
P/NPGO	0.45	<0.01
P/SOI	0.18	<0.01
P/MEI	-0.25	<0.01
Si/PDO	-0.31	<0.01
Si/NPGO	0.37	<0.01
Si/SOI	0.08	0.03
Si/MEI	-0.18	<0.01
C/PDO	0.25	<0.01
C/NPGO	-0.19	<0.01
C/SOI	-0.09	0.02
C/MEI	0.18	0.01
O/PDO	-0.48	<0.01
O/NPGO	0.23	<0.01
O/SOI	0.26	<0.01
O/MEI	-0.39	<0.01

Table B.2: The strongest correlations and their respective lag from the results of cross-correlations for all the variables. Highlighted values are the lag values that are within 18-month cut off and are positive. The ocean state variables are represented by their first letter (and second for silicate).

Variable Pairing	Absolute Cross-correlation	Lag (months)
T/S	0.56	-2
T/N	0.54	-3
T/P	0.55	-3
T/Si	0.47	-5
T/C	0.43	130
T/O	0.70	-3
T/PDO	0.62	-2
T/NPGO	0.58	0
T/SOI	0.43	2
T/MEI	0.44	-2
S/T	0.52	2
S/N	0.82	0
S/P	0.65	0
S/Si	0.51	-3
S/C	0.73	0
S/O	0.54	-1
S/PDO	0.47	-5
S/NPGO	0.54	3
S/SOI	0.32	2
S/MEI	0.31	0
N/S	0.82	0
N/T	0.54	3
N/P	0.95	0
N/Si	0.80	-1
N/C	0.68	0
N/O	0.34	-1
N/PDO	0.43	-3
N/NPGO	0.58	4
N/SOI	0.26	-319
N/MEI	0.27	0
P/S	0.65	0
P/N	0.95	0
P/T	0.55	3
P/Si	0.89	0
P/C	0.59	-1
P/O	0.33	46
P/PDO	0.37	-2
P/NPGO	0.58	5
P/SOI	0.24	31
P/MEI	0.21	0
Si/S	0.51	3
Si/N	0.80	1
Si/P	0.89	0
Si/T	0.47	5
Si/C	0.48	1
Si/O	0.32	-22
Si/PDO	0.34	-316
Si/NPGO	0.54	6
Si/SOI	0.31	-28
Si/MEI	0.15	-3
C/S	0.73	0
C/N	0.68	0
C/P	0.59	1
C/Si	0.49	-1
C/T	0.43	-130
C/O	0.49	-1
C/PDO	0.49	-135
C/NPGO	0.38	-187
C/SOI	0.34	-131
C/MEI	0.34	-133
O/S	0.54	1
O/N	0.34	1
O/P	0.33	-46
O/Si	0.32	22
O/C	0.49	1
O/T	0.70	3
O/PDO	0.49	-2
O/NPGO	0.46	2
O/SOI	0.36	68
O/MEI	0.38	-2

Appendix C: Figures

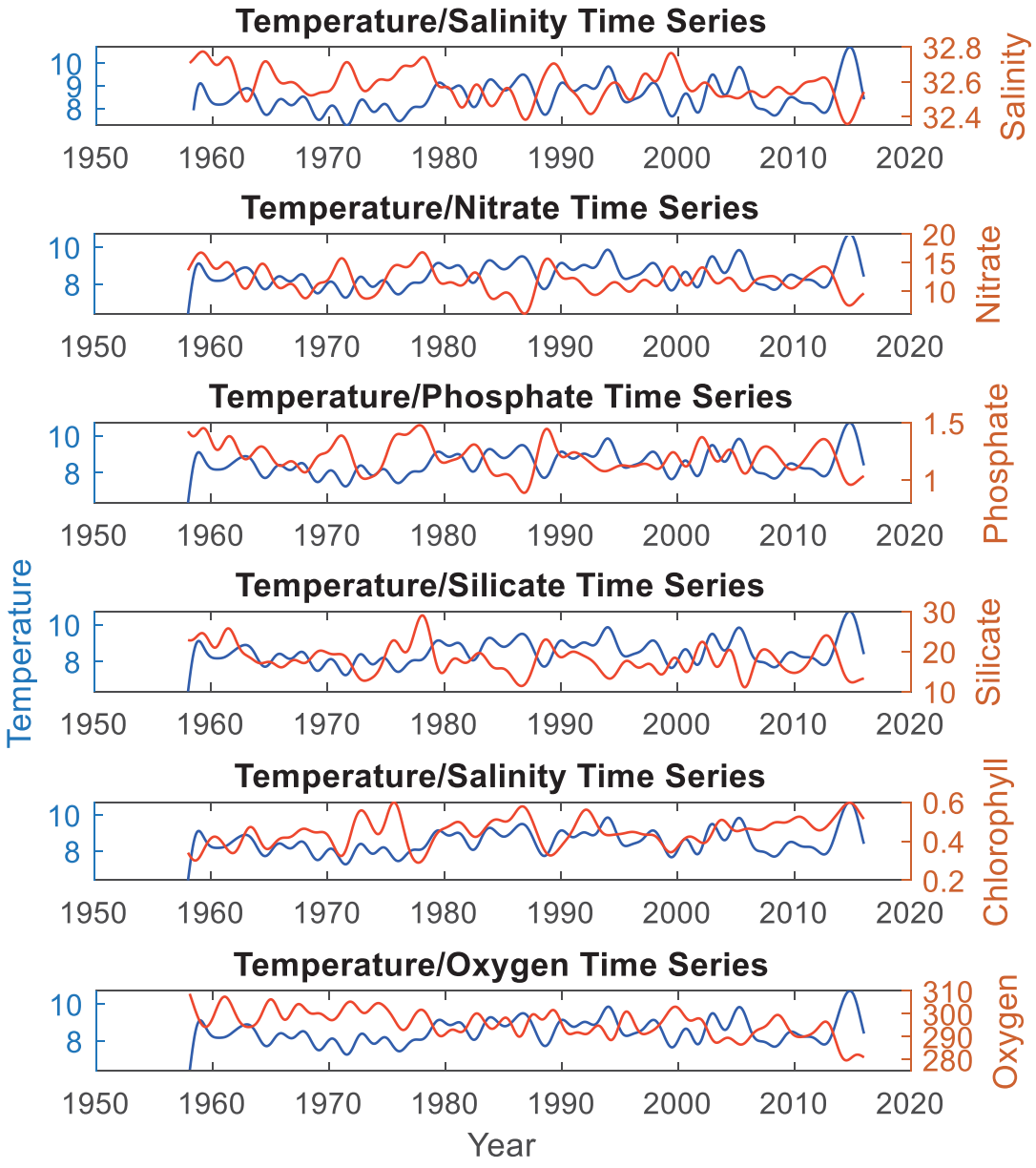

Figure C.1: Temperature vs. ocean state variable time series overplots. Temperature is represented in blue and the ocean state variables are represented in red. Units are as previously established.

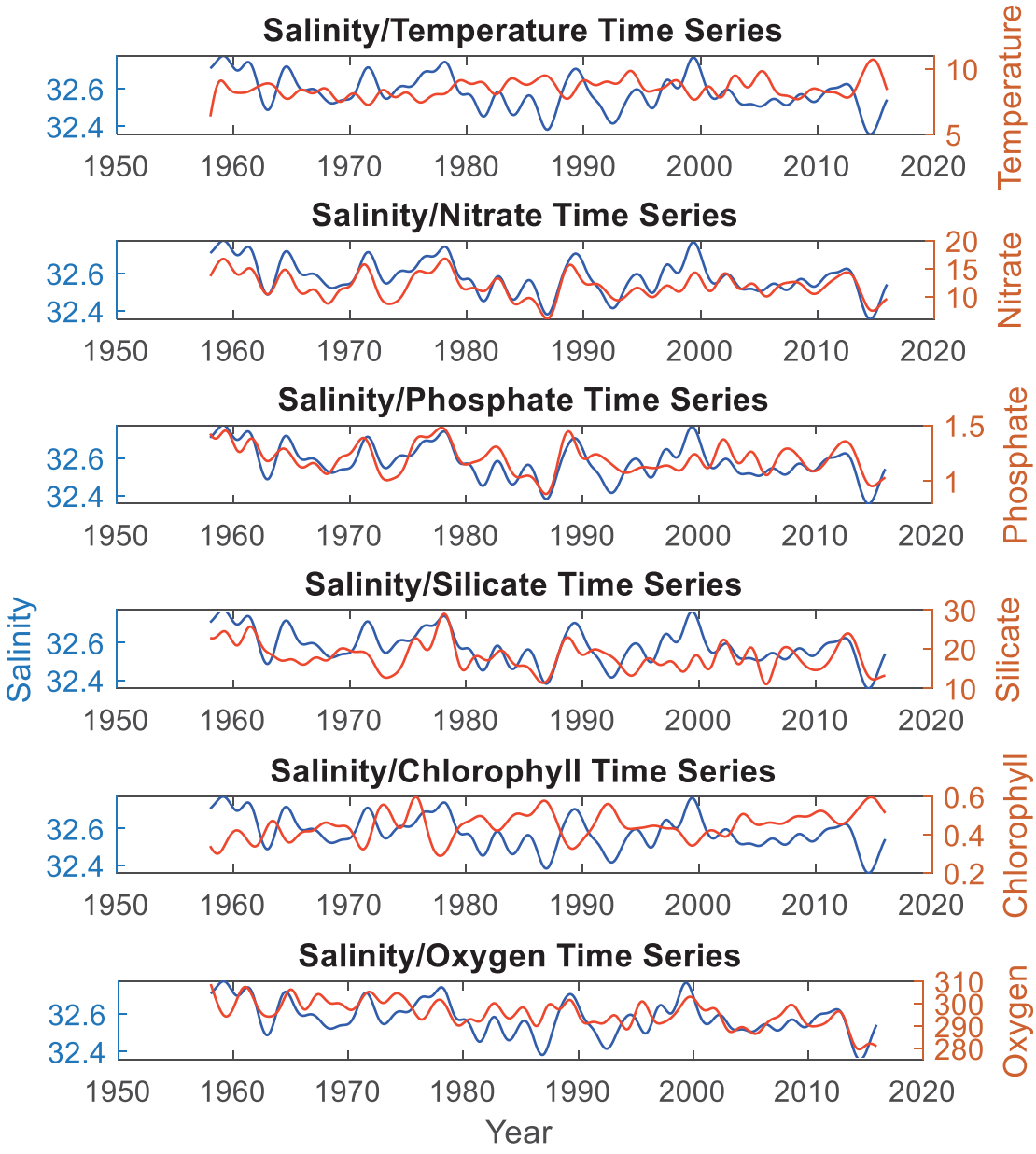

Figure C.2: Salinity vs. ocean state variable time series overplots. Salinity is represented in blue and the ocean state variables are represented in red. Units are as previously established.

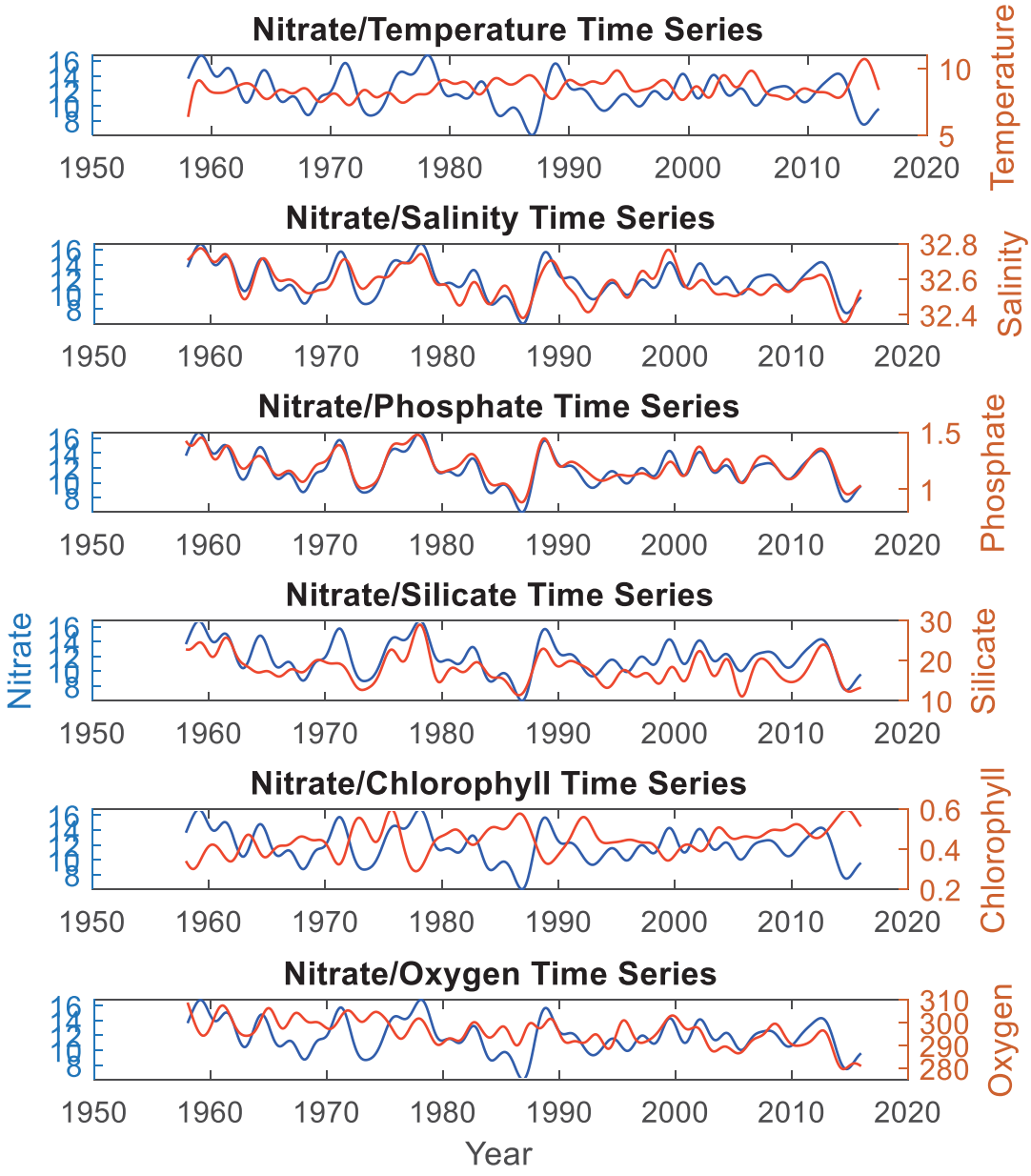

Figure C.3: Nitrate vs. ocean state variable time series overplots. Nitrate is represented in blue and the ocean state variables are represented in red. Units are as previously established.

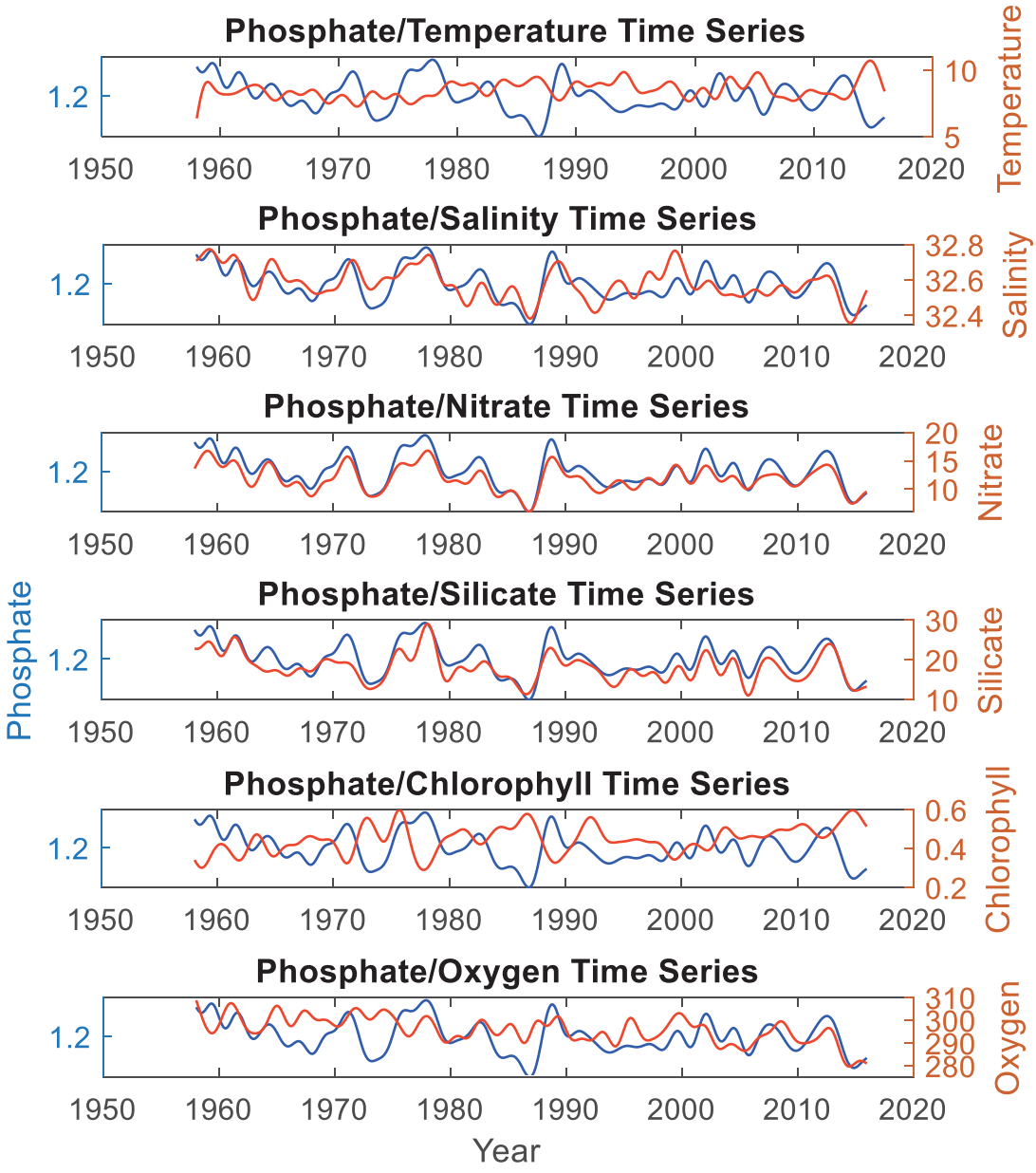

Figure C.4: Phosphatee vs. ocean state variable time series overplots. Phosphate is represented in blue and the ocean state variables are represented in red. Units are as previously established.

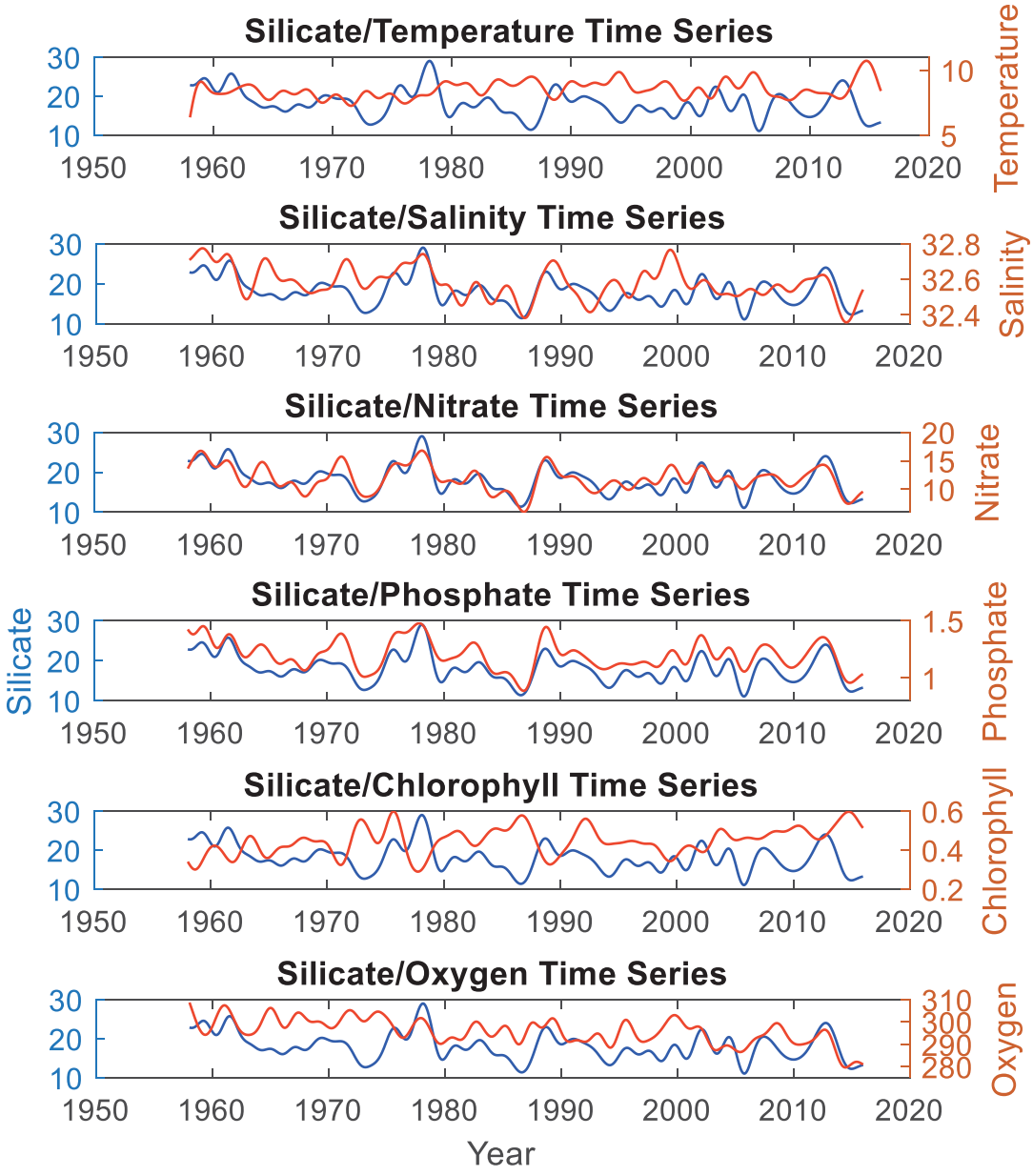

Figure C.5: Silicate vs. ocean state variable time series overplots. Silicate is represented in blue and the ocean state variables are represented in red. Units are as previously established.

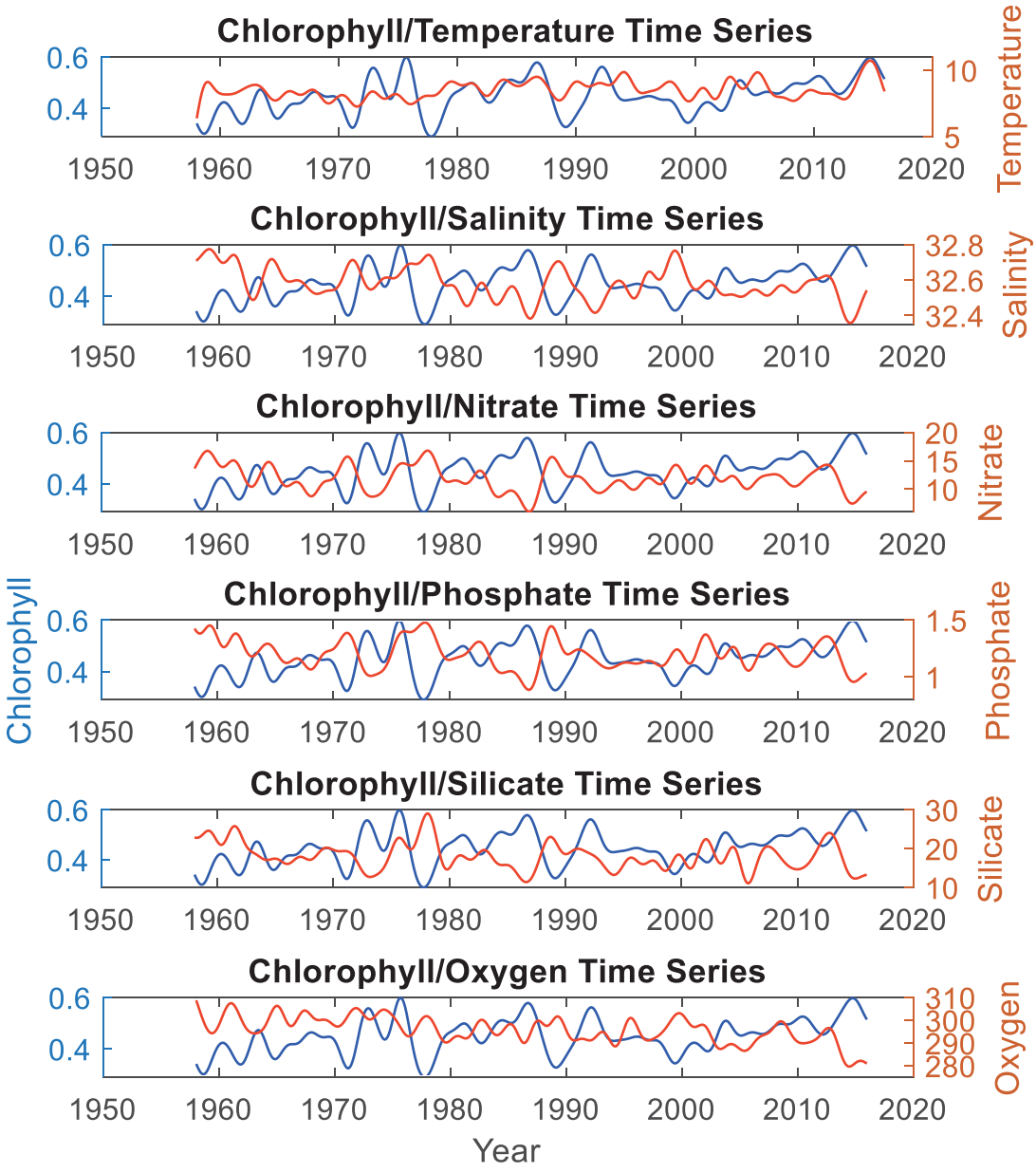

Figure C.6: Chlorophyll vs. ocean state variable time series overplots. Chlorophyll is represented in blue and the ocean state variables are represented in red. Units are as previously established.

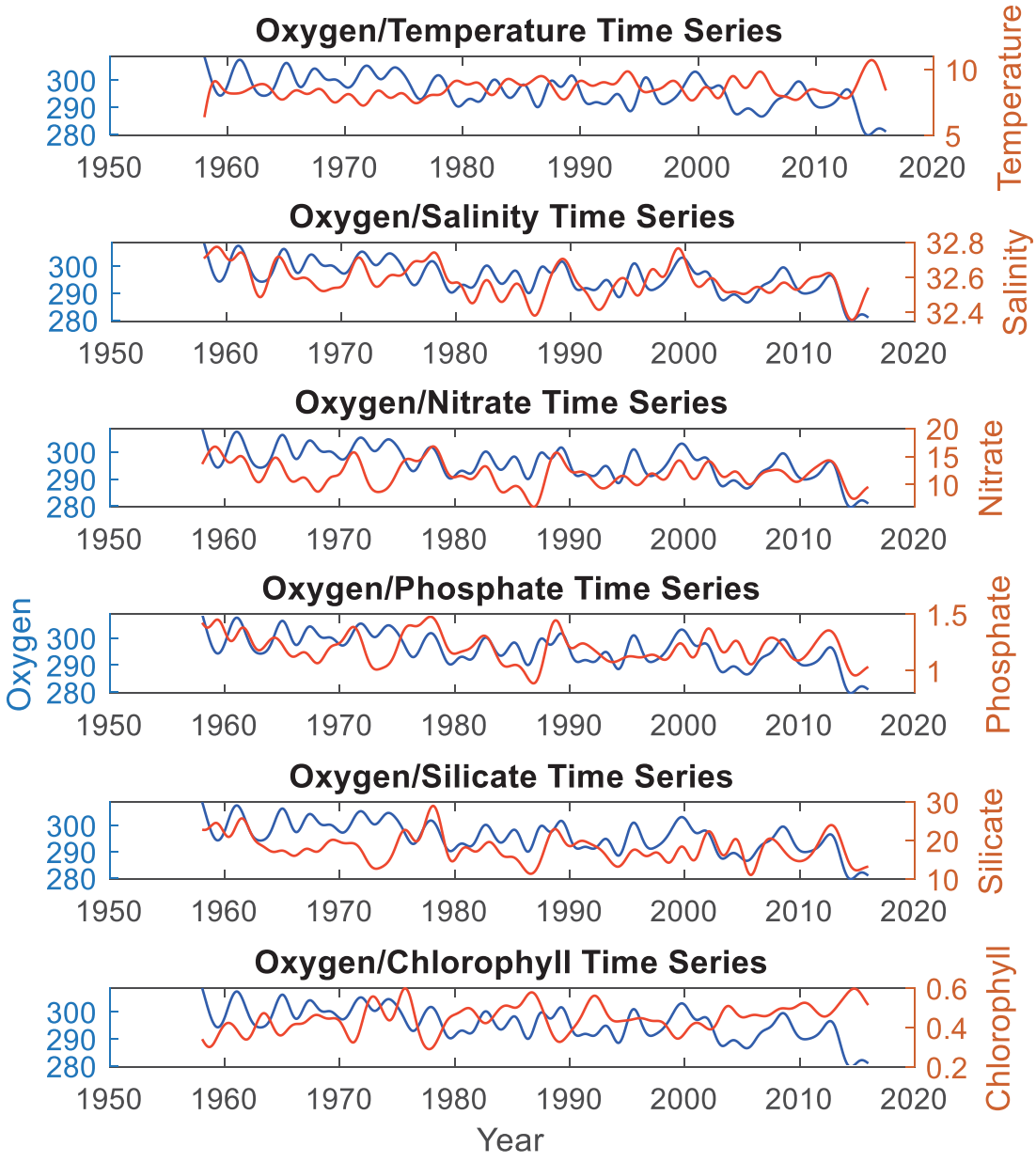

Figure C.7: Oxygen vs. ocean state variable time series overplots. Oxygen is represented in blue and the ocean state variables are represented in red. Units are as previously established.

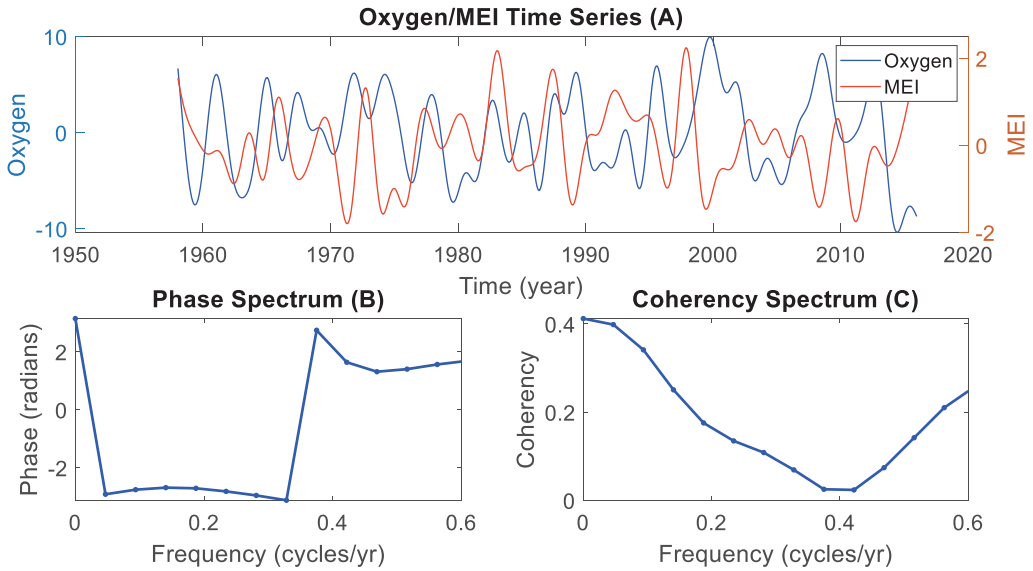

Figure C.8: An overplot of the scaled oxygen and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

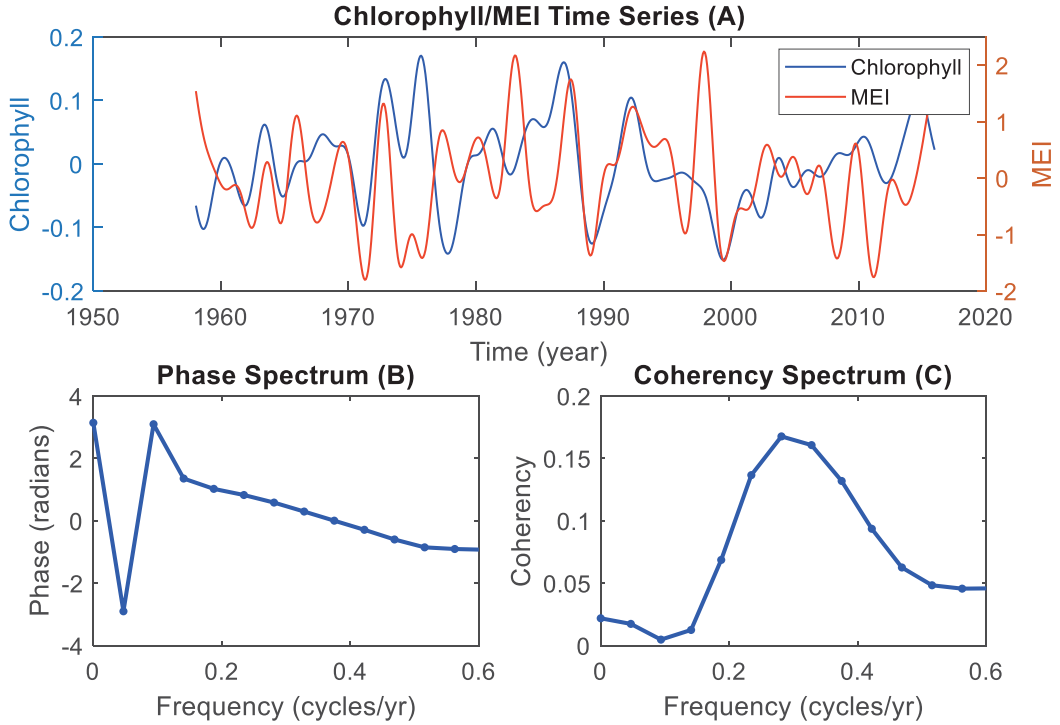

Figure C.9: An overplot of the scaled chlorophyll and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

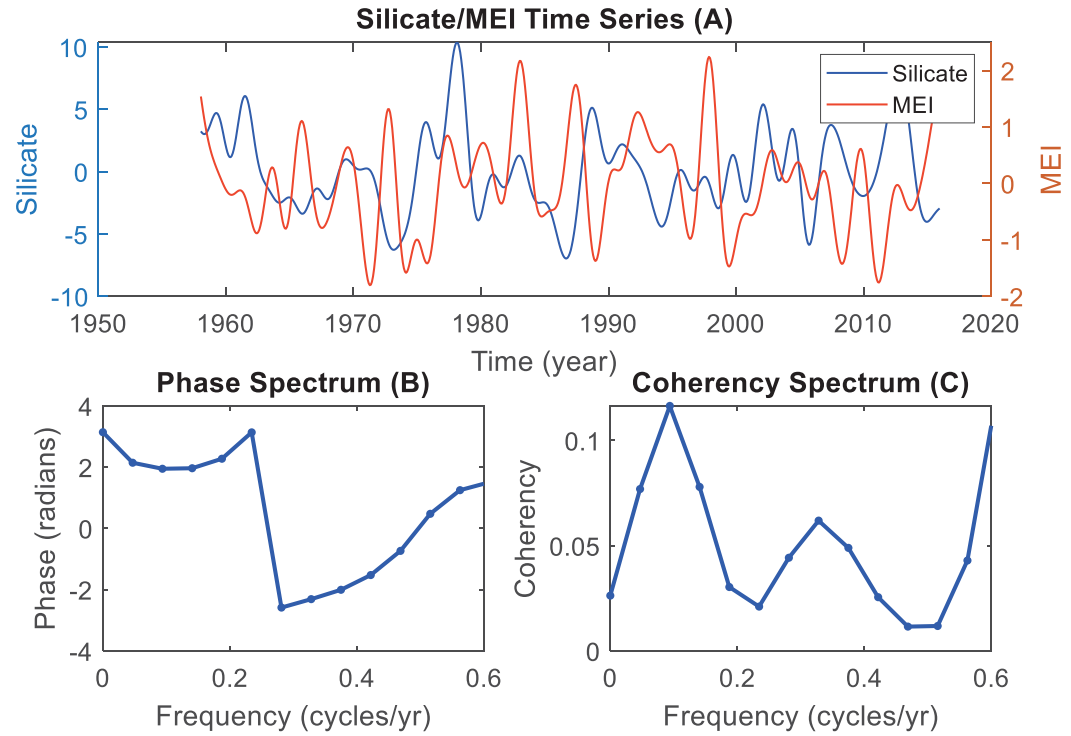

Figure C.10: An overplot of the scaled silicate and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

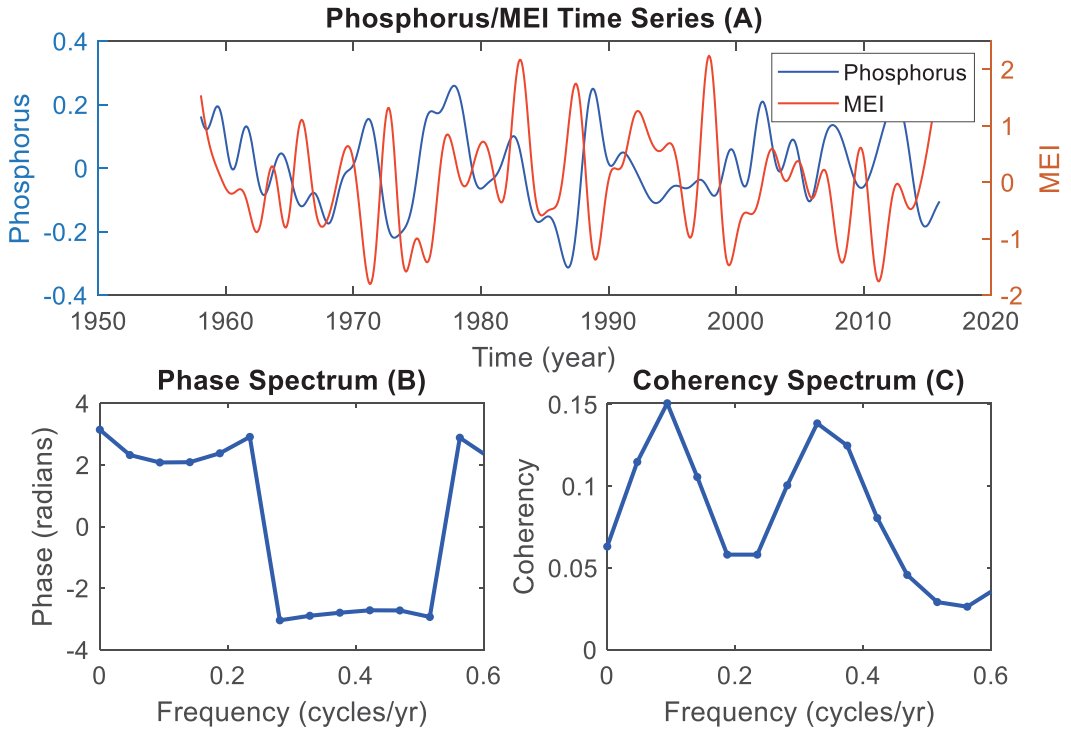

Figure C.11: An overplot of the scaled phosphate and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

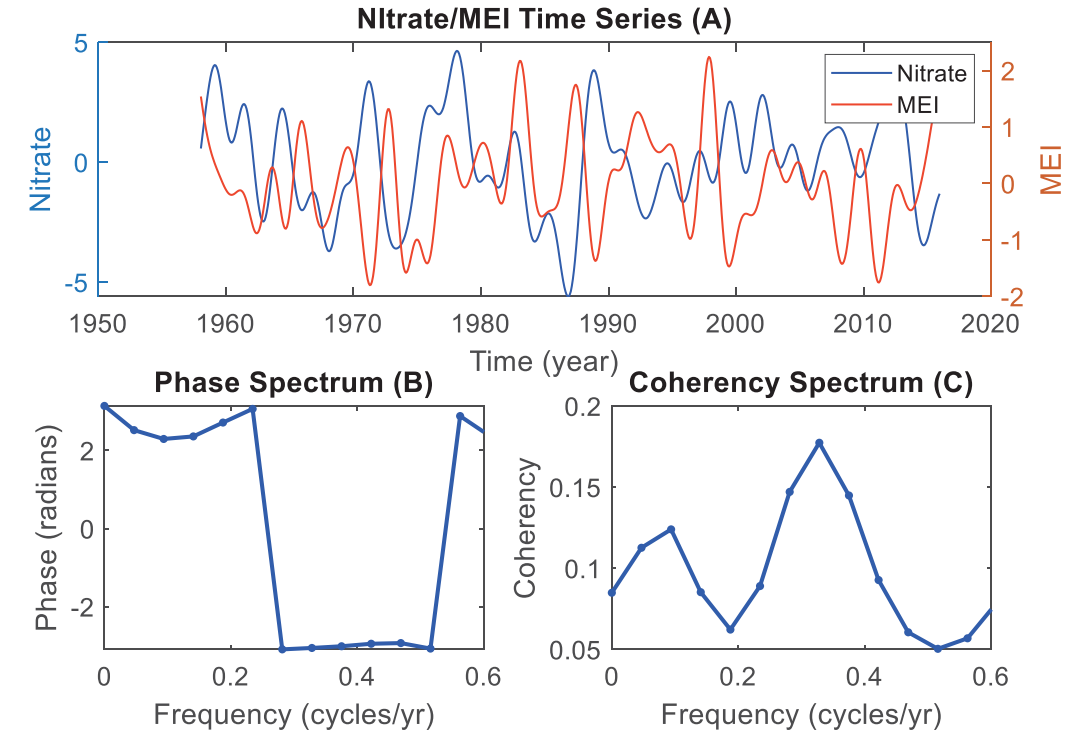

Figure C.12: An overplot of the scaled nitrate and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

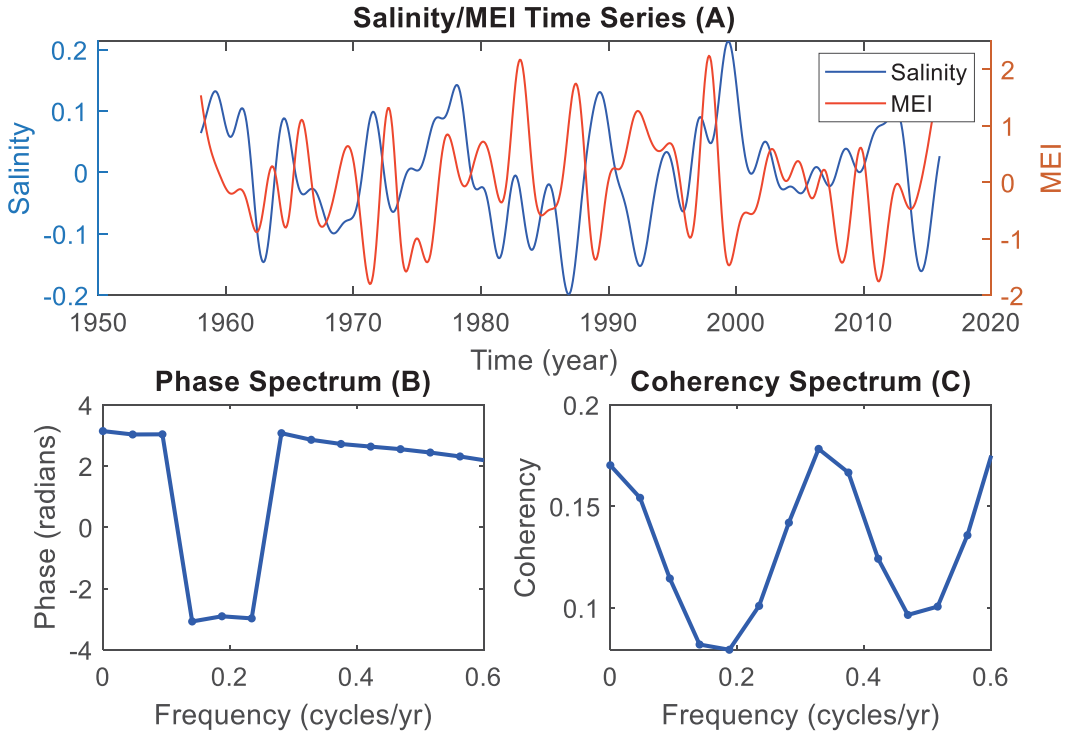

Figure C.13: An overplot of the scaled salinity and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

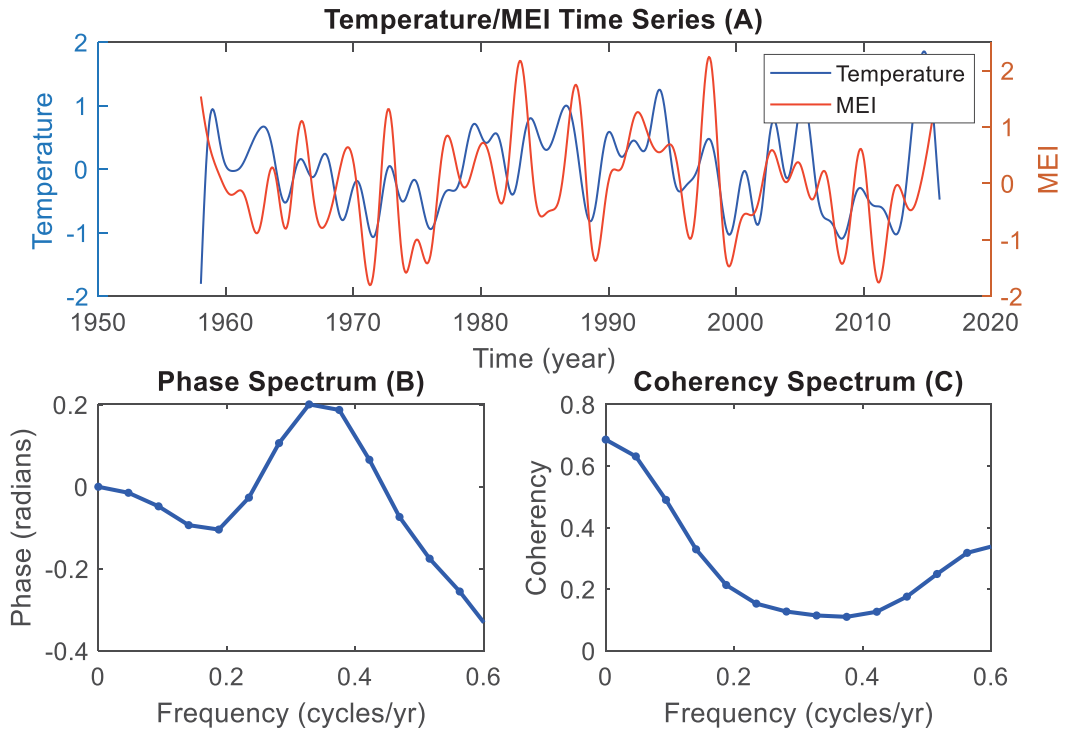

Figure C.14: An overplot of the scaled temperature and MEI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

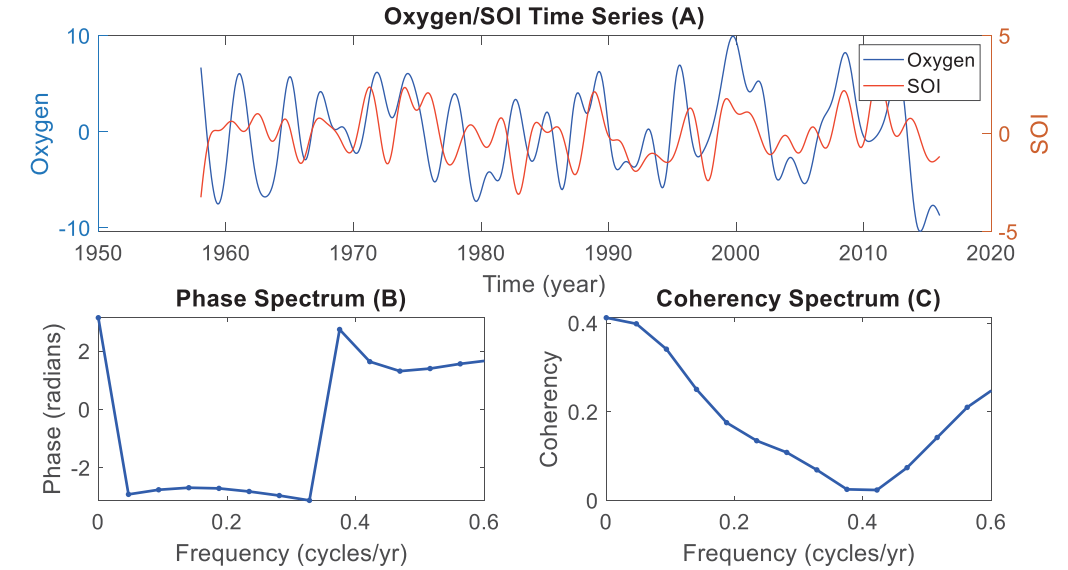

Figure C.15: An overplot of the scaled oxygen and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

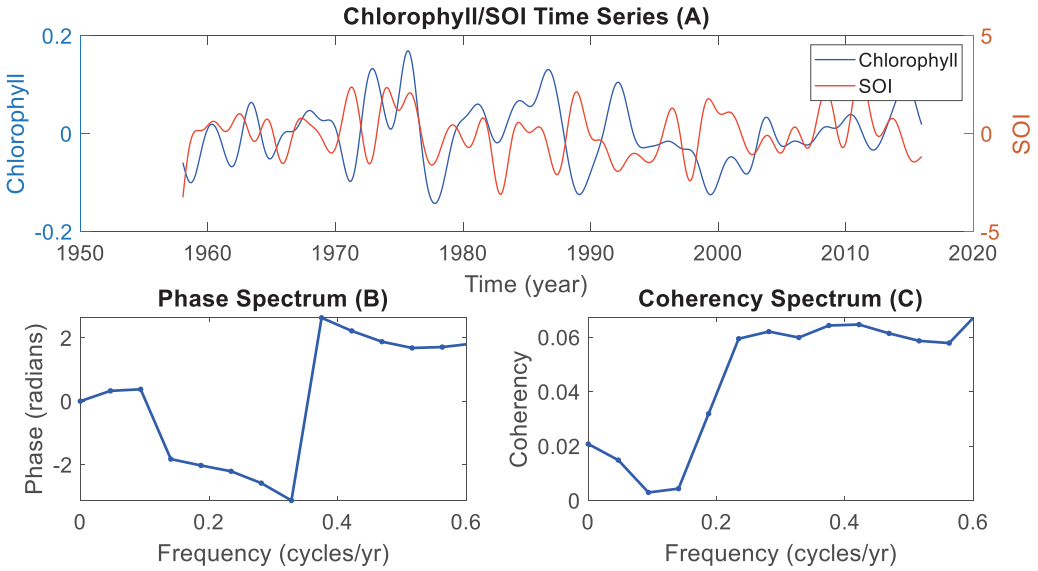

Figure C.16: An overplot of the scaled chlorophyll and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

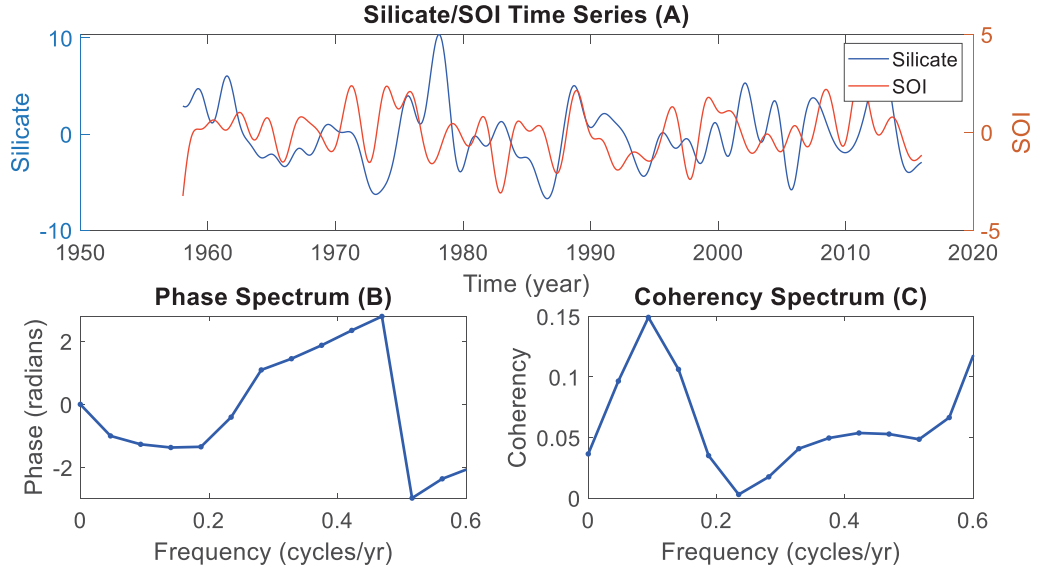

Figure C.17: An overplot of the scaled silicate and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

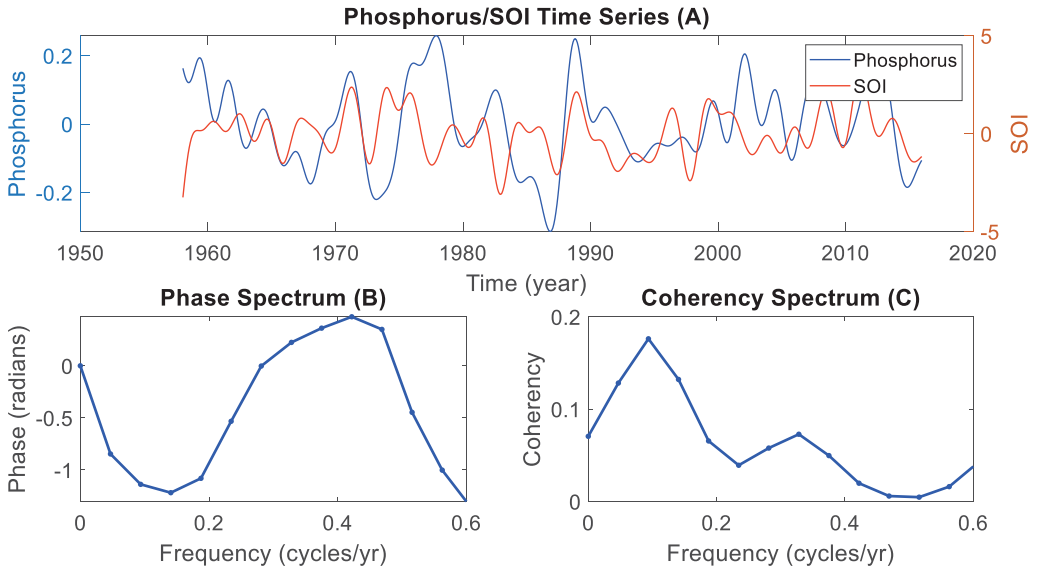

Figure C.18: An overplot of the scaled phosphate and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

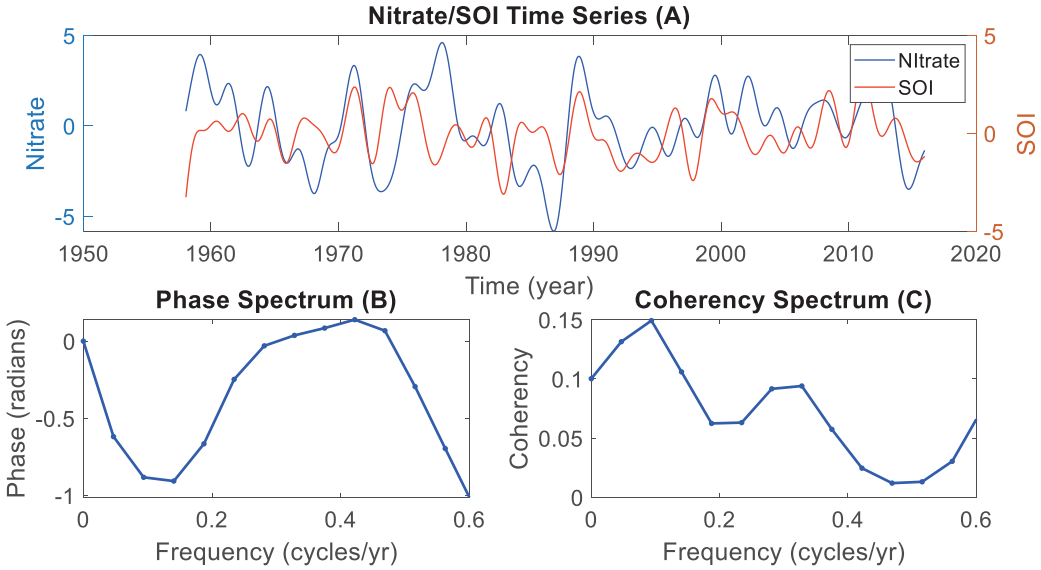

Figure C.19: An overplot of the scaled nitrate and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

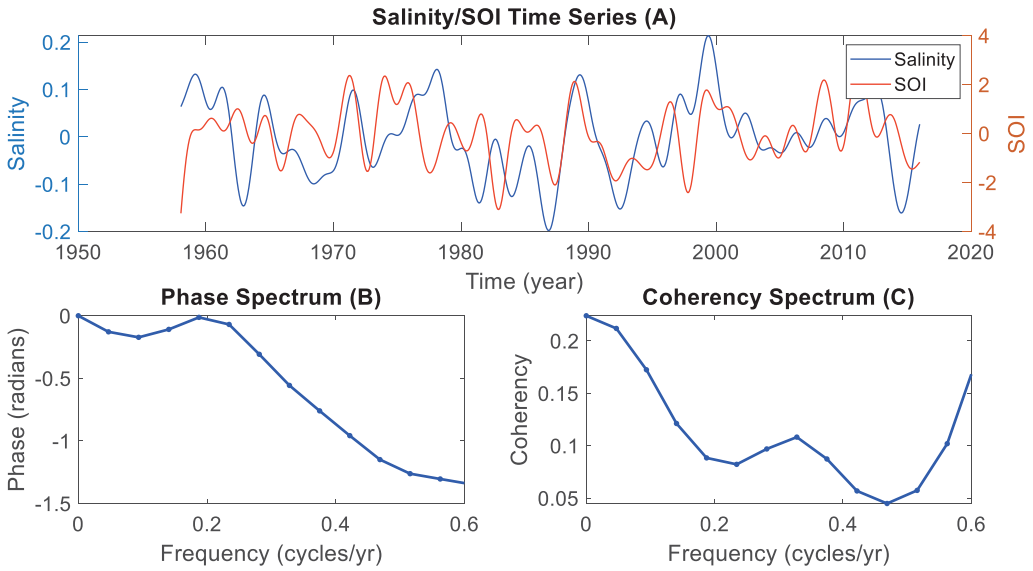

Figure C.20: An overplot of the scaled salinity and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

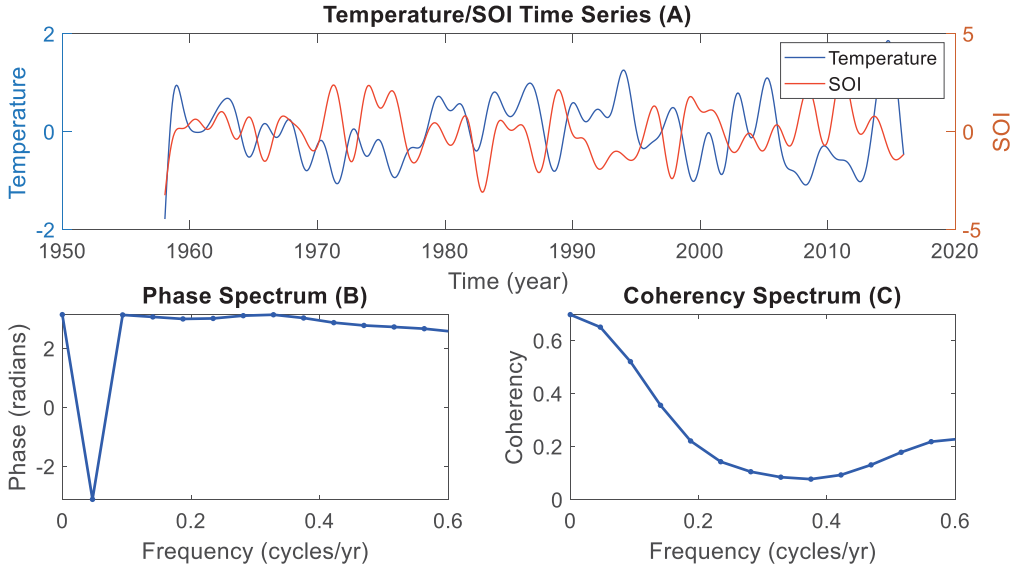

Figure C.21: An overplot of the scaled temperature and SOI time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

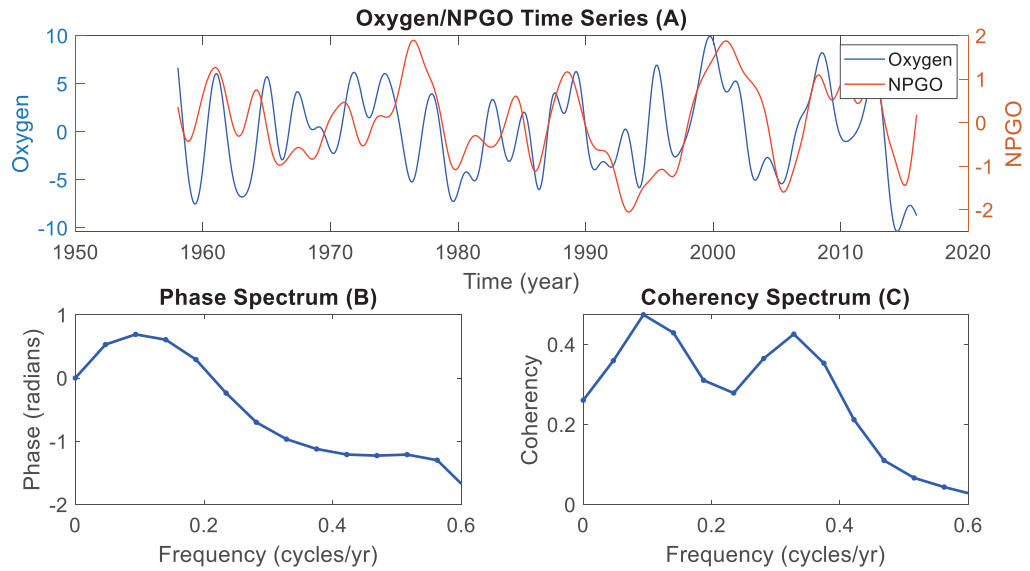

Figure C.22: An overplot of the scaled oxygen and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

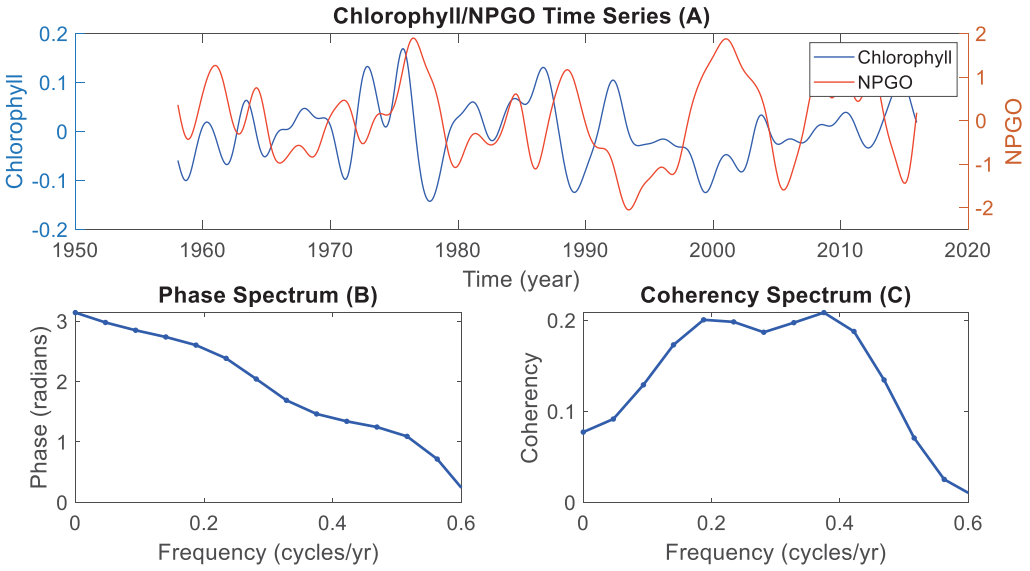

Figure C.23: An overplot of the scaled chlorophyll and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

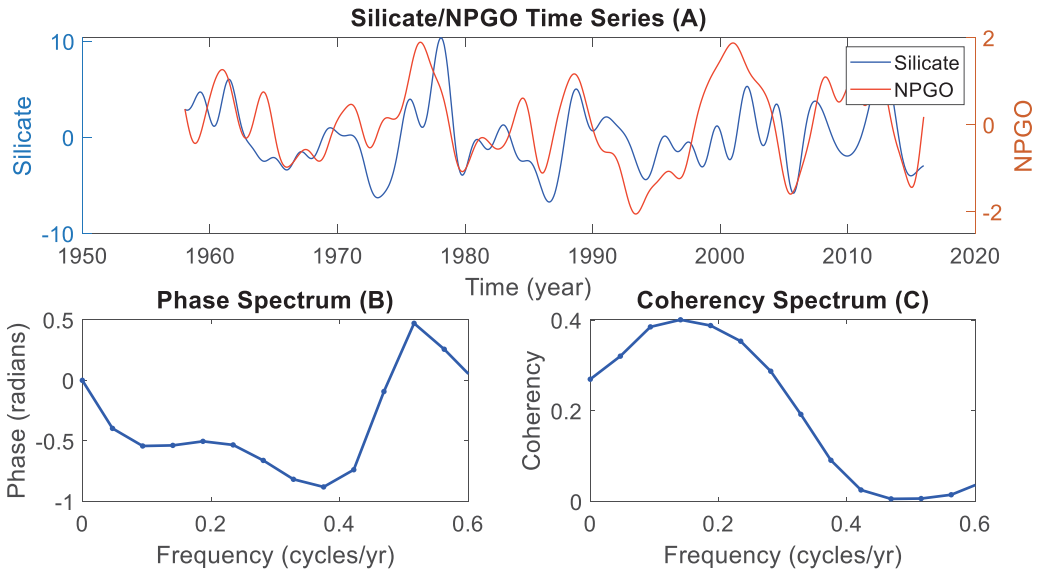

Figure C.24: An overplot of the scaled silicate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

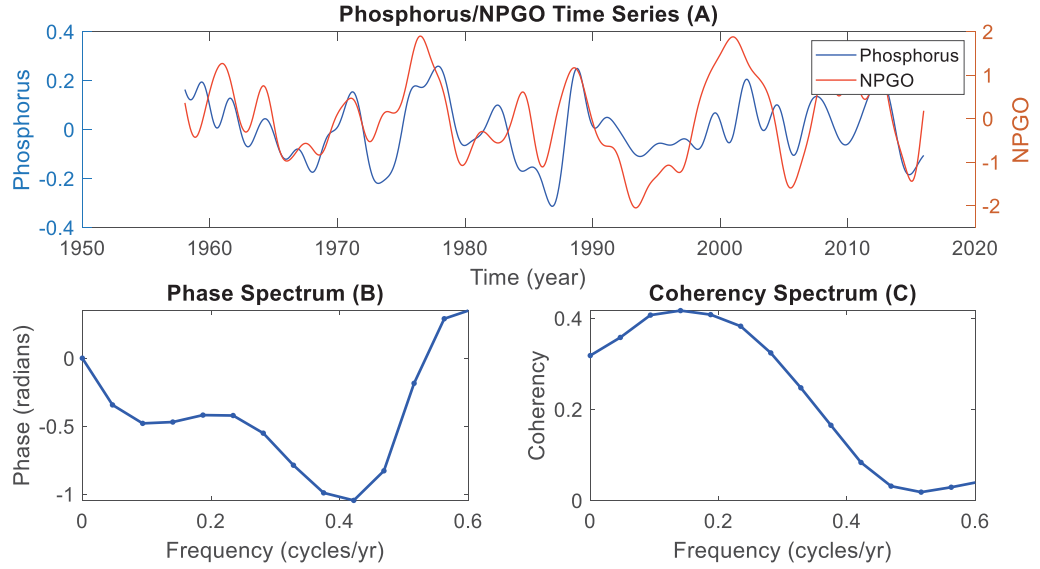

Figure C.25: An overplot of the scaled phosphate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

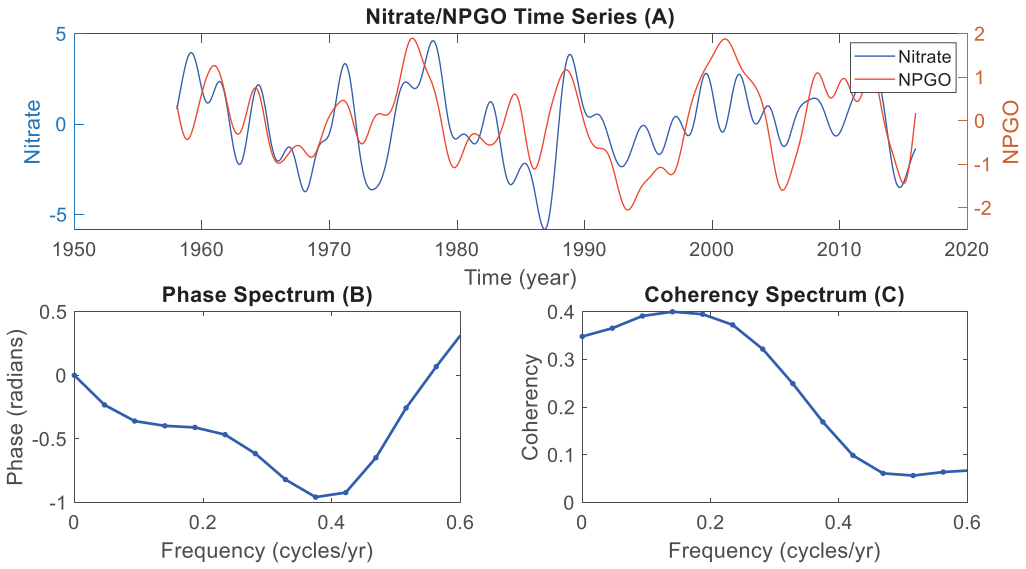

Figure C.26: An overplot of the scaled nitrate and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

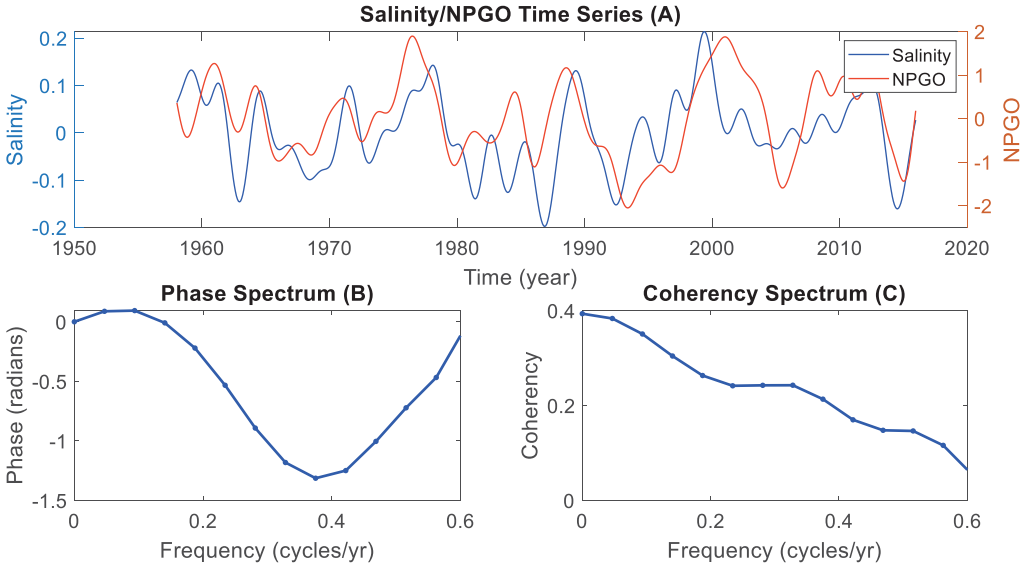

Figure C.27: An overplot of the scaled salinity and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

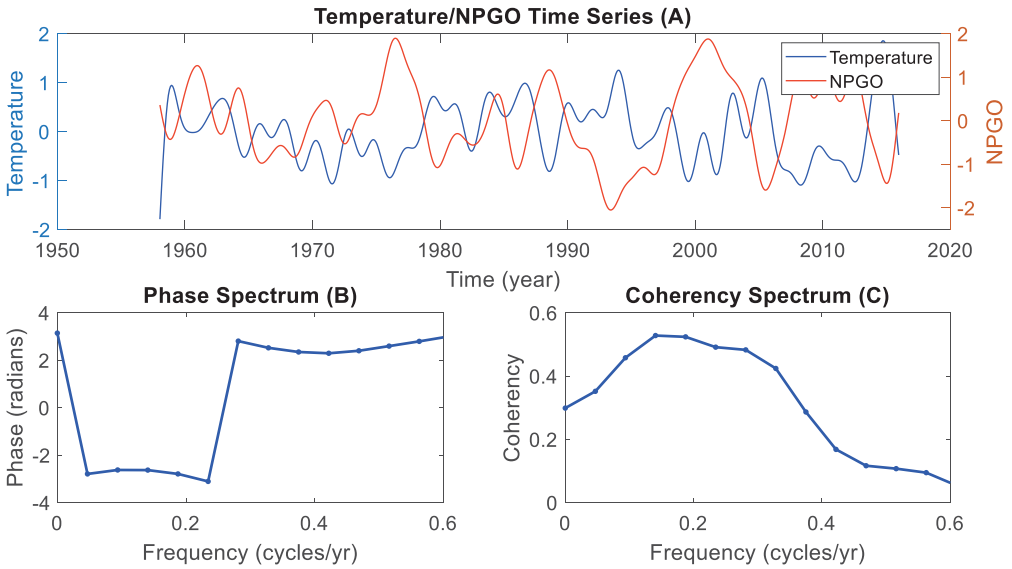

Figure C.28: An overplot of the scaled temperature and NPGO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

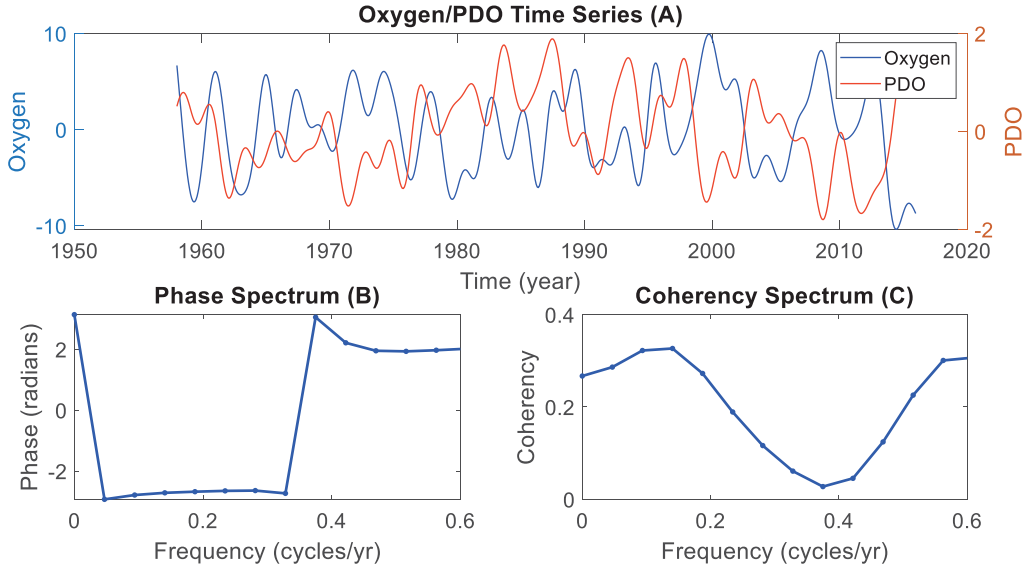

Figure C.29: An overplot of the scaled oxygen and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

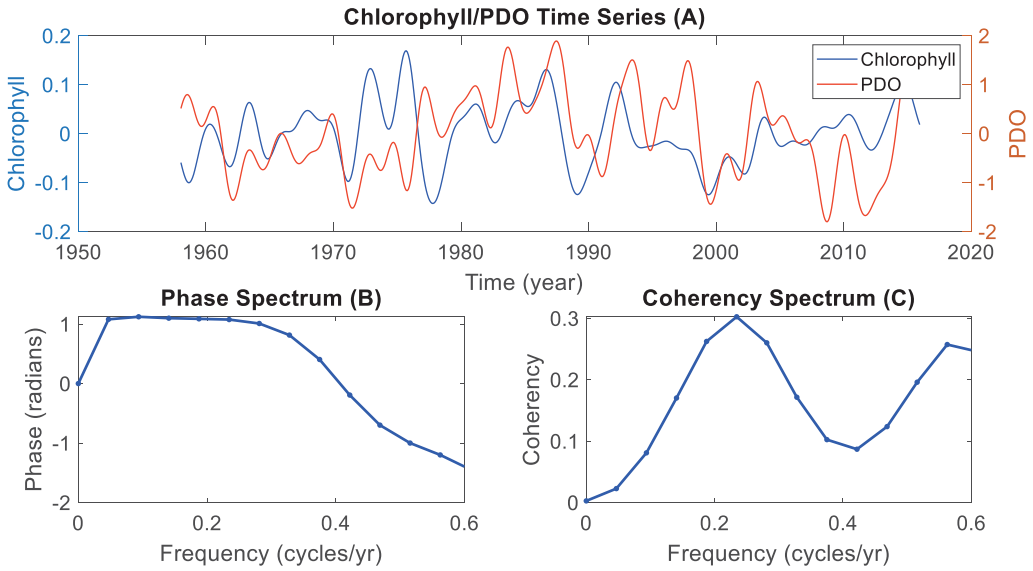

Figure C.30: An overplot of the scaled chlorophyll and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

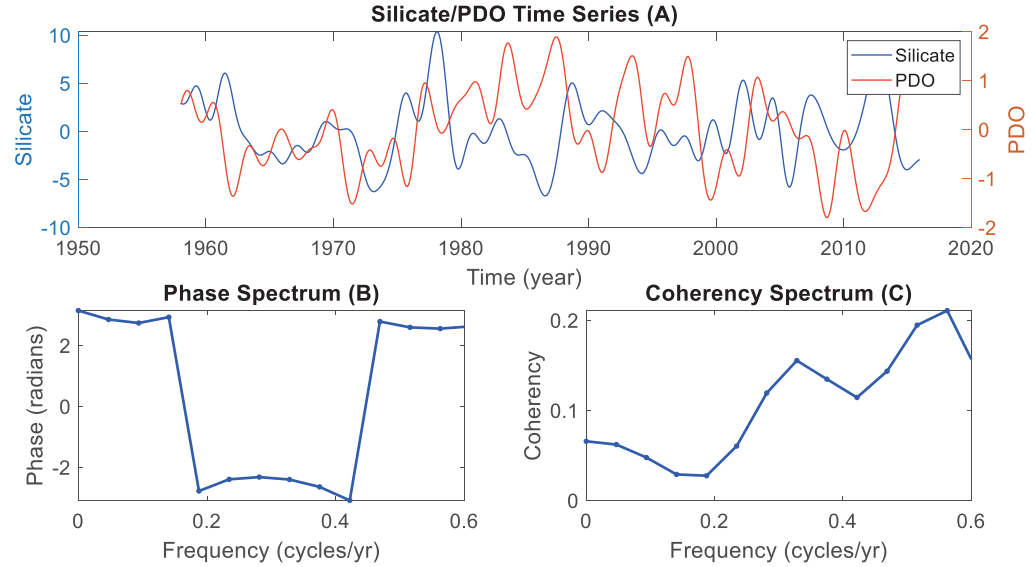

Figure C.31: An overplot of the scaled silicate and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

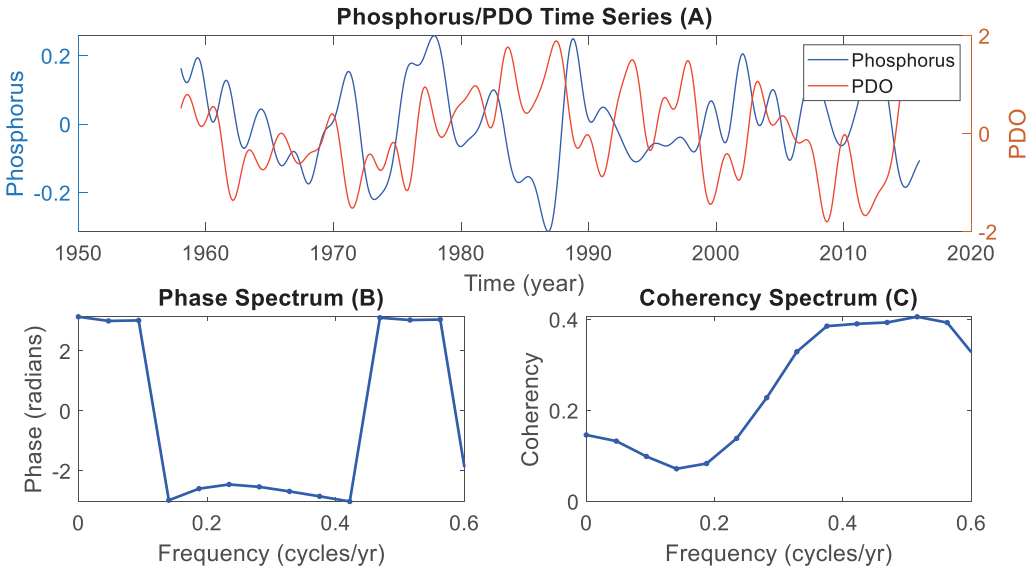

Figure C.32: An overplot of the scaled phosphate and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

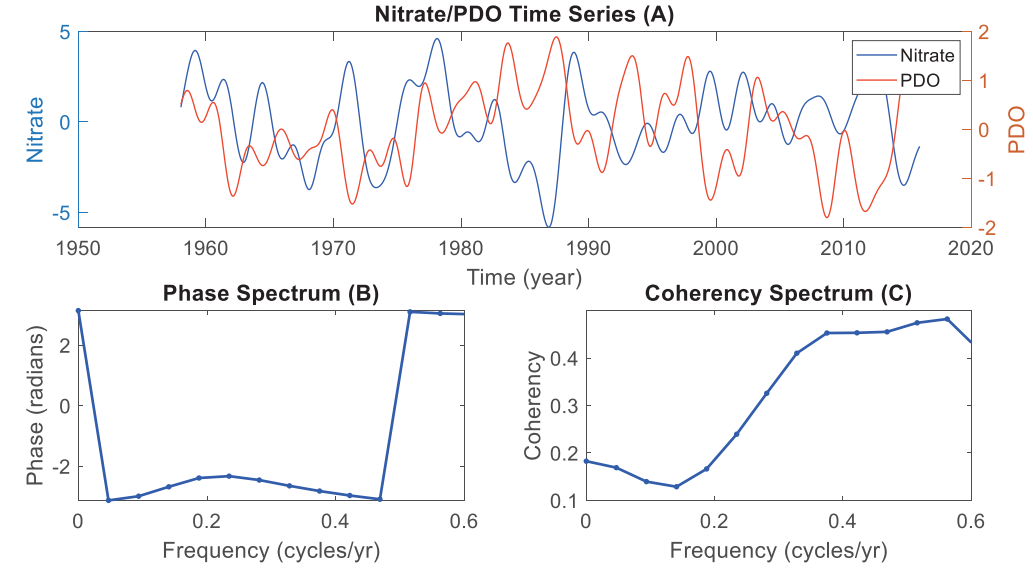

Figure C.33: An overplot of the scaled nitrate and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

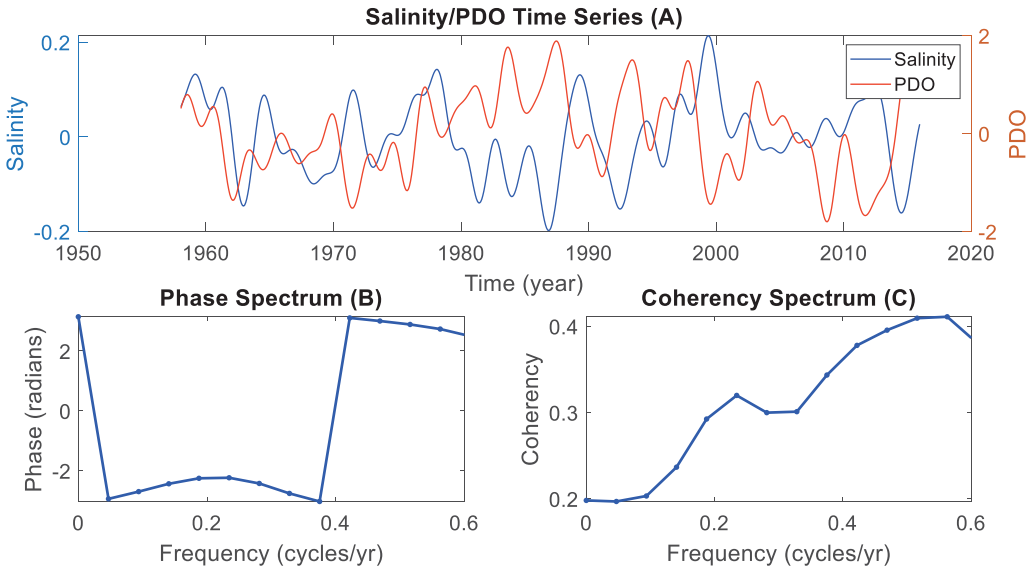

Figure C.34: An overplot of the scaled salinity and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.

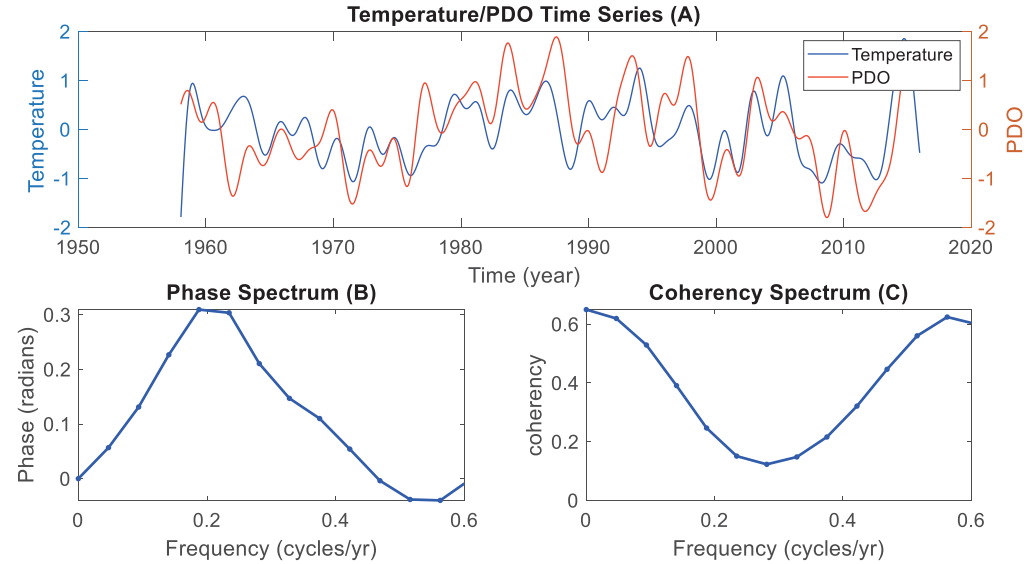

Figure C.35: An overplot of the scaled temperature and PDO time series (panel A), the phase spectrum with a focus on the lower frequencies (panel B) and coherency spectrum with a focus on the lower frequencies (panel C) from the cross-spectral analysis.