The Impact of Freshwater and Phytoplankton on Mixed Layer Heating

Rosalind Emma Echols

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Stephen C Riser, Chair

Gabrielle Rocap

Kyla Drushka

Program Authorized to Offer Degree:

Oceanography

University of Washington

Abstract

The Impact of Freshwater and Phytoplankton on Mixed Layer Heating

Rosalind Emma Echols

Chair of the Supervisory Committee:
Professor Stephen C Riser
Oceanography

Mixed layer heat budgets are governed by a range of horizontal and vertical processes, including net heat flux into the mixed layer. Net heat flux into the mixed layer is determined not only by the values of the heat fluxes at the air-ocean interface, but also by the depth to which shortwave radiation penetrates and the depth of the mixed layer. This dissertation investigates a range of physical and biological processes influencing these latter two factors, using observational data from autonomous profiling floats, as well as machine learning and a one-dimensional model. In Chapter 2, I present a new method for identifying barrier layers, a phenomenon whereby the near-surface density stratification is determined by changes in salinity. Barrier layers are typically associated with shallower mixed layers, and therefore potentially of critical importance in influencing the mixed layer heat budget in regions where they cover a

large percentage of the ocean. The method developed in this Chapter differs from previous methods in that it does not require choice of a temperature threshold, something that has hindered comparison of previous studies, and is therefore applicable without alteration throughout the tropical and subtropical oceans. In Chapter 3 I use this method to identify the prevalence of barrier layers in the Arabian Sea, and show that they are largely associate with decreased heating or increased cooling of the mixed layer because they allow more shortwave radiation to penetrate below the base of the mixed layer. In some cases, they allow enough radiation to penetrate that the net heat flux into the mixed layer shifts from positive to negative. In Chapters 4 and 5, I turn to the question of how the distribution of chlorophyll in the water column affects the absorption of shortwave radiation. Because no climatology of vertical chlorophyll in the global ocean exists, in Chapter 4 I use a machine learning clustering method to identify regional and seasonal patterns in the distribution of chlorophyll. I show that there are distinct patterns in the types of profiles that occur in different regions. I use the patterns developed in Chapter 4 as a tool for a one-dimensional modeling study in Chapter 5, where I quantify the expected impact of various chlorophyll-dependent parameterizations of shortwave radiation. I identify regions where the choice of parameterization is most sensitive to chlorophyll, and specifically those regions where the vertical distribution leads to a large departure from a surface-average based parameterization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

From the bottom of my heart, thank you to everyone who made this work possible, both during the time it took to complete it and in my life leading up to this point. This has been a lifelong dream of mine, and I could not have done it without the support of a large and enthusiastic committee.

Thank you to my advisor, Steve Riser, who took me on as a graduate student very far removed from my undergraduate years and supported me as a whole person throughout my time in graduate school. Between the two of us, we seem to have attempted to cram as many major life events into the last seven years as possible, while also talking about science and exploring projects of interest to us both. I appreciate the latitude for exploration and learning as well as the subtle redirection towards a concrete end goal.

Thank you to my committee, Kyla Drushka, Gabrielle Rocap, Mark Warner, Nathan Kutz, and Meghan Cronin, for your curiosity and breadth of interest. This work would have been far less interesting had it only existed in my own head, and the quality and variety of feedback I have received have made me a better scientist.

Thank you to the agencies that supplied the funding that made this work possible, particularly the National Science Foundation Graduate Research Fellowship, which gave me the independence to explore my own interests at my own pace.

Thank you to the community of the Science Leadership Academy. You taught me how to teach, mentor, and coach; how to open myself up to new experiences and learn from my mistakes; how to admit when I didn't understand and do better each day. I am so fortunate to have been a member of your community. Thank you for helping me develop the courage to speak up about the things that matter to me.

One of the greatest privileges of graduate school has been to be part of a kind and supportive lab group, that has expanded during my time in graduate school and developed into a rich learning community. Earle Wilson, Jesse Anderson, and Tyler Hennon welcomed me enthusiastically to graduate school and helped me figure out how to be a graduate student and scientist after so many years doing something else. I am deeply grateful for their model of kindness in a scientific community, and am delighted with the ongoing support system we have built. Ethan Campbell, Katy Christensen, and Jade Sauvé have all helped me immensely in developing and refining the ideas presented in Chapters 4 and 5 here, and our newest members Nina Buzby, Song Sangmin, Kristen Falcinelli, Treasure Warren, and Hannah Joy-Warren have shown immense curiosity and support as I worked toward the finish line.

So many other graduate students and postdocs helped make this possible, by continuously working to make our department a more welcoming place and by convincing me that I could do this and remain myself. That my ideas were worthy and that my presence was welcome. Shirley Leung demonstrated unmatched enthusiasm for some of the early ideas that developed into Chapter 4 and also showed me what it looks like to be courageous and speak up about what is right, even when it makes others uncomfortable. Hannah Glover and Claire McKinley helped me find a way to advocate for change while also supporting me through a lengthy and isolating pandemic. Sarah Dewey was relentless in making sure I knew that I was capable, even in the most difficult times. Theresa Whorley and Susanna Michael were endlessly supportive, especially as we all tried to finish our dissertations in the middle of a pandemic. And this does not even scratch the surface of all the kind and supportive words and actions of other graduate students throughout my time in graduate school. Everyone deserves the kind of support I have received.

My cohort has been a delight to know and work with. I have learned an enormous amount from the time we have spent together, and deeply value the opportunity to learn about fields outside my own from the most welcoming group of people I've ever encountered. Sasha Seroy, Isaiah Bolden, Christina Rathwell, Andrew Margolskee, Megan Duffy, Dylan Oldenburg, Hillary Scannell, and Caroline Bell have been such incredible colleagues, and made me feel like an expert even when I was drowning in self-doubt. They also showed me how to be well-rounded scientists and people.

Thank you to Mikelle Nuwer and Tansy Burns for helping me to find opportunities to share my love of teaching. It was so important to me not to lose my connection with my previous career, and you both went out of your way to support me in this.

Graduate school has been made easier by the fact that I was returning to the place where I grew up. My wonderful friends, Katie, Erin, Sarah, Erica, and Rachel, who nagged me continuously for 10 years about when I was moving back to Seattle, have helped me through some of the most difficult times of my life over the last seven years while also acting like it was completely obvious that I would be successful in this endeavor, no matter how much I struggled internally. Cora and Hannah, who have known me through many ups and downs, conveniently moved to Seattle at the same time as I did and helped remind that there is life out of grad school.

Being able to spend so much time with my parents, particularly as my family grew to include two children, has been a source of immense joy and offered me a source of support that I desperately needed at times. It is one of my greatest sorrows that my dad, who introduced me to the magic of the ocean as a child, was not able to be with me at the completion of this. My parents have always modeled insatiable curiosity, a feature that prepared me well for the world of graduate school. My mom also taught me the importance of having between one and 7 knitting projects going at any time, both for relaxation and also to allow my brain time to mull over complicated ideas. None of this would have been possible without the many infant and toddler sweaters I knit. My brothers, without whom I would most certainly not be the person I am today, are largely responsible for my eternal need to argue about everything, also preparing me for graduate school.

And finally, thank you to the loves of my life. Brandon, you are the most special person in the world. You uprooted your entire life so I could pursue this dream of mine, which on a day to day basis looked a lot like me complaining. All the time. Your unparalleled support and enthusiasm have helped me through so many difficult times, and the kindness and compassion you model have made me a better person. No one in my life has ever demonstrated such pride in my accomplishments, even the most mundane ones. And our babies, Meredith and Gryffyn, have brought me such great joy. How would I have made it through so many Zoom seminars during the pandemic without Meredith shouting, “It’s the ocean!” every other slide. Being surrounded by love is the best part of doing this work.

DEDICATION

For Meredith—thank you for always reminding me how exciting and wonderful the ocean is.

And for my dad—Thank you for introducing me to the magic of the ocean. I miss you.

Chapter 1. INTRODUCTION

Spatial and temporal patterns of sea surface temperature (SST) play an important role in a wide range of physical phenomena affecting both the ocean and the atmosphere, as well as organisms living in the ocean. The magnitude and distribution of SST, particularly in the tropics, contributes to the structure of large-scale atmospheric circulation (Hoskins et al, 1999; Rao and Sivakumar 1999; Shenoi et al. 1999), submesoscale circulation at fronts (Edwards et al, 2001), and can affect mixing and upwelling when linked to changes in mixed layer depth and stratification (Murtugudde et al, 2002). Regionally elevated SSTs can enhance the strength of tropical storms (Vecchi and Soden, 2007), and passing over oceanic barrier layers can further enhance storm intensity because storm-driven mixing does not cool, and may even warm, the upper ocean (Balaguru et al, 2012). SST also influences the habitability of the upper ocean for organisms, oxygen solubility, and stratification at the base of the mixed layer (and thus access to nutrients from the deep ocean). Understanding the factors that control SST and how they vary both now and under future climate change is therefore essential for understanding the interactions between the ocean and the atmosphere as well as the biological and physical systems in the ocean.

One way to examine the influence of different factors on establishing patterns in SST is through a mixed layer heat budget, which provides a means of accounting for different physical effects. These come in many forms, but generally include terms for the advection of water from one region to another, mixing of different water masses across the base of the mixed layer, and direct heating or cooling due to heat fluxes. Longwave, sensible, and latent heat fluxes are all boundary conditions at the ocean-atmosphere interface, and can only penetrate into the ocean

interior through mixing or diffusion. Thus, we can accurately assume these are absorbed completely within the mixed layer prior to mixing. Shortwave radiation, on the other hand, penetrates into the ocean where it is absorbed and scattered, leading to heating throughout the upper ocean. Shortwave radiation is therefore not necessarily absorbed entirely within the mixed layer, which can significantly impact the balance of heat fluxes and subsequent heating or cooling of the mixed layer. Indeed, several observational studies have shown that penetrative shortwave radiation at the mixed layer depth is a large fraction of the mean air-sea net heat flux (Siegel et al, 1995; Lewis et al, 1990; Echols and Riser, 2020). Modeling studies have illustrated that assuming that all shortwave radiation is absorbed within the mixed layer can significantly shift SST outcomes (e.g. Lewis et al 1990, who found that incorporating penetrative radiation into a model amounted to a reduction in surface temperature of 5-10°C annually, for a 20m mixed layer depth). However accounting for all the factors that influence where the shortwave radiation is absorbed remains complex. The goals of this thesis are to better understand some of the ocean conditions that are likely to significantly alter the ways in which shortwave radiation is absorbed within the upper ocean and the subsequent effects on mixed layer heating.

Whether or not shortwave radiation is absorbed entirely in the mixed layer can be affected by both physical and biological variability in the water column. Absorption of shortwave radiation in the ocean is typically parameterized as an exponential decay, dependent on the optical clarity of the water using what is known as a diffuse attenuation coefficient, K_d. As a result, small differences in mixed layer depth can produce significant differences in the amount of heat retained in the mixed layer owing to the rapid change in absorption with depth. The specific form of the exponential varies from a single exponential to a wavelength and chlorophyll-dependent, vertically varying exponential (**Figure 1.1**; Denman, 1973; Paulson and

Simpson, 1973; Morel, 1988; Morel and Antoine, 1993; Ohlmann et al, 1996, Manizza et al, 2005). Even between these three relatively simple exponential models, we see significant variation in the amount of radiation absorbed above 20m. For mixed layers in that depth range, this could alter the shortwave radiation retained in the mixed layer by 10% depending on the choice of parameterization. The vast range in complexity can present a challenging decision-making process in selecting the best parameterization, particularly when the consequences of selecting different parameterizations are not well quantified in conjunction with what we know about the physical and biological structure of the ocean. Therefore, in this dissertation I explore two scenarios that are likely to heavily influence where shortwave radiation is absorbed in the ocean: regions that experience dominant upper ocean salinity stratification within a deeper isothermal layer, and the spectrum of vertical chlorophyll profiles exhibited in the global ocean.

While thermal stratification dictates the depth of the mixed layer in much of the ocean, some regions that are heavily influenced by river runoff, meltwater, or precipitation exhibit salinity control instead. In these cases, known as barrier layers, the base of the mixed layer can become decoupled from the beginning of the thermocline, as in **Figure 1.2**. Barrier layers can be formed through a variety of mechanisms, including advection of freshwater, tilting of uncompensated salinity fronts, and excess precipitation. The mixed layers associated with barrier layers are frequently shallower than the value produced using a temperature threshold, and depending on the strength of the salinity stratification can create a significant obstacle to surface cooling when deeper water is entrained into the mixed layer. This is particularly significant in tropical regions where it alters the balance of terms in the mixed layer heat budget and increases the importance of the heat flux term, since direct one-dimensional heating is less likely to be counteracted by entrainment cooling. However, previous studies have used such a broad range of

temperature thresholds for determining whether a barrier layer is present that comparison between studies is complicated and the absolute importance of barrier layers is unclear in many regions. Various studies have suggesting that barrier layers may lead to increased warming (Lukas and Lindstrom, 1991), increased cooling (Vinayachandran, 2007), or even minimize the impact of atmospheric forcing on the upper ocean (Vialard and Delecluse, 1998). At least some of this uncertainty relates to the sign of the heat flux retained in the mixed layer, which in turn hinges on how much of the shortwave radiation penetrates below the base of the mixed layer.

Furthermore, chlorophyll concentration also influences where in the water column shortwave radiation is absorbed, and varies over several orders of magnitude at all upper ocean depths. **Figure 1.3** shows decay curves for a range of average chlorophyll concentrations. Between the lowest and highest chlorophyll concentration, there is a difference of 10% of the total incoming radiation absorbed above 20m. Chlorophyll concentration, associated with phytoplankton growth, varies regionally and seasonally, and is not restricted to the surface layers of the ocean that are easily observed by satellites. The concentrations shown in **Figure 1.3** could be observed in a single region at different times throughout the year. Numerous studies have explored the theoretical impact of different parameterizations, and invariably these studies show that the choice of parameterization is important (e.g. Manizza et al, 2005; Oschlies, 2004). These often rely on satellite chlorophyll data or earth system models for chlorophyll-dependent parameterizations, or on bulk regional assessments of an appropriate value for the diffuse attenuation coefficient in the exponential decay (Murtugudde et al, 2002). None of these studies have attempted to use observational data about the vertical distribution of chlorophyll in the global ocean to assess the consequences of selecting different parameterizations, despite

evidence that vertical distribution of chlorophyll can influence upper ocean heating in ways not evident using surface values (Manizza et al, 2005, Lewis et al, 1983).

In this dissertation, I will explore how fresh-water barrier layers and chlorophyll distribution each affect the evolution of SST. In Chapter 2, I propose a new framework using the thermodynamic variable spice for the identification of barrier layers with the goal of eliminating confusion between studies that arises from the wide range in temperature threshold values used to identify differences between the mixed layer depth and isothermal layer depth. I use the Arabian Sea as a case study to show how spice can be used to assess likely mechanisms for the formation and erosion of barrier layers, which influence their expected impact on the evolution of SST.

In Chapter 3, I again focus on the Arabian Sea and explore how the presence of barrier layers is likely to affect the evolution of SST using a mixed layer heat budget. I find that differences in mixed layer depths and the lack of cooling when those mixed layers deepen both play an important role in determining the upper ocean heat balance. Seasonal variations in mixed layer depth, presence of temperature inversions, and barrier layer thickness can shift the effect from positive to negative. These two chapters together highlight the importance of accurately assessing the presence and characteristics of barrier layers in determining SST evolution.

In Chapter 4, I shift to looking at the characteristics of chlorophyll concentration in the ocean, identifying patterns in profile shapes using unsupervised machine learning and the global Argo biogeochemical database. This chapter demonstrates the effectiveness of using machine learning for this application when compared with a more classical approach using curve fitting, and reveals global patterns in the distribution of chlorophyll in the water column. I find some similarities with previous studies that have attempted to categorize global ocean chlorophyll

profiles, but with an expanded set of possible profiles. The geographical distribution of these profiles lends some insight into mechanisms influencing the distribution of phytoplankton in the water column, bearing some resemblance to previous studies of biogeography.

Finally, in Chapter 5 I use the representative average chlorophyll profiles developed in Chapter 4 to explore how a range of parameterizations for the absorption of shortwave radiation influence the evolution of SST in a simple one-dimensional model. The results from these simulations show how atmospheric fluxes, ocean structure, and distribution of chlorophyll in the water column collectively contribute to the evolution of SST. I use the results from these simulations to highlight regions in the ocean that are particularly sensitive to the choice of parameterization and quantify the range of effects likely to be observed.

References

Balaguru, K., Chang, P., Saravanan, R., Leung, L.R., Xu, Z., Li, M., & Hsieh, J.-S. (2012). Ocean barrier layers' effect on tropical cyclone intensification. PNAS, 109, 36, 14343-14347, https://doi.org/10.1073/pnas.1201364109.

Denman, K. L. (1973). A time-dependent model of the upper ocean. Journal of Physical Oceanography, 3(2), 173-184.

Echols, R., & Riser, S.C. (2020). The Impact of Barrier Layers on Arabian Sea Surface Temperature Variability. Geophysical Research Letters, 47, e2019GL085290. https://doi.org/10.1029/2019GL085290.

Edwards, A. M., Platt, T., & Wright, D. G. (2001). Biologically induced circulation at fronts. Journal of Geophysical Research: Oceans, 106(C4), 7081-7095.

Hoskins, B., Neale, R., Rodwell, M., & Yang, G.Y. (1999). Aspects of the large-scale tropical atmospheric circulation. Tellus, 51 A-B, 33-44.

Lewis, M.R., Cullen, J.J., & Platt, T. (1983). Phytoplankton and Thermal Structure in the Upper Ocean: Consequences of Non-uniformity in Chlorophyll Profile. Journal of Geophysical Research, 88, C4, 2565-2570.

Lewis, M.R., Carr, M.E., Feldman, G.C., Esaias, W., & McClain, C. (1990). Influence of penetrating solar radiation on the heat budget of the equatorial Pacific Ocean. Nature, 347.

Lukas, R. and E. Lindstrom (1991). The Mixed Layer of the Western Equatorial Pacific Ocean, Journal of Geophysical Research, 96, Supplement, 3343-3357.

Manizza, M., Le Quéré,C., Watson, A.J., & Buitenhuis, E.T. (2005). Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. Geophysical Research Letters, 32, L05603, doi:10.1029/2004GL020778.

Morel, A. (1988) Optical Modeling of the Upper Ocean in Relation to Its Biogenous Matter Content (Case I Waters). Journal of Geophysical Research, 93, C9, 10,749-10,768.

Morel, A. & Antoine, D. (1993). Heating Rate within the Upper Ocean in Relation to its Bio-Optical State. Journal of Physical Oceanography, 24, 1652-1665.

Murtugudde, R., Beauchamp, J., McClain, C.R., Lewis, M., & Busalacchi, A.J. (2002). Effects of Penetrative Radiation on the Upper Tropical Ocean Circulation. Journal of Climate, 15, 470-486.

Ohlmann, J.C., Siegel, D.A., & Gautier, C. (1996). Ocean Mixed Layer Radiant Heating and Solar Penetration: A Global Analysis. Journal of Climate, 9, 2265-2280.

Oschlies, A. (2004). Feedbacks of biotically induced radiative heating on upper-ocean heat budget, circulation, and biological production in a coupled ecosystem-circulation model. Journal of Geophysical Research, 109, C12031, doi:10.1029/2004JC002430.

Paulson, C.A., & Simpson, J.J. (1977) Irradiance Measurements in the Upper Ocean. Journal of Physical Oceanography, 7, 952-956.

Rao, R.R. & Sivakumar, R. (1999). On the possible mechanisms of the evolution of a mini-warm pool during the pre-summer monsoon season and the genesis of onset vortex in the south-east Arabian Sea. Quarterly Journal of the Royal Meteorological Society, 125, 787-809.

Shenoi, S.S.C, Shankar, D. & Shetye, S.R. (1999). On the sea surface temperature high in the Lakshadweep Sea before the onset of the southwest monsoon, Journal of Geophysical Research, 14, C7, 15703-15712.

Siegel, D.A., Ohlmann, J.C., Washburn, L., Bidigare, R.R., Nosse, C.T., Fields, E., & Zhou, Y. (1995). Solar radiation, phytoplankton pigments and the radiant heating of the equatorial Pacific warm pool. Journal of Geophysical Research, 100, C3, 4885-4891.

Vecchi, G.A., & Soden, B.J. (2007). Effect of remote sea surface temperature change on tropical cyclone potential intensity. Nature, 450, 1066-1070, doi:10.1038/nature06423.

Vialard, J., & Delecluse, P. (1998). An OGCM study for the TOGA decade. Part 1: Role of salinity in the physics of the Western Pacific Fresh Pool. Journal of Physical Oceanography, 28, 1071-1088.

Vinayachandran,P., Shankar, D., Kurian, J., Durand, F., & Shenoi, S. S. C. (2007). Arabian Sea mini warm pool and the monsoon onset vortex. Current Science, 93(2), 203-214.

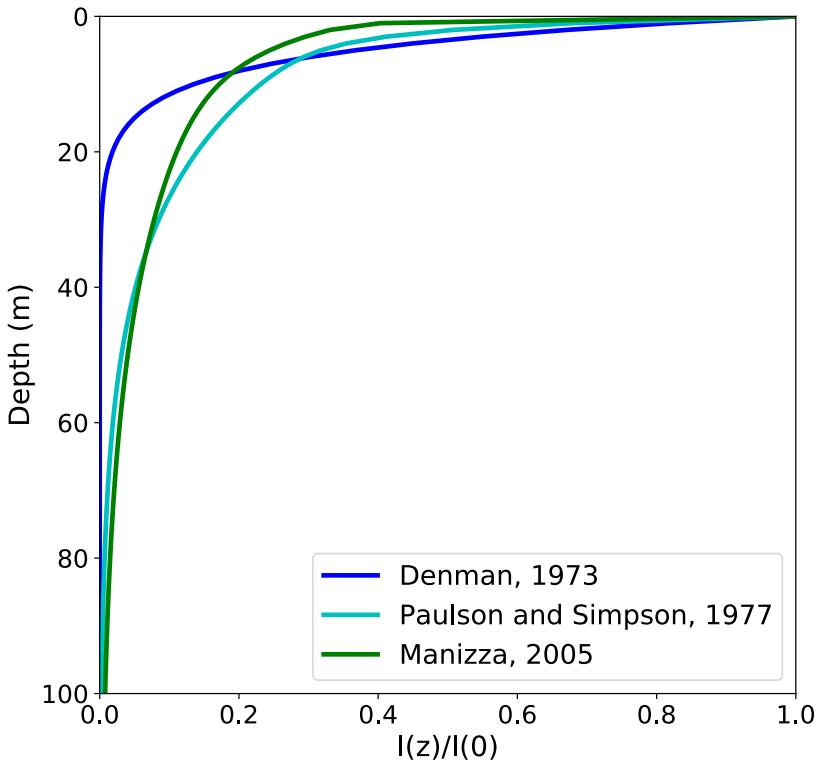

Figure 1.1. Irradiance curves for three different parameterizations: a single exponential with K_d = 0.2 m^{-1} (Denman, 1973); a double exponential with an infrared (visible) fraction of 0.62 (0.38) and infrared K_d = 0.66 m^{-1} (0.05 m^{-1}) (Paulson and Simpson, 1977); and triple exponential incorporating average chlorophyll concentration with [*Chl*]=0.05 mg m^{-3} (Manizza et al, 2005; see Chapter 5 here for details).

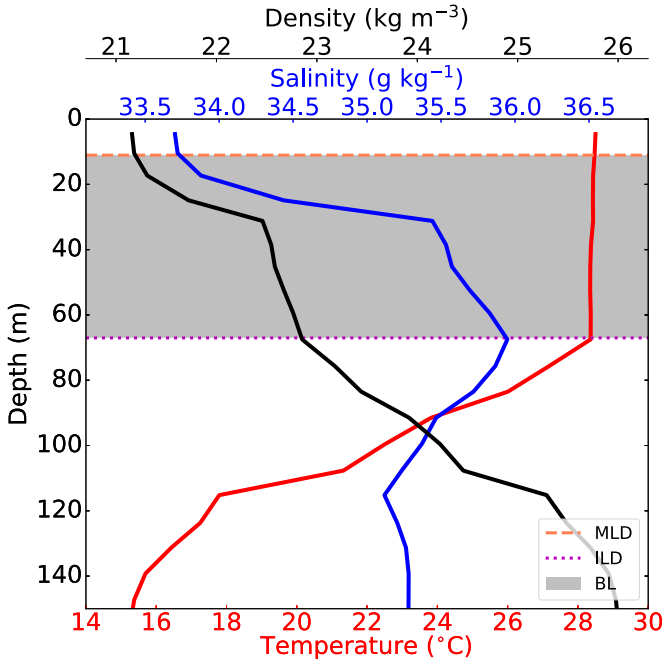

Figure 1.2. Profile exhibiting a barrier layer. Dashed orange line is the mixed layer depth, dotted magenta line is the base of the isothermal layer, and gray-shaded region is the barrier layer, indicating the separation between the base of the mixed and isothermal layers.

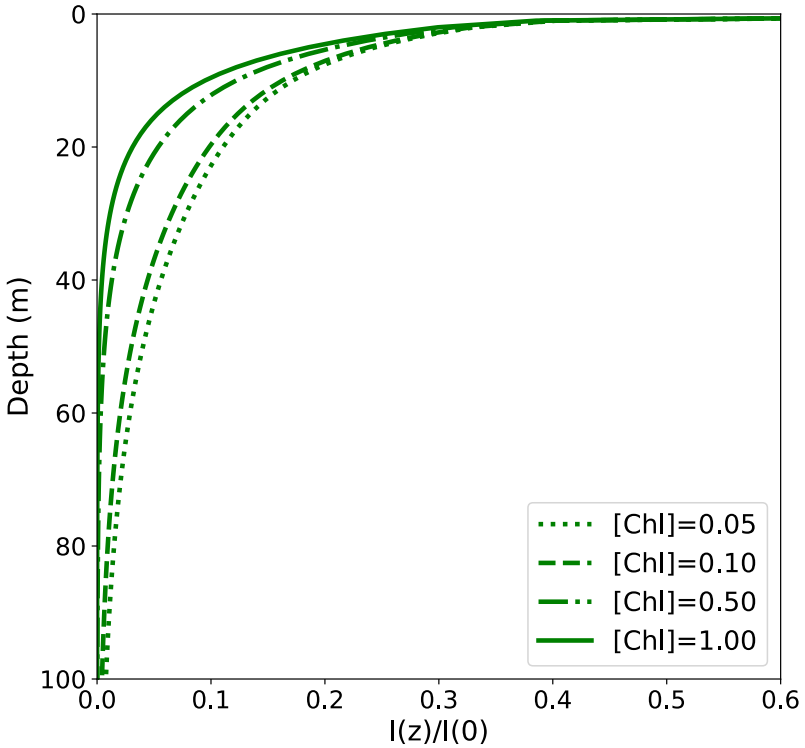

Figure 1.3. Irradiance decay curves for four different concentrations of chlorophyll (mg m^{-3}). All curves computed using a triple exponential decay derived from surface average chlorophyll (Manizza et al, 2005; see details in Chapter 5, Equation 2).

Chapter 2. SPICE AND BARRIER LAYERS: AN ARABIAN SEA CASE STUDY

This manuscript has been previously published as: Echols, R. & Riser, S.C. (2020). Spice and Barrier Layers: an Arabian Sea Case Study. Journal of Physical Oceanography, 50, 695-714, https://doi.org/10.1175/JPO-D-19-0215.1

2.1 ABSTRACT

Barrier layers (BL) are a frequent occurrence in low-latitude oceans, but variations in identification methods and quantitative descriptors used, as well as analyses of their lifetimes, lead to differing views on their overall impact. Herein a new method is proposed for identifying BLs that relies on the vertical spice profile rather than arbitrary temperature thresholds. Using 13 years of Argo profiling float data from the Arabian Sea, this method is shown to produce reasonable assessments of BL characteristics in the region. Similarly, use of the spice variable to identify formation mechanisms shows promise for using limited observational data and climatology to reproduce mechanisms proposed from modeling studies. Upper ocean stability calculations combining profiling float and atmospheric forcing data are used to suggest the most likely one-dimensional mechanisms for erosion of BLs, and to calculate the expected lifetimes of BLs throughout the year. Consistent with the evidence about the seasonal spatial extent and frequency of BLs, their duration is expected to be 2-3 times longer during the northeast monsoon than the southwest monsoon. However, the most likely erosion mechanisms vary widely throughout the year, as do the associated changes to the upper ocean structure and thus likely impacts on SST.

2.2 INTRODUCTION

Although much of the ocean is thermally stratified near the surface, there are regions where salinity stratification dominates and can create a density mixed layer (ML) within a much deeper isothermal or thermally inverted layer. The region between the bases of these layers is known as a barrier layer, so termed because it creates a mixing barrier between the base of the ML and the cooler waters of the thermocline (**Figure 2.1**a). BLs were first identified in the Western Pacific Warm Pool (Lukas and Lindstrom 1991) and have subsequently been identified throughout the Indian Ocean (e.g. Sprintall and Tomczak 1992; Qu and Meyers 2005) as well as in subtropical (e.g. Katsura et al. 2015) and high latitude regions (Kara et al. 2000). Since their discovery, it has been suggested that they may enhance ML warming (Lukas and Lindstrom 1991), minimize the impact of atmospheric forcing on sea surface temperatures (SST) (Vialard and Delecluse 1998), or even lead to a decrease in SST (Vinyaychandran et al. 2007), and that they may increase tropical storm intensity (Balaguru et al. 2012; Ffield 2007). This is of particular interest for the high SST in the Arabian Sea that contributes to the onset of the southwest monsoon (Rao and Sivakumar 1999; Masson et al. 2005; Shenoi et al. 1999).

One of the challenges of observational BL studies is that several different definitions are employed for the ML and isothermal layer. Lukas and Lindstrom (1991) used a density gradient criterion ($\leq$0.01 kg m^{-4}) to identify the lower boundary of the ML, and temperature and salinity gradient criteria with roughly equivalent impacts on density to identify the thermocline and halocline. Subsequent work has primarily used a change in temperature to identify the depth of the isothermal layer, using the corresponding density change to denote the ML depth. However, the values used for the temperature threshold vary among 0.2°C (de Boyer Montegut et al. 2014; hereafter CdBM2014), 0.5°C (Vialard and Delecluse 1998), 0.8°C (Qu and Meyers 2005), and

1°C (Durand et al. 2004; Durand et al. 2007; Thadathil et al. 2008; Agarwal et al. 2012). As a result, each of these studies references a different segment of the water column and produces different ML and BL characteristics. Arabian Sea studies alone exhibit a wide range of observations of the frequency and thickness of these layers, differing substantially within a single season. The precise details of the ML definition have implications for estimates of how long BLs persist, what causes their ultimate erosion, and their impact on ML heat budgets; BLs of the same thickness located at different depths or with different degrees of stratification are likely to have different responses to atmospheric forcing. Thus, disparities in the portion of the water column referenced as the "BL" and the characteristics of this layer can be significant.

Due to the near-isothermal (but not isohaline) nature of the upper water column when a BL is present (**Figure 2.1**a), the near-surface temperature-salinity relationship of a BL appears on a *T*-*S* diagram (**Figure 2.1**b) as a roughly horizontal portion of the near-surface profile (**Figure 2.1**b, red segment). To quantify the "flatness" of this tail, we can examine the density ratio

$$R_\rho = \frac{\alpha(T_C)_z}{\beta(S_A)_z} \tag{1}$$

where $\alpha = -\frac{1}{\rho}(\frac{\partial \rho}{\partial T_C})$ and $\beta = \frac{1}{\rho}(\frac{\partial \rho}{\partial S_A})$, ρ is potential density referenced to the sea surface, T_C is conservative temperature, S_A is absolute salinity, and subscript z refers to a vertical derivative (IOC et al. 2010), with z defined as positive upwards. Here R_ρ indicates the relative contributions of temperature and salinity stratification to changes in density (Schmitt, 1981; Schmitt 1990). In a BL, where vertical changes in salinity dominate the changes in density and $(T_C)_z \approx 0$, we expect that $R_\rho \approx 0$. The only exception to this would be in the presence of a relatively strong

temperature inversion, in which case R_ρ would be small and positive [$(T_C)_z$ and $(S_A)_z$ are both negative, but $|\alpha(T_C)_z| < |\beta(S_A)_z|$ for a salinity stratified, stable water column]. Similarly, a small negative density ratio ($0 > R_\rho \gg -1$) would indicate weak, positive temperature stratification with a stable salinity stratification dominating. The density ratio also indicates conditions in which double-diffusion in the form of convective overturning or salt-fingering are likely to occur (Schmitt 1990), both of which are possibilities within BLs. When a temperature inversion occurs, relatively cool, fresh water overlays warmer, saltier water within the BL, resulting in conditions possibly favorable to convective overturning. The bottom of the BL coincides with a region of relatively high temperature and salinity; if the bottom of the BL coincides with the beginning of the main thermocline, as originally defined by Lukas and Lindstrom (1991), the water is cooler and fresher below, setting up conditions possibly favorable to salt-fingering (Schmitt 1990). We do not expect to be able to observe these dynamical processes directly in data from in situ data collected from ships or profiling floats; additionally the effects of shear may be important in many cases (Schmitt 1990), but nonetheless BLs may be susceptible to erosion in this way.

While BLs are not the only phenomenon associated with subsurface salinity maxima, they are unique among upper ocean stratification regimes in mid- and low latitudes in that the saltiness of the water column increases below the ML and peaks near the bottom of a roughly isothermal layer, inviting the use of the variable known as *spice* to describe them. Along an isopycnal, increases (decreases) in temperature must be counteracted by increases (decreases) in salinity (Jackett and McDougall 1985) and spice characterizes those changes in the relative warmth and saltiness of a water parcel (Veronis 1972; Flament 2002). Although the concept of spice has appeared in several different forms since its inception and construction of a full

definition requires a specific reference point, here we will follow the lead of Shcherbina et al. (2009) in referencing the approximate differential form

$$d\tau = \rho(\alpha d\theta + \beta dS) \quad (2)$$

where τ denotes spice, ρ represents the reference density, and the parentheses indicate multiplication. For a given ρ, increases (decreases) in potential temperature, salinity, or both, can lead to an increase (decrease) in spiciness. This is a direct analog of the differential form for density (when neglecting pressure contributions), but with the sign of the $\alpha d\theta$ term reversed.

Vertical variations in spice can indicate the layering of different water types (Shcherbina et al. 2009). Just as BLs occur when $R_\rho \approx 0$ at the base of the mixed layer, they likewise occur when spice increases with depth primarily due to salinity, since spice and density increase concurrently with increasing salinity in an isothermal layer. Temperature inversions can contribute to an increase in spice with depth, but this is only possible in a stable water column if $0 < R_\rho < 1$ (that is, the water column is stabilized by salinity). Thus, BLs in mid- and low latitudes can be characterized as occurring when spice increases with depth (or density) from the surface as shown in **Figure 2.1**c. All other scenarios of relative temperature and salinity stratification lack this feature.

In this study, we aim to remove some of the ambiguity surrounding BL identification and use this to improve the quantitative estimates of BL presence in the Arabian Sea. We do this by proposing a method for identifying the presence of a BL relying on the vertical spice profile. We also address questions of BL formation, persistence, and erosion mechanisms more quantitatively than was previously possible from observational data, using spice to assess variations in temperature and salinity simultaneously. We focus on the Arabian Sea as the study area due to its importance in the monsoon cycle, particularly the elevated sea surface

temperatures observed during the spring intermonsoon, which previous studies have suggested may be influenced by BL (e.g. Durand et al. 2004). The region also exhibits broad variation in the salinity stratification observed throughout the year, rendering of suitably determining the BL somewhat complex.

2.3 DATA

This study uses data from Argo-type floats in the Arabian Sea, here considered to be the region bounded by the Equator in the south and the Persian Gulf in the North, spanning longitudes of 40°E and 80°E (**Figure 2.2**). Initially, all available temperature and salinity profiles from 2004-2016 were considered, including both delayed mode and real time/adjusted profiles. Since not all of these profiles had completed full quality control measures, the utility of each profile was individually assessed. Any profile lacking data in the top 10 meters was discarded, as were profiles with temperature (T) and salinity (S) values outside a reasonable range for this location ($0< T <35$°C; $20< S <40$). Profiles whose maximum pressure was less than 100 meters, or profiles with large gaps in the pressure data, were also excluded from the analysis. After these quality control measures (which resulted in discarding roughly 34% of the original profiles) there were a total of 34,311 useable profiles of temperature and salinity. In contrast with previous studies looking at the seasonal patterns of BLs in the Arabian Sea (e.g. Thadathil et al. 2008), most 2°×2° boxes in the primary study area, the Southeast Arabian Sea (SEAS), have over 100 profiles annually per box, and many have considerably more. Here, we use a slightly larger definition of the SEAS from other studies, focusing on the region between 60°E and 80°E and the Equator to 15°N, as this definition captures the region where BLs are seen most frequently across all seasons (black outlined region in **Figure 2.2**) and may have an impact on the development of the mini-warm pool during the spring intermonsoon.

We used the Monthly Isopycnal/Mixed-Layer Ocean Climatology (MIMOC) conservative temperature and absolute salinity fields to represent long-term averages in the temperature and salinity fields (Schmidtko et al. 2013). Absolute Salinity (S_A) indicates the mass fraction of salt in seawater (g kg^{-1}). Although the differences between conservative and potential temperature, and practical and absolute salinity, are small near the surface where this study is conducted, both conservative temperature (T_C) and absolute salinity are necessary for using the Thermodynamic Equations of State-2010 (TEOS-10) (Feistel 2003; IOC et al. 2010). In this paper, we use TEOS-10 to calculate spice (via the variability of isopycnal temperature and salinity; McDougall and Krzysik, 2015) and potential density (the density a fluid parcel would acquire if it were brought adiabatically to the surface), as well as to convert in-situ temperature and salinity to T_C and S_A. A comparison of the results of this spice calculation with the Flament (2002) formulation yield very similar results. The National Centers for Environmental Prediction/Department of Energy (NCEP-DOE) Reanalysis 2 daily average heat fluxes, precipitation, and wind fields were used to calculate approximate real-time fluxes associated with individual profiles, and NCEP/NCAR Reanalysis 1 monthly long term mean (1981-2010) data were used to represent long term averages of these same quantities to develop a climatology. Both of these are based on state-of-the-art reanalysis/forecast with data assimilation (Kanamitsu et al. 2002).

2.4 Identifying and Describing Barrier Layers

The established method for identifying BLs involves approximating the bottom of a near-isothermal layer with a temperature threshold and the bottom of the density ML (if different) with the corresponding change in density,

$$z_{IL} = z(T = T_r + \Delta T)$$

$$z_{ML} = z(\sigma_\theta = \sigma_{\theta,r} + \Delta T \frac{\partial \sigma_\theta}{\partial T})$$

$$h_{BL} = z_{IL} - z_{ML} \quad (3)$$

where z_{IL} represents the depth of the isothermal layer, T_r is the temperature at a reference pressure (typically 10dbar), z_{ML} is the depth of the ML, $\sigma_{\theta,r}$ is the density at a reference pressure r, and h_{BL} is the thickness of the BL. ΔT is the change in temperature selected to indicate the bottom of the isothermal layer. As noted earlier, we find a range of values between –0.2°C and –1.0°C being used for ΔT in previously published BL studies in the Arabian Sea. Also, seasonal variations in the strength of stratification within the BL suggest that a particular threshold may not capture comparable segments of a water column in different seasons, making comparisons between seasons less meaningful. There are also situations in which temperature and salinity contribute nearly equally to the density stratification in such a way as to create a mismatch between the isothermal layer and ML as defined above, without necessarily signifying a genuine BL.

An alternative approach is to examine the properties of the vertical spice profile. For a BL to occur in a gravitationally stable profile, a low spice, low density water mass must be layered above a high spice, higher density water mass (**Figure 2.1**); the bottom of the BL corresponds to a transition to a regime in which thermal stratification dominates, therefore returning to a lower spice regime (**Figure 2.1**c). Each of these spice transitions occurs diapycnally, as the BL is defined by a change in density (**Figure 2.1**a). To identify suitable layer definitions for the mixed and isothermal layers, we can therefore look for inflection points in the profile of spice vs. density, indicating transitions between stratification regimes. In doing so, we are identifying just the portion of the water column where salinity stratification controls the density as the BL; this excludes profiles where temperature and salinity contribute equally, and

does not require the pre-selection of a temperature threshold or gradient that is compatible with the strength of the salinity stratification for identifying BL. We use a peak detection method as discussed by Shcherbina et al. (2009), who referred to the curvature in a spice vs. density profile as diapycnal spiciness curvature, $\tau_{\sigma\sigma}$, where the subscripts denote derivatives with respect to potential density. High absolute values of $\tau_{\sigma\sigma}$ correspond to sharp diapycnal transitions in water mass properties (Shcherbina et al. 2009). A positive $\tau_{\sigma\sigma}$ peak indicates a transition point from lower to higher spice (as happens at the base of the ML when a BL is present), whereas a negative peak indicates a transition from higher to lower spice (as happens at the bottom of the barrier/isothermal layer). These patterns can be observed qualitatively in **Figure 2.3**, where a clear low-spice signal and high $\tau_{\sigma\sigma}$ dominate the mixed layer, whereas the reverse is true within the BL. This is similar in spirit to the method developed by Lorbacher et al. (2006) for ML identification, but relies on the spice profile rather than the temperature profile.

In our analysis we linearly interpolate each spice profile to a 0.01 kg m^{-3} grid and then apply a Gaussian filter with a standard deviation of 0.01 kg m^{-3} to remove some of the noise. For older float profiles with lower vertical measurement resolution this might not be necessary, but this method is effective for applying the same algorithm to all available measurements. Using a smaller grid (0.005 kg m^{-3}) produces a greater than 97% agreement in both BL identification and thickness; the choice of grid primarily needs to be small enough to resolve the density features. For comparison to the threshold method, $\Delta T = -0.2$°C translates to $\Delta\rho \approx 0.06$ kg m^{-3} for the region considered here. Any profiles where a maximum in spice was found in the two uppermost measured values of *T* and *S* were labelled non-BLs and excluded from the remainder of the algorithm; for these profiles, a temperature threshold of $\Delta T = -0.2$°C was found to be sufficient for identifying the base of the ML in this region. $\tau_{\sigma\sigma}$ was calculated using a central difference

formula and smoothed again using the same Gaussian filter as described above to remove noise introduced by taking derivatives. A profile that has the spice characteristics of a BL has a positive first peak in $\tau_{\sigma\sigma}$, corresponding to a transition from low spice to high spice; this is the base of the ML (orange line, **Figure 2.4**b). A deeper negative peak, corresponding to a transition from high spice to low spice indicates the base of the isothermal layer. When more than one possible negative peak existed, the peak nearest the depth of maximum spice was used as this represented the depth to which the salinity stratification dominates density. **Figure 2.4**b shows a profile with a weak negative $\tau_{\sigma\sigma}$ peak much shallower than the maximum spice, and another much larger peak at the same depth as the maximum spice; the latter was selected as the base of the BL. Once the two peaks were identified in the spice vs. density profile, the corresponding pressure levels were then used as the ML depth and the equivalent of the isothermal layer depth. The BL thickness is defined as in Equation 3. The results of this method are shown on a standard depth profile in **Figure 2.1**a. Due to the nature of the data used, which has limited measurements within the top 10m, the analysis here focuses primarily on monthly and seasonal patterns. However, if data with sufficient near-surface resolution were available, this method could also be applied to explore shorter time scales without major modification as it would provide the depths of changes in stratification.

In previous studies, the quantity *porosity* has been used as a measure of the longevity and spatial extent of BLs (Mignot et al. 2009; CdBM2014). Here, instead, we quantify the frequency with which BLs are observed by dividing profiles into 2°×2° boxes and calculating the percentage of profiles in a given box that are identified as BLs (effectively 100% - *porosity*). Mignot et al. (2009) used 5 profiles per 2°×2° box as their minimum number for robust statistics; the data coverage available here easily meets this minimum in the SEAS outside of coastal areas.

Figure 2.5 shows the BL frequency by season. In comparing this with the spatial distribution of BLs shown in other studies (Agarwal et al. 2012; CdBM2014), we find that the results are generally most similar during the northeast monsoon. We find that the BLs reach farther west than suggested by some others, who found very few BLs west of 68°E (Agarwal et al. 2012). Differing spatial distributions of the datasets used may account for the nuances in some of the differences. In comparison with Thadathil et al. (2008), we find that the distribution of BLs is shifted towards the southeast, where little data were available at the time of that study. In comparison with CdBM2014 we find evidence of BLs farther north and east than in their South Central Arabian Sea study box.

We also calculate BL characteristics for comparison to previous studies. BLs in the Arabian Sea demonstrate a seasonal cycle that strongly correlates with the southwest and northeast monsoons, with peaks in thickness in February and August. Figure 6b shows the thickness and frequency averaged over the entire basin and suggests that the peak average thickness of BLs in February and August are roughly comparable, with values of approximately 35 meters. Previous studies have identified values ranging from 5-60m for the southwest monsoon (CdBM2014 and references therein), and 30-50m for the northeast monsoon (Shenoi et al. 2004; Durand et al. 2004; Agarwal et al. 2012). In this regard, the spice method produces results that fall within the range of previously calculated values. Comparing **Figure 2.5** to previous plots of barrier layer thickness (CdBM2014; Agarwal et al. 2012; Durand et al. 2007) indicates that the barrier layers found here tend to have the most similar distribution and thickness during the northeast monsoon, although we do not find regions with BL thickness > 60m, when averaged in a 2°×2° box; during the southwest monsoon, we find them to be thicker and existing farther west. A major difference between the results here and those reported

previously is that the depths calculated for the ML and BL here specifically indicate transitions in the stratification regime.

BL thickness does not fully capture the strength of BLs, important when considering their impact and ultimate erosion. Chi et al. (2014) suggested BL potential energy (P_{BL}) as a more suitable measure of the strength of a BL. P_{BL} incorporates the thickness of a BL, but also considers ML depth and strength of stratification within the BL. It is defined as $P_{BL} = P_{mix} - P_0$, where

$$P_{mix} = -g \int_{z_{ILD}}^{0} \overline{\rho(z)} z dz = -\frac{1}{2} g \overline{\rho(z)} \cdot z_{ILD}^2 \tag{4}$$

and

$$P_0 = -g \int_{z_{ILD}}^{0} \rho(z) dz = -g \left(\int_{z_{MLD}}^{0} \rho(z) z dz + \int_{z_{ILD}}^{z_{MLD}} \rho(z) z dz \right) \tag{5}$$

Here P_{mix} is the potential energy the water column would have after mixing downwards to the base of the isothermal layer, and P_0 is the initial potential energy of the water column. The overbar indicates a vertical average. A large positive value of P_{BL} would indicate a large increase in potential energy associated with mixing away the BL, requiring significant input of kinetic energy. This formulation gives a better quantitative estimate of the relative difficulty of eroding BLs through the input of mechanical energy than the thickness. **Figure 2.6**c shows that despite comparable BL thicknesses in February and August, ML depths are shallower yet the overall P_{BL} is substantially higher during February. **Figure 2.6**d-f focus exclusively on the Southeast Arabian Sea; the results show similar patterns to **Figure 2.6**a-c, consistent with the fact that more than two-

thirds of the BL profiles observed are in the SEAS. These results suggest that the new BL identification method described here produces reasonable results.

2.5 Formation of Barrier Layers

Applications of spice to BLs are not limited to identifying these layers. **Figure 2.7** illustrates the possible mechanisms by which BLs could be formed (Cronin and McPhaden 2002, hereafter CM2002), illustrated in terms of the necessary spice and density conditions of the water parcels involved. In **Figure 2.7**a, a horizontal spice front (τ_{low} and ρ_{low} to the left; τ_{high} and ρ_{high} to the right) is tilted into the vertical by a vertical shear in the horizontal velocity. The simultaneous low spice and low density come about due to a lower salinity water mass on the left, uncompensated by temperature. After tilting, the vertical profile at the original front location will have low spice above higher spice, necessarily setting up a BL. **Figure 2.7**b shows a low spice, low density, water mass being advected into a region of higher spice and density, establishing a BL where the two water masses overlap. The reverse process, where high spice, high density water is advected and ultimately subducted under the lower spice and lower density water mass would also form a barrier layer. The final formation mechanism comes about as a result of rainfall, illustrated in **Figure 2.7**c. As with the previous two cases, the surface water in this case must have lower spice and density than the water mass on which it is superimposed in order to form a BL. In all cases, low salinity largely controls both the density and the spice. Cooler water can contribute to the low spice, but in order to establish the necessary stable vertical profile, $\left|R_\rho\right| < 1$.

To explore these formation processes more quantitatively, we can first examine the vertical derivatives of the temperature and salinity tendency as done by CM2002,

$$S_{zt} = -\boldsymbol{U} \cdot \nabla S_z - wS_{zz} - \boldsymbol{U}_z \cdot \nabla S - w_z S_z - (\overline{w'S'})_{zz} \quad (6)$$

$$T_{zt} = -\boldsymbol{U} \cdot \nabla T_z - wT_{zz} - \boldsymbol{U}_z \cdot \nabla T - w_z T_z - (\overline{w'T'})_{zz} + \frac{1}{\rho c_p}(Q_{rad})_{zz} \quad (7)$$

In these equations $\boldsymbol{U}$ represents the horizontal velocity, w the vertical velocity, and ∇ the horizontal gradient operator. $(\overline{w'S'})_{zz}$ is the vertical turbulent flux of salinity, equal to $S_0(P - E)$ at $z = 0$, where P and E indicate precipitation and evaporation and S_0 is the surface salinity. $(\overline{w'T'})_{zz}$ represents the vertical turbulent flux of temperature. $(\overline{w'T'})_{zz}$ at $z = 0$ represents the contributions of latent, sensible, and longwave radiation, while Q_{rad} represents the shortwave radiation.

This formulation highlights the processes shown in **Figure 2.7** in terms of the temperature and salinity tendencies, including the advection of freshwater (term 1 on the right side of each equation), rain (term 5 on the right side of each equation), and tilting of a horizontal salinity gradient through vertical shear (term 3 on the right side of each equation) (CM2002). As pointed out in CM2002, the terms involving vertical velocity do not contribute to the initial formation of BLs; however, w_z can stretch a BL that already exists and w can vertically displace an existing BL without altering its structure otherwise. The reverse of each of these processes can also contribute to BL erosion, including through mechanisms such as upwelling and advection of high salinity water (Shenoi et al. 2004).

Rather than considering salinity and temperature separately, we return to the consideration of spice as an indicator of a BL, consistent with the schematic representation of **Figure 2.7**. We use a linear combination of Equations (6) and (7) above to examine the vertical derivative of the spice tendency, τ_{zt}, (neglecting variations in α and β), to find that

$$\tau_{zt} = \rho(\alpha\theta_{zt} + \beta S_{zt}) \quad (8)$$

Substituting equations (6) and (7) into equation (8) yields:

$$\tau_{zt} = \rho(\alpha(-U \cdot \nabla T_z - wT_{zz} - U_z \cdot \nabla T - w_z T_z - (\overline{w'T'})_{zz} + \frac{1}{\rho c_p}(Q_{rad})_{zz}) + \beta(-U \cdot \nabla S_z - wS_{zz} - U_z \cdot \nabla S - w_z S_z - (\overline{w'S'})_{zz}))$$

(9)

or

$$\tau_{zt} = \rho(-U \cdot (\alpha \nabla T_z + \beta \nabla S_z) - w(\alpha T_{zz} + \beta S_{zz}) - U_z \cdot (\alpha \nabla T + \beta \nabla S) - w_z(\alpha T_z + \beta S_z) - (\alpha((\overline{w'T'})_{zz} - \frac{1}{\rho c_p}(Q_{rad})_{zz}) + \beta(\overline{w'S'})_{zz})$$

(10)

In final form, this yields

$$\tau_{zt} = -U \cdot \nabla \tau_z - w\tau_{zz} - U_z \cdot \nabla \tau - w_z \tau_z - \rho(\alpha((\overline{w'T'})_{zz} - \frac{1}{\rho c_p}(Q_{rad})_{zz}) + \beta(\overline{w'S'})_{zz})$$

(11)

BLs are the only gravitationally stable condition with $\tau_z < 0$ near the surface, since the change in salinity with depth that produces increasing spice (βdS) also increases the density. When considering a mechanism such as the tilting of a salinity front by vertical shear of the horizontal velocity (term 3 on the right side of Equation (11)), identifying a salinity front is therefore insufficient for locating likely areas of BL formation. A salinity front fully compensated by temperature (cool on the fresh side, warm on the salty side) would not establish any density stratification when tilted into the vertical, while a salinity front aligned with a more significant temperature front would not lead to vertical stratification dominated by salinity.

Therefore, regions of BL formation by this method must have a spice front (fresher on one side, saltier (spicier) on the other) aligned with a density front such that

$$\frac{\partial \rho}{\partial \tau} = \frac{-\alpha dT + \beta dS}{\alpha dT + \beta dS} > 0 \qquad (12)$$

across the spice front. When the front is tilted, the less spicy (less dense) water overlays the spicier (more dense) water, as shown by **Figure 2.7**a, producing the slope in the density-spice plot observed in **Figure 2.1**c. (Were the front tilted the other direction by the shear, it would create a gravitationally unstable stratification, not a BL). Creation of new BLs by advection of an existing BL (term 1 on the right side of Equation 11) requires $\tau_z < 0$ in the source water. Advection of a low salinity water mass into a high salinity region (more akin mathematically to term 3 on the right side of Equation 11) requires only a difference in spice and density and horizontal velocities. At the point of formation, $\tau_z < 0$ either due to entering a region with warmer background temperatures or cooling of the surface layer (**Figure 2.7**b). Therefore, in order for a BL to form, τ_{zt} must be negative because any non-BL configuration will have $\tau_z \geq 0$.

Previous work has suggested likely formation mechanisms, primarily derived from modeling results. We use the spice framework outlined by Equation 11 to investigate climatological patterns to infer whether these mechanisms are consistent with the observed patterns in spice. For the advection mechanism (term 1), **Figure 2.8** shows climatological values of τ_z averaged over the top 50 meters seasonally, based on the MIMOC climatology. We find a strongly negative vertical derivative of spice, $\tau_z < 0$, in the region of greatest BL prevalence during the northeast monsoon (DJF; **Figure 2.5**), indicating an increase in spice with depth (**Figure 2.8**). This pattern begins in the Bay of Bengal during the previous southwest monsoon (August) and gradually spreads into the Arabian Sea, peaking during the northeast monsoon. Prevailing near-surface currents in the SEAS are generally to the north and west during the

northeast monsoon (Schott and MacCreary 2001; Beal et al. 2013), suggesting advection of more negative τ_z into regions of less negative τ_z. This corresponds to term 1 on the right side of Equation (11), and is consistent with previous work identifying advection of fresh water from the Bay of Bengal within the isothermal layer as the primary mechanism for BL formation in the SEAS during the northeast monsoon (Durand et al. 2007). That τ_z appears strongly negative in the SEAS data on climatological scales indicates the dominance of BLs over a more traditional stratification regime in this region. Only isolated patches of $\tau_z < 0$ appear in other seasons, consistent with the lack of large freshwater input that sets up the necessary vertical patterns on a basin scale.

Previous work on BL formation through frontal tilting (term 3 in Equation 11) has been focused primarily on the location of salinity fronts (e.g. CdBM2014) or the relative magnitude of horizontal salinity and temperature changes (Katsura et al. 2015). Examining the locations and alignment of spice and density fronts provides more complete information about whether the conditions meet those outlined in **Figure 2.7**a and Equation 12. Although we expect the position of fronts to move on a range of time scales, climatological data provide insight into whether those conditions are typically met. **Figure 2.9** shows the average location of surface spice and density contours based on MIMOC data. In the South Central Arabian Sea during the southwest monsoon (August) where frontal tilting has been previously discussed as a likely mechanism, (CdBM2014), we do find that the surface spice and density contours are aligned (notably between 65ºE and 69º E, around 7ºN), with density and spice increasing in the same direction. Analysis on shorter time scales using satellite data shows similar results, with frontal alignment frequently coinciding with measurement of BLs by Argo floats (not shown). Due to the strong influx of freshwater from the Bay of Bengal alternating with the advection of high salinity water

from the Northern Arabian Sea, we expect strong fronts in both spice and density to occur in part of the basin during all seasons, and thus this mechanism may contribute to BL formation throughout the year.

Rain has been examined as a mechanism of BL formation in 1-dimensional modeling studies (e.g., You 1998). The final group of terms in Equation 11 constrain the conditions in which a BL could be formed due to surface freshwater forcing when used in conjunction with a stability scaling, such as the Monin-Obukhov length for the ocean (Equation 13). The length can be interpreted as the length scale at which vertical wind mixing and atmospheric (buoyancy) forcing balance one another, with positive values indicating a stable water column. We use the definition from Large (1998) for this length, L, in the ocean

$$L = \frac{u_*^3}{\kappa B_0} \tag{13}$$

where $$B_0 = -g(\alpha\overline{(w'T')}_0 - \beta\overline{(w'S')}_0) \tag{14}$$

$$u_* = \sqrt{\frac{\Upsilon}{\rho}} \tag{15}$$

Here Υ is the wind stress and κ is the von Karman constant (0.4). From the definitions for $\overline{(w'T')}_0$ and $\overline{(w'S')}_0$ described above, we derive the expression for buoyancy flux

$$B_0 = B_H + B_W = \frac{g}{\rho_0}\left[\frac{\alpha Q_0}{c_p} + \rho_0 \beta S_0 (P - E)\right] \tag{16}$$

where B_H and B_W indicate the buoyancy forcing due to heat and freshwater flux, respectively, $(P - E)$ is expressed in units of m s^{-1} and B_0 has units of m^2 s^{-3}. B_0 indicates the rate at which buoyancy is removed from the water column or available potential energy is supplied due to surface heat and freshwater fluxes. In the absence of atmospheric heat and freshwater fluxes, the denominator approaches zero and $L \to \infty$, indicating indefinite deepening of the ML. Positive

buoyancy forcing is associated with processes that increase the stability of the ML (e.g. rain or heating) whereas negative buoyancy forcing is associated with processes that decrease the stability of the ML (e.g. evaporation or cooling). Within the classification as stable or unstable, we can attribute the stability (instability) to specific mechanisms: excess heating (cooling), excess rain (evaporation), or a combination of both. The combination of excess heating and rain will always lead to more stable conditions, just as the combination of excess cooling and evaporation will always lead to decreased stability. Simultaneous heating and evaporation or cooling and rain could cause stability to tend in either direction, depending on which forcing mechanism dominates. Stability requires that the buoyancy forcing term be positive, which further implies that

$$-\alpha \frac{Q_0}{\rho_0 c_p} < \beta S_0 (P - E) \tag{17}$$

establishing the relative allowed amounts of rain and cooling or heating and evaporation that will permit a stable upper ocean. To form a BL from buoyancy flux, *P*-*E* must be positive (i.e. rain dominates), as must the final term from Equation 11 above, indicating changes to the vertical spice profile due to turbulent fluxes of freshwater and heat

$$(\alpha((\overline{w'T'})_{zz} - \frac{1}{\rho c_p}(Q_{rad})_{zz}) + \beta(\overline{w'S'})_{zz}) > 0. \tag{18}$$

A positive value for Equation 18 would produce a negative tendency in τ_Z as required. Between the considerations of stability (Eq. 17), necessary freshwater forcing (*P*–*E* > 0), and the constraints on the spice tendency (Eq. 18), these equations define the warming or cooling that can occur in conjunction with rain if a BL is to form. Although positive heat flux lends itself to stability, this effect must be minor in comparison to the freshwater flux for a BL to form. At the other extreme, however, excessive cooling at the surface will lead to gravitational instability.

Figure 2.10 shows regions in which rainfall may lead to BLs in the Arabian Sea, highlighted in lightest green ($B_{heat} < 0$, $B_{P\text{-}E} > 0$); these are the regions in which cooling and freshening occur together but the freshening contributes more to changes in density, permitting stability. On a seasonal basis, the regions where BL development might be possible via rain are in the near-equatorial (medium and light green in **Figure 2.10**) regions during the southwest monsoon. The medium green areas in **Figure 2.10**($B_{heat} > 0$, $B_{P\text{-}E} > 0$), showing regions of comparable rain and heating on average, may allow for BL formation if the rain overlaps with transient cooling events, such as night-time cooling, cooling during a storm, relatively cold rainwater falling, or if the heating is very mild. In addition, the presence of BLs may insulate the ML from the effects of strong atmospheric heating by allowing increased heating below the ML by shortwave radiation (Vialard and Delecluse 1998). Regardless, it is unlikely that freshwater influx from rain could establish the same intense stratification observed due to other mechanisms, and therefore BLs formed in this way are expected to be short-lived. Thus, it seems clear that the dominant mechanisms forming BLs in this region are two-dimensional: primarily advection of freshwater and previously established BLs during the northeast monsoon (Durand et al. 2007), and restratification due to tilting of frontal regions during the summer monsoon (CdBM2014).

2.6 Stability and Erosion of Barrier Layers

The upper ocean stability described by the Monin-Obukhov length also provides insight into the possible fate of BLs once formed. This is of interest because their duration and demise dictate their likely impact on SST. **Figure 2.11**shows the seasonal distribution of BLs superimposed on the region of expected stability ($L > 0$, hatched region), inferred by using NCEP/NCAR Reanalysis 1 Long Term mean atmospheric fluxes to estimate B_0 in Equation (16). We observe that BLs are most frequently found within the stable regions during the northeast

monsoon (DJF); with the exception of several small patches in regions with low BL frequency near 60ºE, BLs found in the stable region also tend to be thicker than outside of it (**Figure 2.5** contours; **Figure 2.11** line plots). During the southwest monsoon (JJA), BLs are often found outside of the stable region, and tend to be thicker towards the northwest (**Figure 2.5** contours; **Figure 2.11** line plots). This may in part be due to the deeper MLs typically found farther north, which would establish a deeper isothermal layer as the background for the BL prior to frontal tilting (i.e., the initial depth of the ML on the right side of the front in **Figure 2.7**a would be deeper, leaving a deeper lower layer after tilting).

The profiling float dataset cannot in general be used for examining possible two-dimensional mechanisms for eroding BLs. Shenoi et al. (2004) used a model to suggest the role of upwelling in eroding BLs at the end of the northeast monsoon. Here, we focus on the one-dimensional turbulent fluxes, the final term in the temperature, salinity, and spice tendency equations (Equation 11). We examine the range of conditions that might erode a BL in order to suggest some broad patterns about likely mechanisms and the associated BL duration.

The limited duration of BLs derived from porosity calculations during the southwest monsoon are sometimes attributed to strong winds (e.g., CdBM2014). While winds in the northwestern portion of the basin are quite strong climatologically, particularly within the Findlater Jet (Schott and McCreary 2001), very few BLs are observed in that part of the basin. The winds in the dominant region of BL formation during the southwest monsoon are of a similar magnitude to those found during the northeast monsoon. However, **Figure 2.6** showed that the BL potential energy, P_{BL}, is roughly twice as large during the northeast monsoon (despite similar BL thickness), suggesting significantly greater resistance to wind-driven erosion. Input of turbulent kinetic energy to overcome the BL potential energy (P_{BL}) will have a greater effect

with lower P_{BL} and higher kinetic energy input. We can estimate the rate of wind driven erosion by using a parameterization of ML deepening (McPhaden and Hayes 1991) as

$$w_e = \frac{2m(\frac{\rho_a C_D}{\rho_0})^{3/2}|U|^3}{-g\left(\frac{\Delta\rho}{\rho_0}\right)h} \quad , \tag{19}$$

where m is empirically determined to represent efficiency (here we use a value of 1.25, as discussed in the work of Niiler and Kraus (1977); Lee et al (2000) use a different formulation for w_e and a smaller value of m, but the two methods yield similar magnitudes for w_e), $\Delta\rho$ is the jump in density across the base of the ML, g is gravity, and ρ_a, C_D, and U are the air density, drag coefficient, and wind velocity respectively. This parameterization only considers wind work, neglecting shear and buoyancy effects. We can create a separate estimate directly from float profiles by examining the change in ML depth between consecutive profiles, $w = \frac{dh}{dt}$, which would potentially incorporate changes in ML depth due to a variety of factors. Consecutive profiles are identified by sorting all float data by WMOID number, and then sorting by date, allowing examination of all profiles for an individual float in sequence. For both cases, we only use values that indicate ML deepening, since shoaling of the ML would not contribute to BL erosion by wind. We then calculate the expected duration of the BL with

$$\Delta t = \frac{h_{BL}}{w} \tag{20}$$

where Δt is the duration of the BL, h_{BL} is the thickness of the BL, and w is the rate of ML deepening. This is done for each individual BL profile, and then averaged by month. **Figure 2.12**a shows the expected duration of BLs in the SEAS using Equation 20 and the different estimates of w. Although the estimates vary in magnitude, all suggest peak BL duration at the end of the northeast monsoon when P_{BL} peaks, and a minimum duration some time during the southwest monsoon (although timing differs between the methods). The w calculation made

directly from float profiles (green line) leads to consistently longer expected durations during the northeast and southwest monsoon. When comparing with the observed frequency (**Figure 2.5**), the period of longest duration (the northeast monsoon) matches the period when BL are most frequent; in addition, durations on the order of 20-30 days are compatible with an observed BL frequency exceeding 60%.

Wind mixing is not the only possible one-dimensional mechanism for BL erosion. From the examination of stability conditions in the region, it is expected that much of the basin will be unstable due to buoyancy forcing during the summer and northeast monsoons (**Figure 2.10**). During the two intermonsoons, however, nearly the entire basin is characterized by weak winds and strong positive buoyancy forcing, primarily supplied by atmospheric heat flux. In the latter case, neither buoyancy loss nor wind-driven mixing is expected to contribute significantly to BL erosion. However, changes to the spice profile dictated by increasing temperatures at the surface could result in a thermally stratified surface layer, effectively removing the BL. **Figure 2.3** shows several short-lived events in which the ML depth remains roughly the same but SST increases sufficiently to produce a dominant temperature stratification and high surface spice. Basin-wide temperature increases occur during both intermonsoon periods, with temperatures reaching their annual maximum prior to the onset of the southwest monsoon (Rao and Sivakumar 1999).

To perform a comparable analysis to that done with wind erosion (**Figure 2.12**a), we define two BL characteristics, the buoyancy deficit and the spice deficit (denoted as Θ and Π), as

$$\Theta = \frac{g}{\rho_0}\int_0^{z_{ILD}}(\sigma(z_{ILD}) - \sigma)dz \tag{21}$$

$$\Pi = \frac{g}{\rho_0}\int_0^{z_{ILD}}(\tau(z_{ILD}) - \tau)dz \tag{22}$$

Θ and Π both have units of $m^2\ s^{-2}$. Θ indicates the buoyancy loss necessary to eliminate the BL, homogenizing the water column down to the isothermal layer depth. Π indicates the spice gain required to overcome the effects of the salinity stratification; this does not necessarily lead to homogenization of the water column as it could occur due to a strong temperature increase at the surface, simply creating a shallow thermally stratified surface layer. We also define a parameter analogous to the buoyancy forcing to denote the effect of atmospheric forcing on the spice,

$$\Sigma_0 = \Sigma_H + \Sigma_W = \frac{g}{\rho_0}\left[\frac{\alpha Q_0}{c_p} - \rho_0 \beta S_0 (P - E)\right] \quad (23)$$

Σ_H and Σ_W indicate the spice flux due to heat and freshwater flux. The only difference between B_0 (Equation 16) and Σ_0 (Equation 23) is the sign of the effect of the freshwater flux, consistent with its opposite effect on buoyancy and spice. Θ and Π can both be calculated for any BL profile, but to examine erosion times, the former is relevant for conditions in which surface buoyancy loss is occurring, the latter for situations where the ocean boundary layer is stable but the surface forcing will act to increase the surface spice. An increase in salinity at the surface would affect the buoyancy as well as the spice; however, it is the resulting changes in density that would be responsible for changes to the BL, and in this case we would focus on buoyancy forcing as a driver of BL erosion. This encompasses all forcing conditions except those associated with BL formation during simultaneous rain and cooling.

In cases where the overall buoyancy forcing is negative (establishing conditions favorable to convective overturn), we therefore estimate the time required for this to provide the necessary buoyancy loss in the BL as

$$\Delta t = \frac{\Theta}{B_0} \quad . \quad (24)$$

In cases where the spice forcing, Σ_0, is positive, we estimate the time required to provide the necessary spice gain in the BL as

$$\Delta t = \frac{\Pi}{\Sigma_0} \quad . \tag{25}$$

Determination of which equation is appropriate for estimating BL erosion time is based on float location relative to the regions in Figure 10. The resulting estimates are for distinct sets of profiles and forcing conditions within each month.

Figure 2.12b shows the calculated monthly erosion times derived from Equations 24 and 25, using gridded daily fluxes interpolated to the position of the profiles. We exclude any estimated durations longer than the maximum duration observed in the float data (~120 days) from the analysis. This eliminates single profiles that wildly skew the trend and does little to alter the overall patterns observed. As with wind-driven ML deepening, erosion due to buoyancy and spice forcing mechanisms both show some seasonality. Since most of the basin is dominated by positive heat fluxes in March, erosion time via buoyancy loss is quite long at that time. By contrast, the intermonsoon periods, which include March, April, and October, are times of the most rapid erosion by means of the spice mechanism (on the order of 10 days in October, for example). Regardless of mechanism, BLs are likely to persist longer during the northeast monsoon than the southwest, as all estimates of duration shown in **Figure 2.12** tend to be larger during the northeast monsoon, reaching the near-monthly timescales one would expect from their high prevalence in the SEAS. This reflects not only differences in the atmospheric forcing but also the BL characteristics shown in Figure 6c and 6f, and is consistent with float data showing that BLs are more likely to be observed on successive profiles during the northeast monsoon.

While the calculations that produced **Figure 2.12** provide insight into seasonal patterns of BL duration based on possible erosion mechanism, assessing an actual erosion mechanism associated with a specific profile is complicated. As a water column becomes vertically homogenized, we expect the magnitude of the $\tau_{\sigma\sigma}$ peaks to decrease. In the case of a purely

wind-eroded BL, we would expect to see a gradual diminishing of the positive $\tau_{\sigma\sigma}$ peak signaling the base of the ML, as the ML and BL waters were entrained together, accompanied by ML deepening. If eroded by spice, we also expect a decrease in the magnitude of the $\tau_{\sigma\sigma}$ peak as thermal stratification takes over, but without the same ML deepening. This is qualitatively visible in **Figure 2.3**. The more typical 5- or 10-day interval between measurements for most Argo-type floats does not allow for this type of detailed examination on short time scales. However, by comparing ML depths during and after the presence of a BL we can examine a range of scenarios. In both wind-driven erosion as well as erosion due to buoyancy loss (or a combination of the two), we expect to see significantly deeper MLs once the BL is eroded, consistent with the expected homogenization (in salinity and density) of the isothermal layer. However, if the BL erodes due (primarily) to increased temperature and therefore spice in the ML, we do not anticipate the same ML deepening unless also in the presence of substantial winds.

We therefore examine the ML depth after BL erosion in comparison with the ML and isothermal layer associated with the BL (**Figure 2.13**). We assume negligible horizontal advection between measurements and therefore that observed changes are changes to the water column itself and not the measurement of an entirely different portion of the ocean. Comparing ML depth during and immediately after the existence of a BL (**Figure 2.13**a, c), a value of one indicates no change to ML depth during BL erosion. This suggests a mechanism for BL erosion through warming in the ML that was not associated with ML deepening (i.e. increased surface spice). We find this to be most commonly the case during the fall and spring intermonsoon periods (**Figure 2.13**c), as expected from the previous analysis. These are times when strong heat flux and low winds dominate, and therefore it is unsurprising that the ML deepening is minimal and ML depths remain

much shallower than the previously existing isothermal layer (values ~0.5, **Figure 2.13**d). The most dramatic increase in ML depth upon BL erosion occurs during the southwest monsoon. This is also the season in which final ML depth most closely approximates the previous isothermal layer depth (**Figure 2.13**d), suggesting complete erosion of the BL and the restoration of deep MLs. This is compatible with the short durations expected with wind driven erosion. The changes in ML density accompanying the patterns in ML depth support this assessment of likely mechanisms (not shown). In both cases, the scatter plots make the spread of the data appear quite large (**Figure 2.13**c, d). However, from the histograms (**Figure 2.13**a, b) we see that the majority of the data is clustered near the median values discussed above, pointing to the relevance of using the median to describe the seasonal trends.

2.7 Summary and Discussion

This study focuses on the Arabian Sea due to its potential impact on the nature of the southwest monsoon. The results here are consistent with many previous studies, but also provide more detailed observational evidence for the presence and characteristics of barrier layers. We take advantage of the increased data provided by the Argo program and use spice as a key descriptor of barrier layers. However, the methods here could be readily applied to other low- and mid-latitude regions of interest, and preliminary work in this regard captures the expected features in the Bay of Bengal, Western Pacific Warm Pool, and the Amazon River plume, including likely formation mechanisms.

The concept of spice has been employed previously to expand insight into mixing processes in the ocean (Veronis 1972; Cole and Rudnick 2012), as well as to describe interleaving and double-diffusive processes (Flament 2002; Shcherbina et al. 2009). The present

work differs from the previous studies in the type of data used (profiling floats instead of gliders and towed instruments), as well as the portion of the water column serving as the primary focus (ML instead of below the ML). Although certainly of interest when considering the duration and fate of barrier layers, using spatial and temporal changes in diapycnal spiciness curvature, $\tau_{\sigma\sigma}$, to explore mixing becomes more complicated when examining the part of the ocean exposed to the atmosphere. However, using $\tau_{\sigma\sigma}$ of a single profile to identify barrier layers has several advantages over the standard threshold method used for identifying barrier layers. The central advantage is that the stratification that accompanies barrier layers has a distinct spice signature. In addition, using $\tau_{\sigma\sigma}$ avoids some of the subjectivity inherent in selecting an "appropriate" threshold, and relates to water mass evolution (Shcherbina et al. 2009), a feature that may provide insight into barrier layer evolution and behavior in the future. Unlike the threshold method, this method explicitly focuses on the barrier itself via the use of a diapycnal characteristic; thus, although the average barrier layer thicknesses identified are comparable to those found in previous studies, we suggest that there is more physical meaning to the individual values (for thickness and other characteristics) than in cases reliant on a particular threshold. It is also clear from the results here that thickness alone cannot adequately describe a barrier layer, and that other measures that characterize the resistance to erosion by various methods (e.g. barrier layer potential energy, spice deficit, and buoyancy deficit) provide more insight.

On rare occasions, the base of what is traditionally termed the "isothermal layer" may not coincide with the beginning of a true thermocline. Since $\tau_{\sigma\sigma}$ captures diapycnal changes in spice, an isopycnal layer below the barrier layer (i.e. a region with both constant temperature and constant salinity) will not be included in the barrier layer. This may help to explain some of the higher ratios seen in **Figure 2.13**. In some cases, the ML depth after barrier layer erosion is up to

three times deeper than the previously identified “isothermal” layer. In many cases, the means of erosion (e.g. surface warming) is such that the post-barrier layer ML depth coincides closely with that of the barrier layer ML, and such deeper low-stratification regions are not relevant for erosion, nor is the main thermocline accessible as a result of the disappearance of the barrier layer. This is important when considering the impact of barrier layers on SST.

This work does not address all possible aspects of barrier layer formation and erosion; specific to the Arabian Sea, we have neglected consideration of the downwelling and upwelling in the Lakshadweep region that contribute to moving the depth of the thermocline throughout the year (Shankar et al. 2004; Shenoi et al. 1999; Shenoi et al. 2004), and more generally have not considered two-dimensional erosion mechanisms. However, one can easily imagine how the analysis that produced Figure 8 for barrier layer formation might produce a similar result for barrier layer erosion under the appropriate circumstances: by reversing the desired sign of τ_{zt} to positive, we can identify situations in which water with high spice near the surface, or an existing barrier layer with less negative τ_z, is carried into an area with barrier layers. In the Arabian Sea, this occurs when the currents switch direction and high salinity water with $\tau_z > 0$ is advected from the Northern Arabian Sea to the south starting during the spring intermonsoon. This mechanism may work in tandem with the one-dimensional mechanisms explored here, particularly if strong mixing with higher salinity water occurs (Shenoi et al, 2004). The spice framework thus allows for the identification and analysis of barrier layers using limited datasets that we expect to be applicable to other regions of the ocean beyond the Arabian Sea.

2.8 ACKNOWLEDGMENTS

We thank K. Drushka and M. Cronin, as well as two anonymous reviewers, for their constructive feedback and suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant DGE-1762114. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also gratefully acknowledge the generous support of Office of Naval Research through the NASCar grant N00014-15-1-2254 and NOAA Grant NA15OAR4320063 to the University of Washington, through the Joint Institute for the Study of the Ocean and the Atmosphere. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. NCEP/DOE 2 Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/.

References

Agarwal, N., R. Sharma, A. Parekh, S. Basu, A. Sarkar, and V.K. Agarwal, 2012: Argo observations of barrier layer in the tropical Indian Ocean, *Adv. Space Res.*, **50**, 642-654. http://dx.doi.org/10.1016/j.asr.2012.05.021.

Balaguru, K., P. Chang, R. Saravanan, L.R. Leung, Z. Xu, M. Li, and J.-S. Hsieh, 2012: Ocean barrier layers' effect on tropical cyclone intensification, *PNAS*, **109**, 36, 14343-14347, https://doi.org/10.1073/pnas.1201364109.

Beal, L., V. Hormann, R. Lumpkin, and G.R. Foltz, 2013: The response of the surface circulation of the Arabian Sea to monsoonal forcing, *J. Phys. Oceanogr.*, **43**, 2008-2022, https://doi.org/10.1175/JPO-D-13-033.1.

de Boyer Montégut, C., F. Durand, R. Bourdallé-Badie, and B. Blanke, 2014: Role of fronts in the formation of Arabian Sea barrier layers during summer monsoon, *Ocean Dynamics*, **64**, 809-822, https://doi.org/10.1007/s10236-014-0716-7.

Chi, N.-H., R.-C. Lien, E. A. D'Asaro, and B. B. Ma, 2014: The surface mixed layer heat budget from mooring observations in the central Indian Ocean during Madden–Julian Oscillation events, *J. Geophys. Res.: Oceans*, **119**, 4638–4652, https://doi.org/10.1002/ 2014JC010192.

Cole, S.T. and D.L. Rudnick, 2012: The spatial distribution and annual cycle of upper ocean thermohaline structure, *J. Geophys. Res.*, **117**, C02027, https://doi.org/10.1029/2011JC007033.

Cronin, M. F., and M. J. McPhaden, 2002: Barrier layer formation during westerly wind bursts, *J. Geophys. Res.*, **107**, C12, 8020, https://doi.org/10.1029/2001JC001171.

Durand, F., D. Shankar, C. de Boyer Montégut, S.S.C. Shenoi, B. Blanke, and G. Madec, 2007: Modeling the barrier-layer formation in the southeastern Arabian Sea. *J. Climate*, **20**, 2109-2120. https://doi.org/10.1175/JCLI4112.1

Durand, F., S. R. Shetye, J. Vialard, D. Shankar, S. S. C. Shenoi, C. Ethe, and G. Madec, 2004: Impact of temperature inversions on SST evolution in the South-Eastern Arabian Sea during the pre-summer monsoon season, *Geophys. Res. Lett.*, **31**, L01305, https://doi.org/10.1029/2003GL018906.

Feistel, R., 2003: A new extended Gibbs thermodynamic potential of seawater, *Prog. Oceanogr.*, **58**, 43-114.

Ffield, A., 2007: Amazon and Orinoco river plumes and NBC rings: Bystanders or participants in hurricane events? *J. Climate*, **20**, 316-333, https://doi.org/10.1175/JCLI3985.1.

Flament, P., 2002: A state variable for characterizing water masses and their diffusive stability: spiciness, *Prog. Oceanogr.*, **54**, 493-501.

IOC, SCOR and IAPSO, 2010: The international thermodynamic equation of seawater – 2010: Calculation and use of thermodynamic properties. Intergovernmental Oceanographic Commission, Manuals and Guides No. 56, UNESCO (English), 196 pp. Available from http://www.TEOS-10.org.

Jackett, D.R. and T.J. McDougall, 1985: An oceanographic variable for the characterization of intrusions and water masses, *Deep-Sea Res.*, **32**, 10, 1195-1207.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2), *Bull. Amer. Meteor. Soc.*, **83**, 1631-1643, https://doi.org/10.1175/BAMS-83-11-1631.

Kara, A. B., P. A. Rochford, and H. E. Hurlburt, 2000: Mixed layer depth variability and barrier layer formation over the North Pacific Ocean. *J. Geophys. Res*., **105** (C7), 16783–16801.

Katsura, S., E. Oka, and K. Sato, 2015: Formation mechanism of barrier layer in the Subtropical Pacific, *J. Phys. Oceanogr.*, **45**, 2790- 2805, https://doi.org/10.1175/JPO-D-15-0028.1.

Large, W., 1998: Modeling and parameterizing the ocean planetary boundary layer. *Ocean Modeling and Parameterization.* E.P. Chassignet and J. Verron, Eds., Kluwer Academic Publishers, 81-120.

Lorbacher, K., D. Dommenget, P. P. Niiler, and A. Köhl, 2006: Ocean mixed layer depth: A subsurface proxy of ocean- atmosphere variability, *J. Geophys. Res.*, **111**, C07010, https://doi.org/10.1029/2003JC002157.

Lukas, R. and E. Lindstrom, 1991: The Mixed Layer of the Western Equatorial Pacific Ocean, *J. Geophys. Res.*, **96**, Supplement, 3343-3357.

Masson, S. and Coauthors, 2005: Impact of barrier layer on winter-spring variability of the southeastern Arabian Sea, *Geophys. Res. Lett.*, **32**, L07703, https://doi.org/10.1029/2004GL021980.

McDougall, T.J., and O.A. Krzysik, 2015: Spiciness, *J. Mar. Res.,***73**, 141-152, https://doi.org/10.1357/002224015816665589.

McPhaden, M.J., and S.P. Hayes, 1991: On the variability of winds, sea surface temperature, and surface layer heat content in the Western Equatorial Pacific, *J. Geophys. Res.*, **96**, Supplement, 3331-3342.

Mignot, J., C. de Boyer Montégut, and M. Tomczak, 2009: On the porosity of barrier layers, *Ocean Sci.*, **5**, 379-387.

Price, J.F., R.A. Weller, and R. Pinkel, 1986: Diurnal cycling: observations and models of the upper ocean response to diurnal heating, cooling, and wind mixing, *J. Geophys. Res.*, **91**, C7, 8411-8427.

Qu, T., and G. Meyers, 2005: Seasonal variation of barrier layer in the southeastern tropical Indian Ocean, *J. Geophys. Res.*, **110**, C11003, https://doi.org/10.1029/2004JC002816.

Rao, R.R. and R. Sivakumar, 1999: On the possible mechanisms of the evolution of a mini-warm pool during the pre-summer monsoon season and the genesis of onset vortex in the south-east Arabian Sea, *Quart. J. Roy. Meteor. Soc.*, **125**, 787-809.

Roemmich, D., M. Morris, W.R.Young, and J.R. Donguy, 1994: Fresh equatorial jets, *J. Phys. Oceanogr.*, **24**, 540-558.

Schmitt, R.W., 1981: Form of the temperature-salinity relationship in the Central Water: Evidence for double-diffusive mixing, *J. Phys. Oceanogr.*, **11**, 1015-1025.

Schmitt, R.W., 1990: On the density ratio balance in the Central Water, *J. Phys. Oceanogr.*, **20**, 900-906.

Schmidtko, S., G. C. Johnson and J. M. Lyman, 2013: MIMOC: A Global Monthly Isopycnal Upper-Ocean Climatology with Mixed Layers, *J. Geophys. Res.*, **118**, 1658-1672, https://doi.org/10.1002/jgrc.20122.

Schott, F.A., and J.P. McCreary, 2001: The monsoon circulation of the Indian Ocean, *Prog. Oceanogr.*, **51**, 1-123.

Shankar, D., and Coauthors, 2004: Observational evidence for westward propagation of temperature inversions in the southeastern Arabian Sea. *Geophys. Res. Lett.*, **31**, L08305, https://doi.org/10.1029/2004GL019652

Shcherbina, A.Y., M.C. Gregg, M.H. Alford, and R.R. Harcourt, 2009: Characterizing thermohaline intrusions in the North Pacific Subtropical Frontal Zone, *J. Phys. Oceanogr.*, **39**, 2735-2756, https://doi.org/10.1175/2009JPO4190.1.

Shenoi, S.S.C, D. Shankar, and S.R. Shetye, 1999: On the sea surface temperature high in the Lakshadweep Sea before the onset of the southwest monsoon, *J. Geophys. Res.*, **14**, C7, 15703-15712.

Shenoi, S. S. C., D. Shankar, and S. R. Shetye, 2004: Remote forcing annihilates barrier layer in southeastern Arabian Sea, *Geophys. Res. Lett.*, **31**, L05307, https://doi.org/10.1029/ 2003GL019270.

Sprintall, J. and M. Tomczak, 1992: Evidence of the barrier layer in the surface layer of the tropics, *J. Geophys. Res.*, **97**, C5, 7305-7316.

Thadathil, P. and Coauthors, 2008: Seasonal variability of the observed barrier layer in the Arabian Sea, *J. Phys. Oceanogr.*, **38**, 624-638, https://doi.org/10.1175/2007JPO3798.1.

Veronis, G., 1972: On properties of seawater defined by temperature, salinity, and pressure, *J. Mar. Res.*, **30**, 2, 227-255.

Vialard, J., and P. Delecluse, 1998: An OGCM study for the TOGA decade. Part 1: Role of salinity in the physics of the Western Pacific Fresh Pool, *J. Phys. Oceanogr.*, **28**, 1071-1088.

You, Y., 1998: Rain-formed barrier layer of the western equatorial Pacific warm pool: A case study, *J. Geophys. Res.*, **103**, C3, 5361-5378.

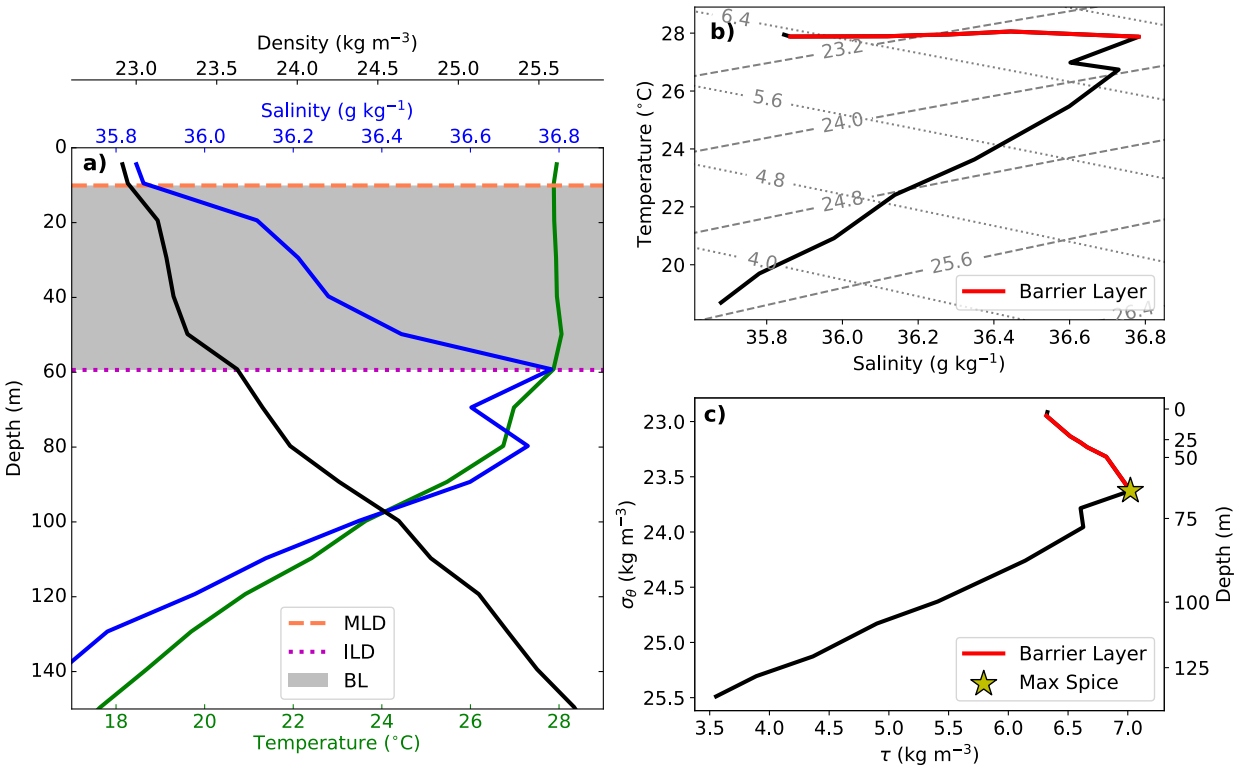

Figure 2.1. Data from Argo float 2900258, 1/28/2004, 11.98°N, 69.719°E: (a) temperature, salinity, and density profiles from the float. Mixed layer depth is indicated by the orange dashed line, with isothermal layer depth indicated by the magenta dotted line. Gray shaded area is the section of the water column termed the "BL". (b) *T*-*S* diagram corresponding to profile in (a). Red portion of line indicates the BL. Dashed gray lines are potential density contours, dotted gray lines are spice contours (both kg m^{-3}). (c) Potential density-spice diagram. Red-shaded section as in (b). Yellow star indicates the location of maximum spice. Plots show different representations of the same segment of the water column.

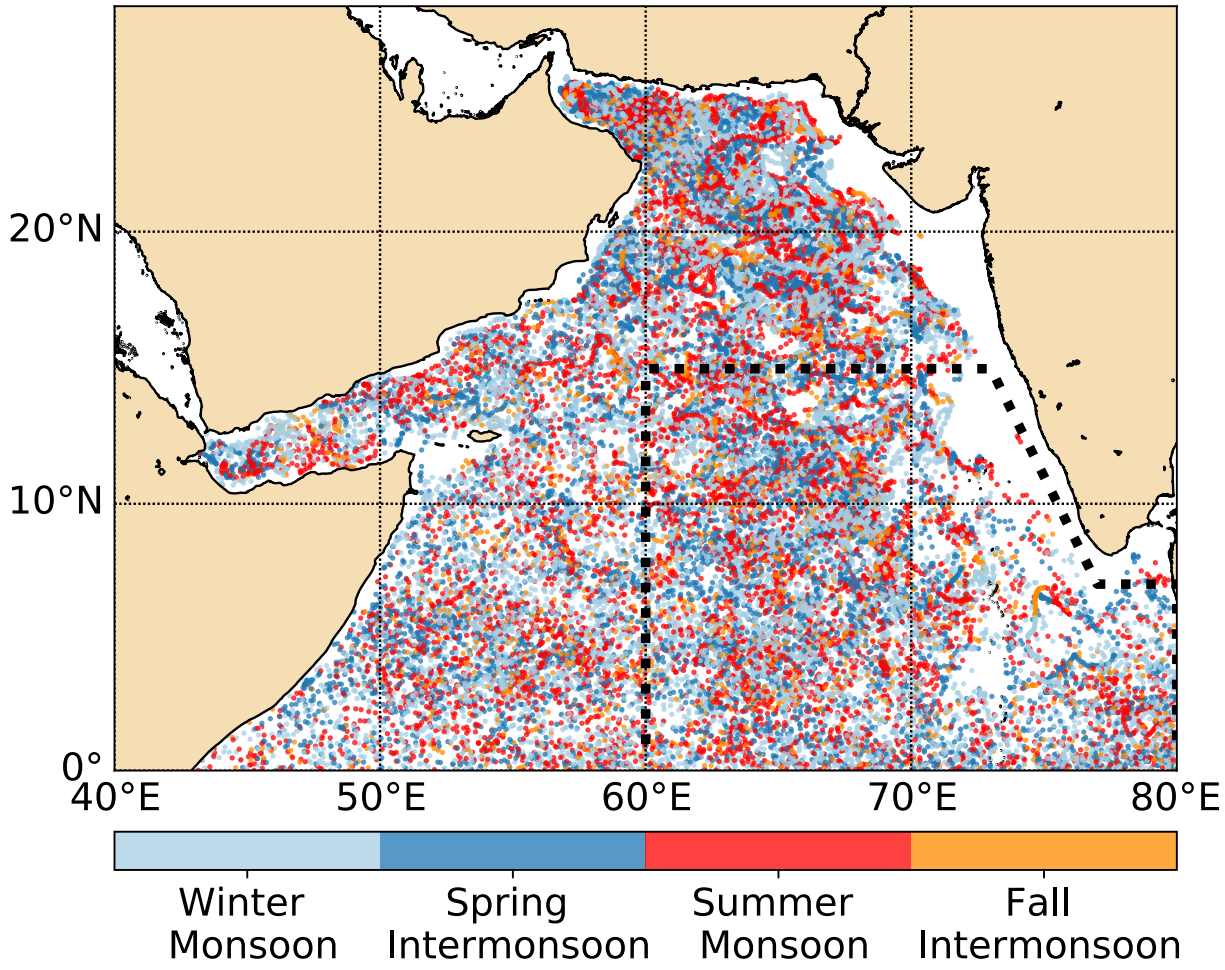

Figure 2.2. Distribution of profiling float data in the Arabian Sea, from 2004-2016. Each dot indicates one profile, color-coded by season. Black-dashed area indicates SEAS-definition for use in this study.

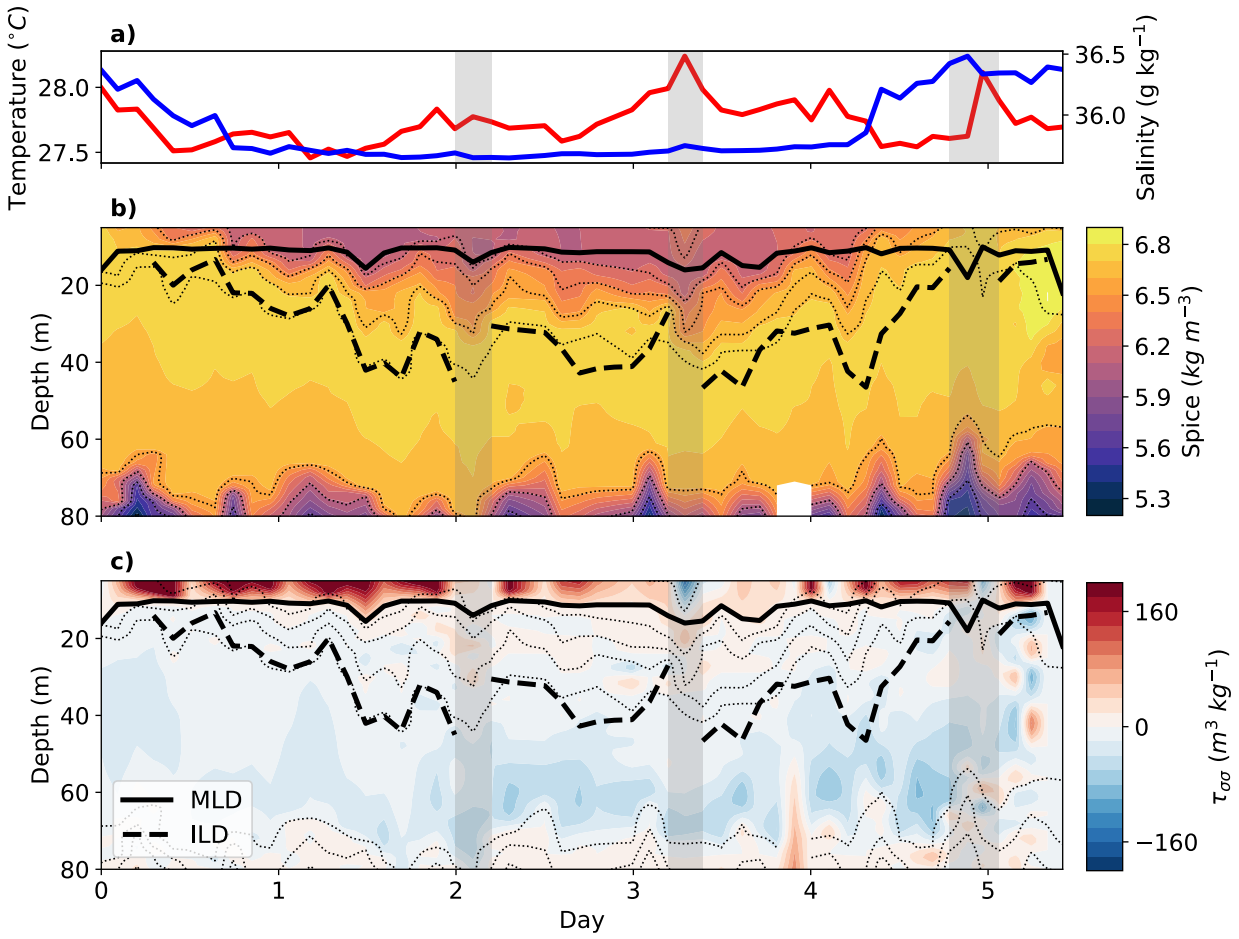

Figure 2.3. Data from part of a 12-day, ~2-hourly cycle routine during February, 2015 near 13°N, 69°E (profiling float 2902174). (a) Surface temperature (red) and salinity (blue); (b) spice; and (c) diapycnal spiciness curvature. Solid black line is mixed layer depth and dashed black line is isothermal layer depth. Both depths determined using spice method. Highlighted regions depict temporary BL erosion.

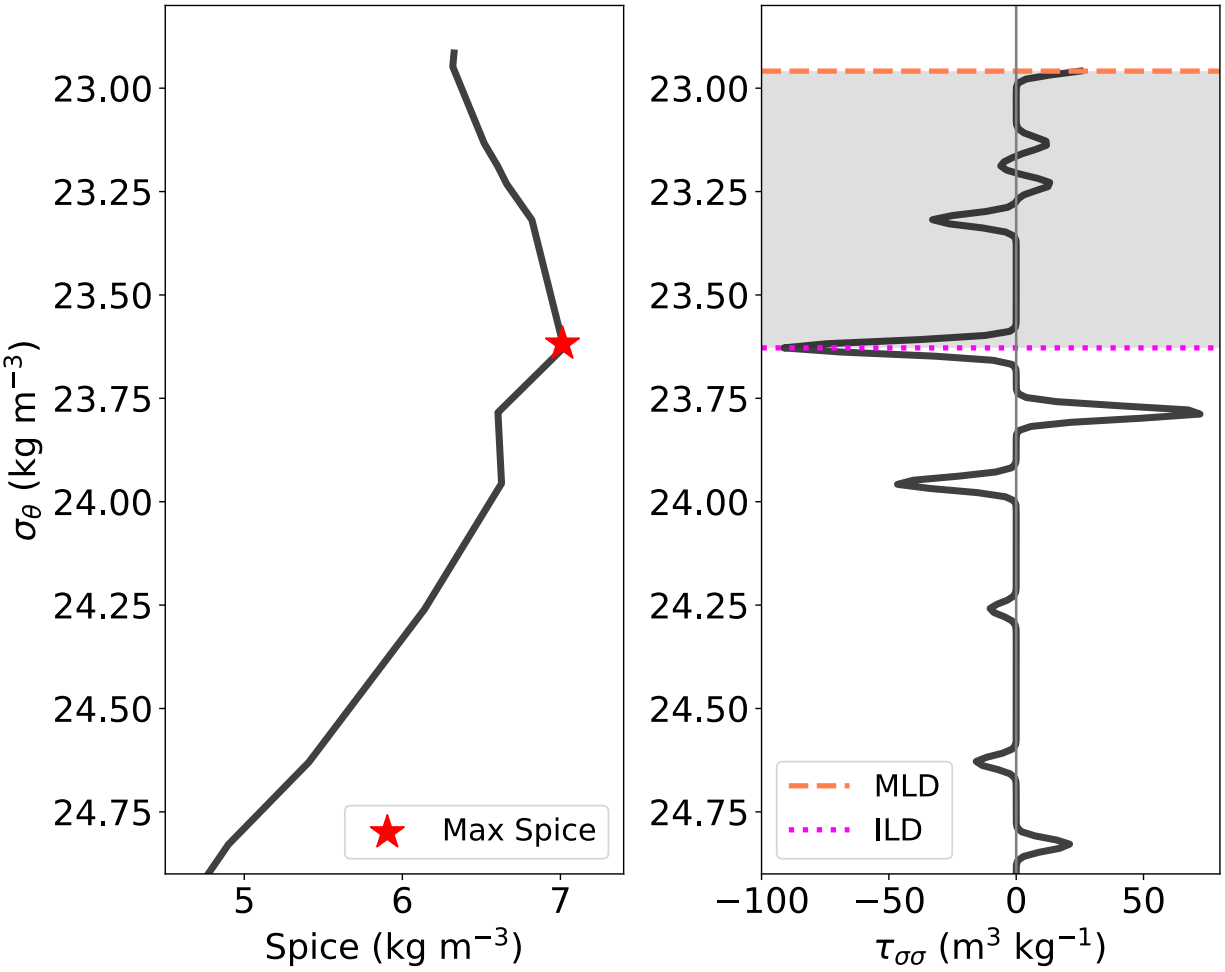

Figure 2.4. $\tau_{\sigma\sigma}$ peak locations used for defining the mixed layer depth (orange dotted line) and isothermal layer depth (magenta dashed line). Note the correspondence between the maximum spice (red star) and isothermal layer depth. This profile shows a relatively non-noisy profile with clearly dominant peaks near the densities of interest. Gray shaded region corresponds to the BL. Same float profile as for Figure 1.

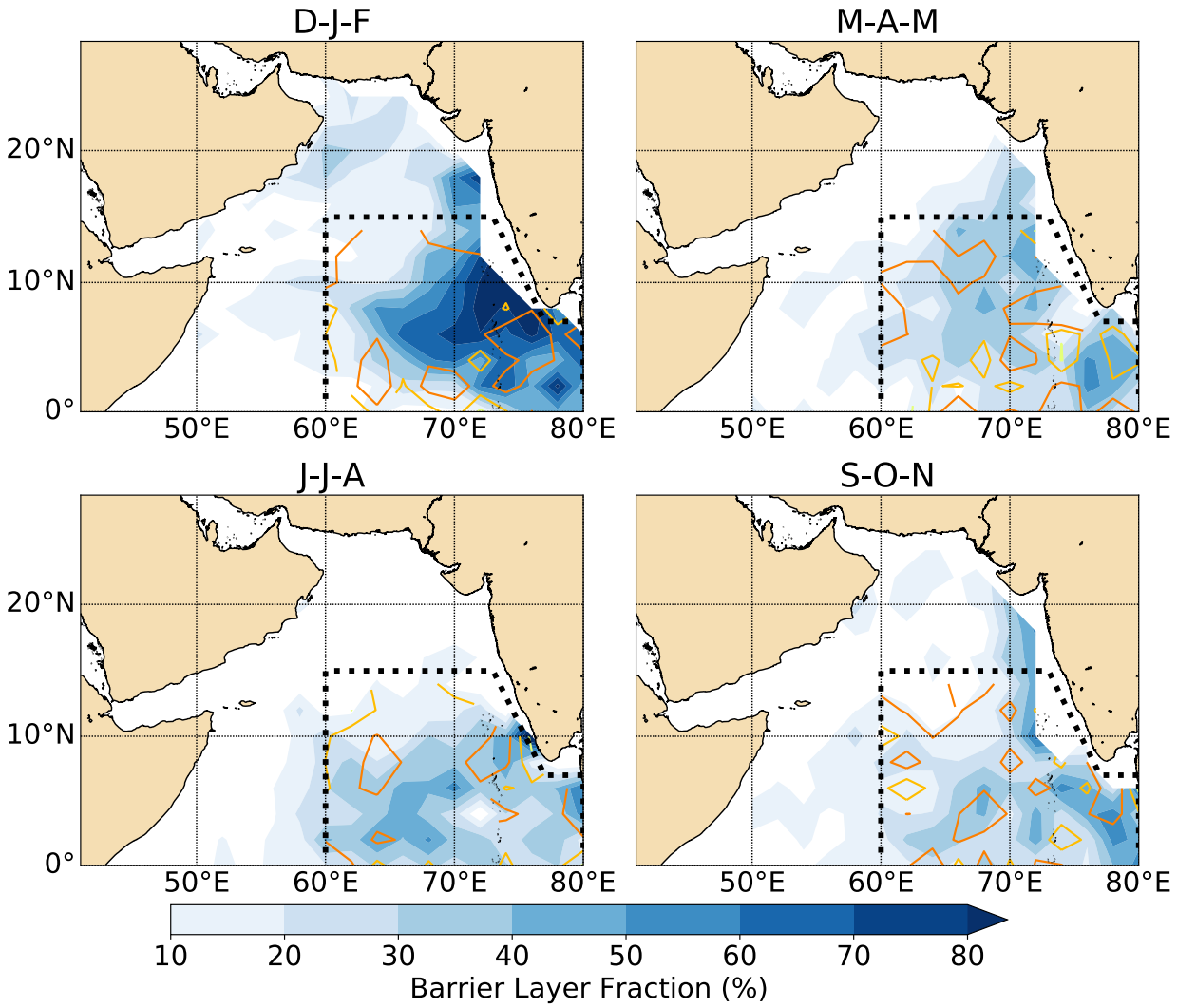

Figure 2.5. Frequency (blue filled contours) and thickness (orange line contours, SEAS only) of BL by season. Thickness contours are at 30m (light orange) and 40m (dark orange). Black-dashed area indicates SEAS-definition for use in this study.

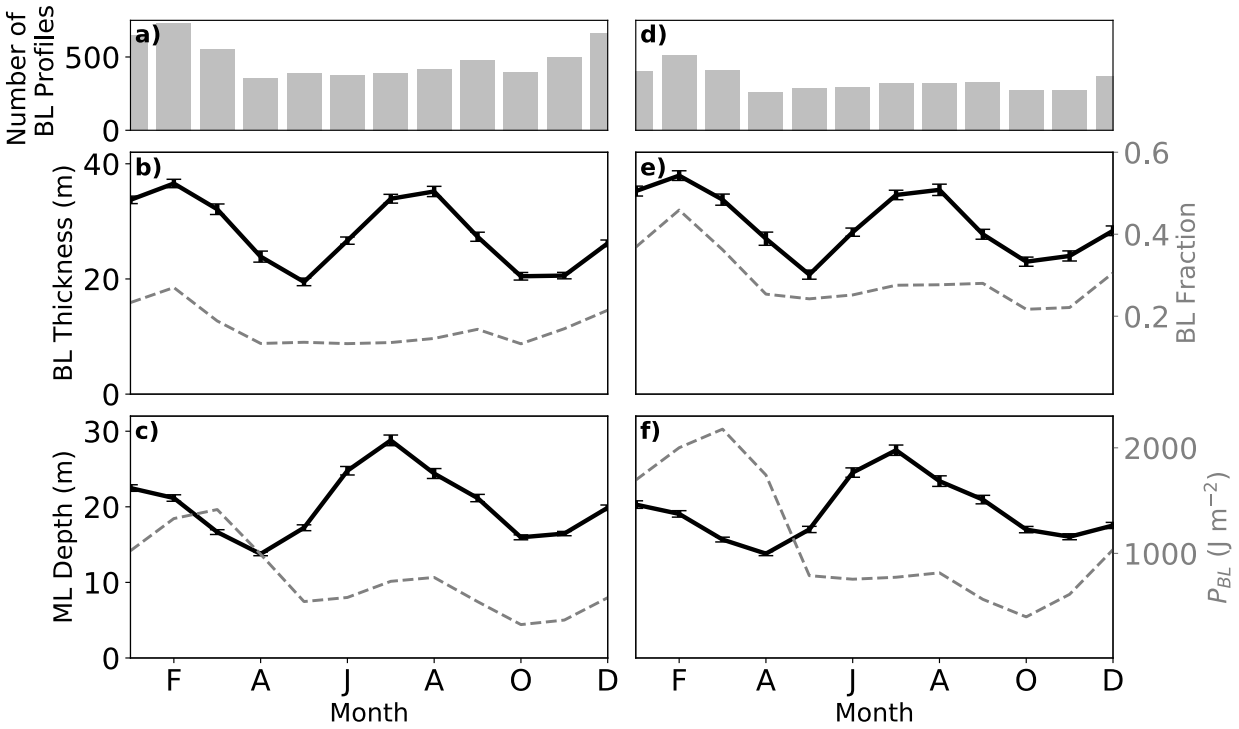

Figure 2.6. BL statistics for the Arabian Sea. a) Number of BL profiles observed monthly in Arabian Sea. b) BL thickness indicated by black line (left y-axes), dashed gray line indicate BL fraction (right y-axes). c) ML depth indicated by black line (left y-axes), dashed gray line indicates P_{BL} (right y-axes). Error bars are standard error. d)-f): Same as a)-c) but for SEAS.

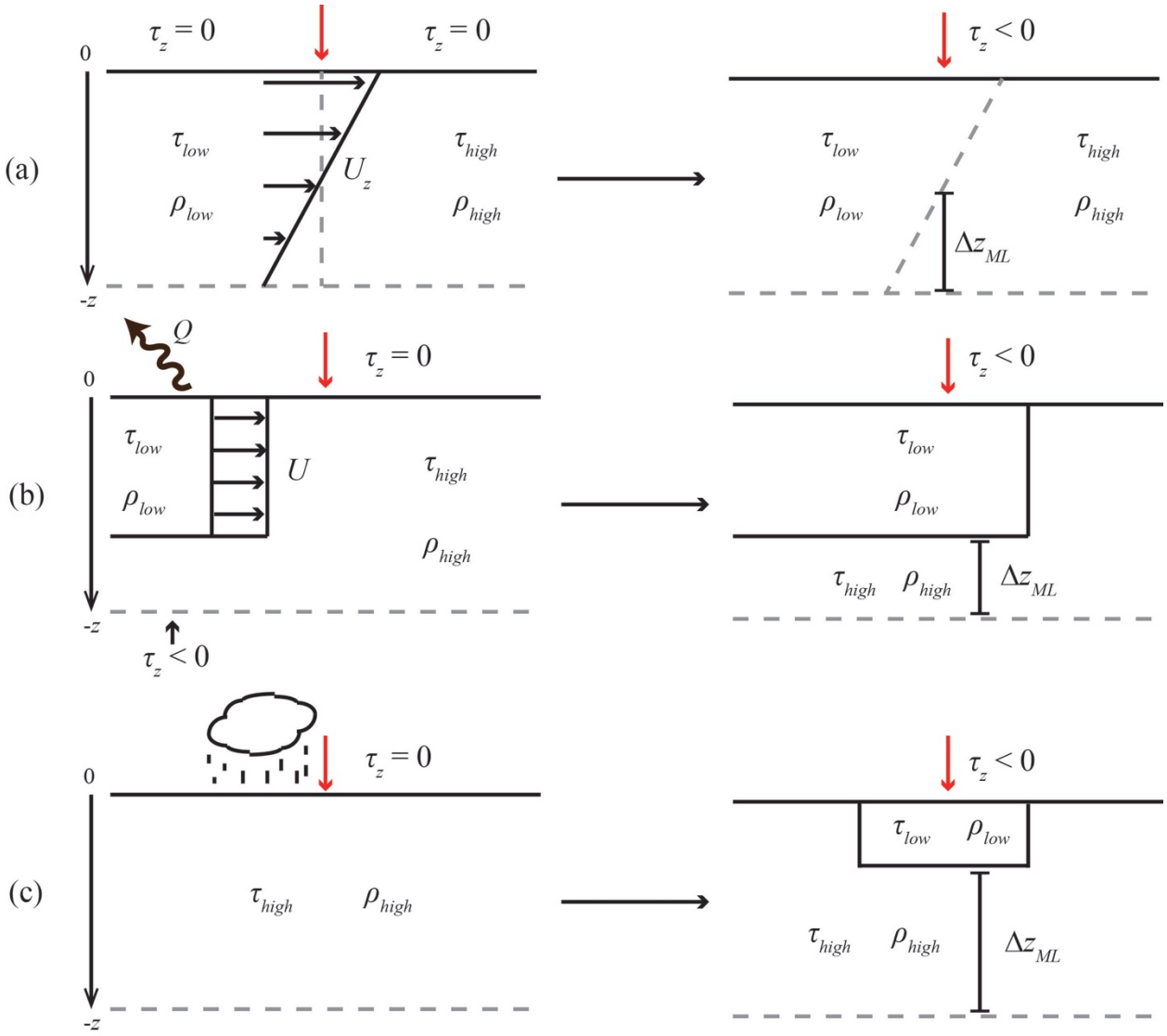

Figure 2.7. Schematic showing mechanisms of BL formation in terms of spice. (a) corresponds to frontal tilting (Term 3 on right side of Equation 11); (b) corresponds to advection of freshwater (Term 1 on right side of Equation 11), with possible surface cooling (Term 5); (c) corresponds to rain-formed BL, requiring some wind-stirring to mix the freshening downwards (Term 5 on right side of Equation 11). Red arrow shows the location of observation.

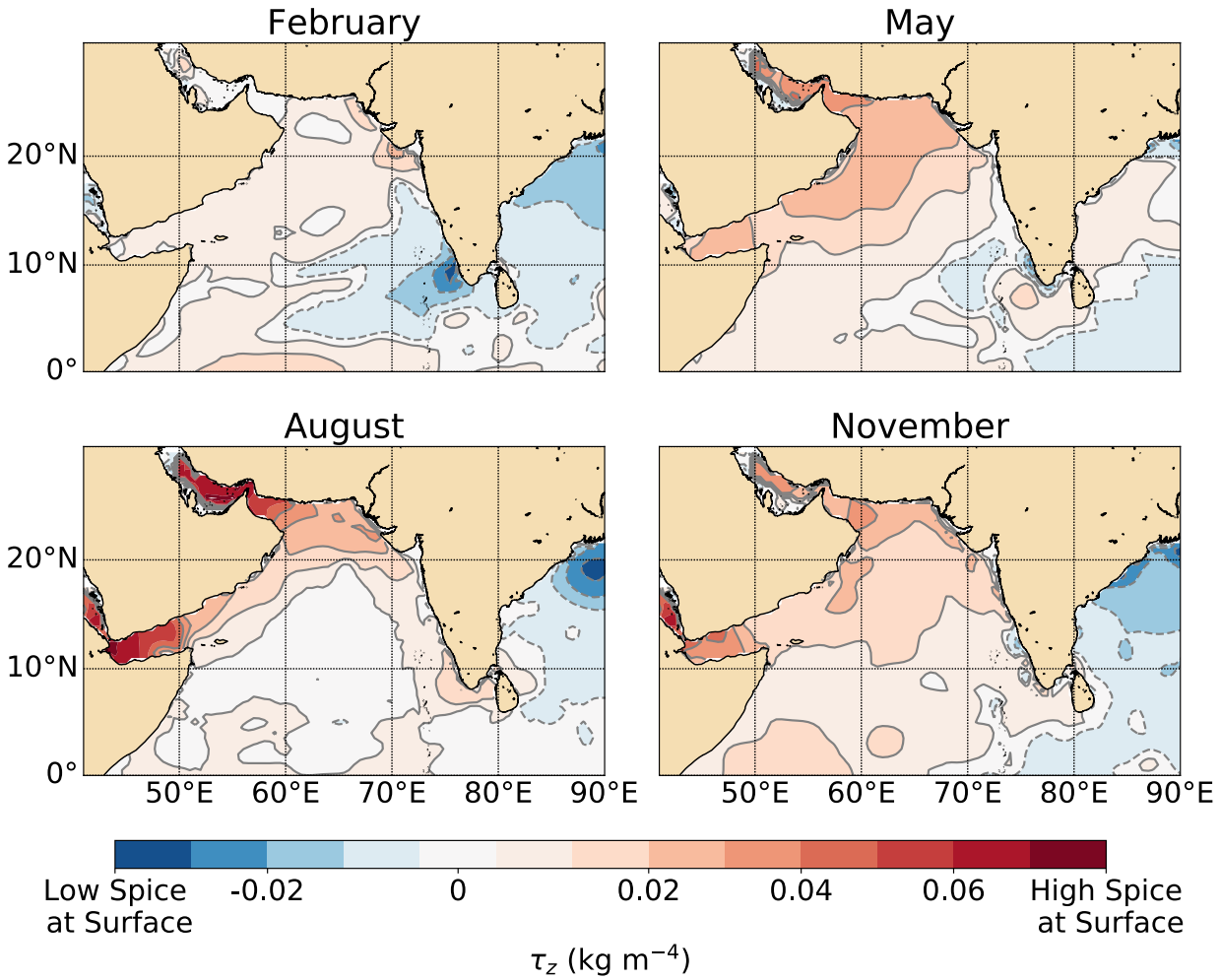

Figure 2.8. Plots of vertical derivative of spice, τ_z,derived from MIMOC climatology. Blue regions indicate locations where the climatological conditions exist to allow for advection of an existing BL into an adjacent region, given suitable velocities (term 1 on right side of Equation 11).

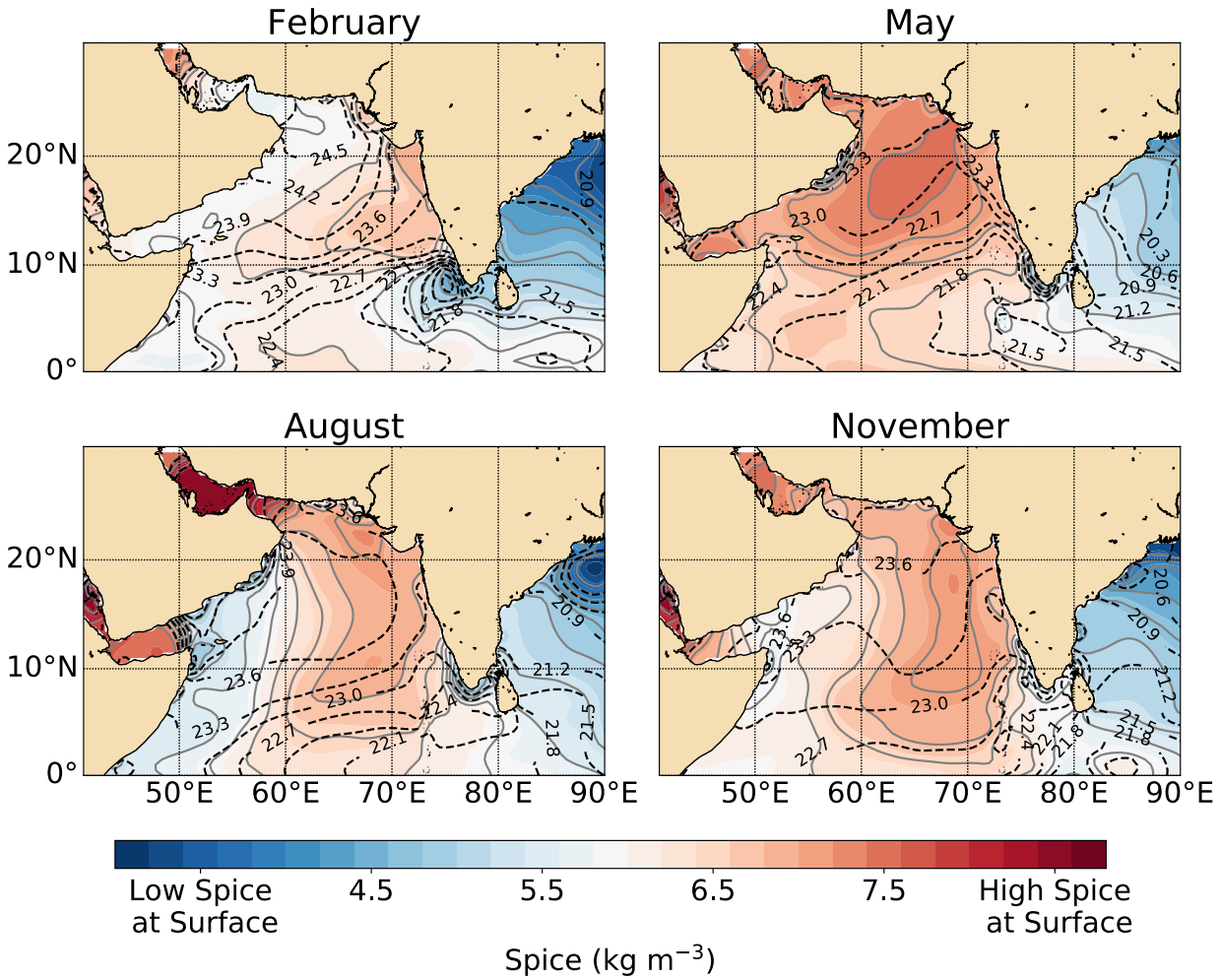

Figure 2.9. Contours of spice (color shading and gray solid lines) and potential density (black dotted lines) derived from MIMOC.

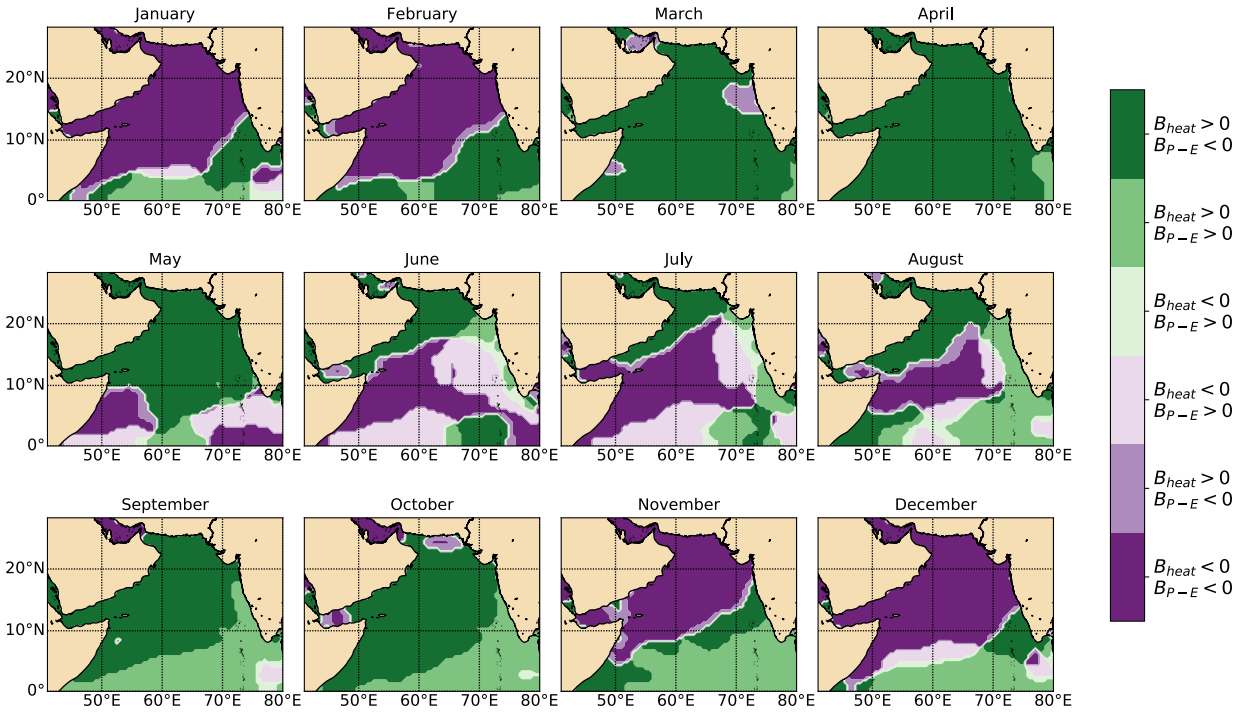

Figure 2.10. Regions of Monin-Obukhov stability, shaded according to the balance of buoyancy forcing due to heat flux and buoyancy forcing due to evaporation-precipitation. Green areas are stable (consistent with hatched areas in Figure 11), purple areas are unstable.

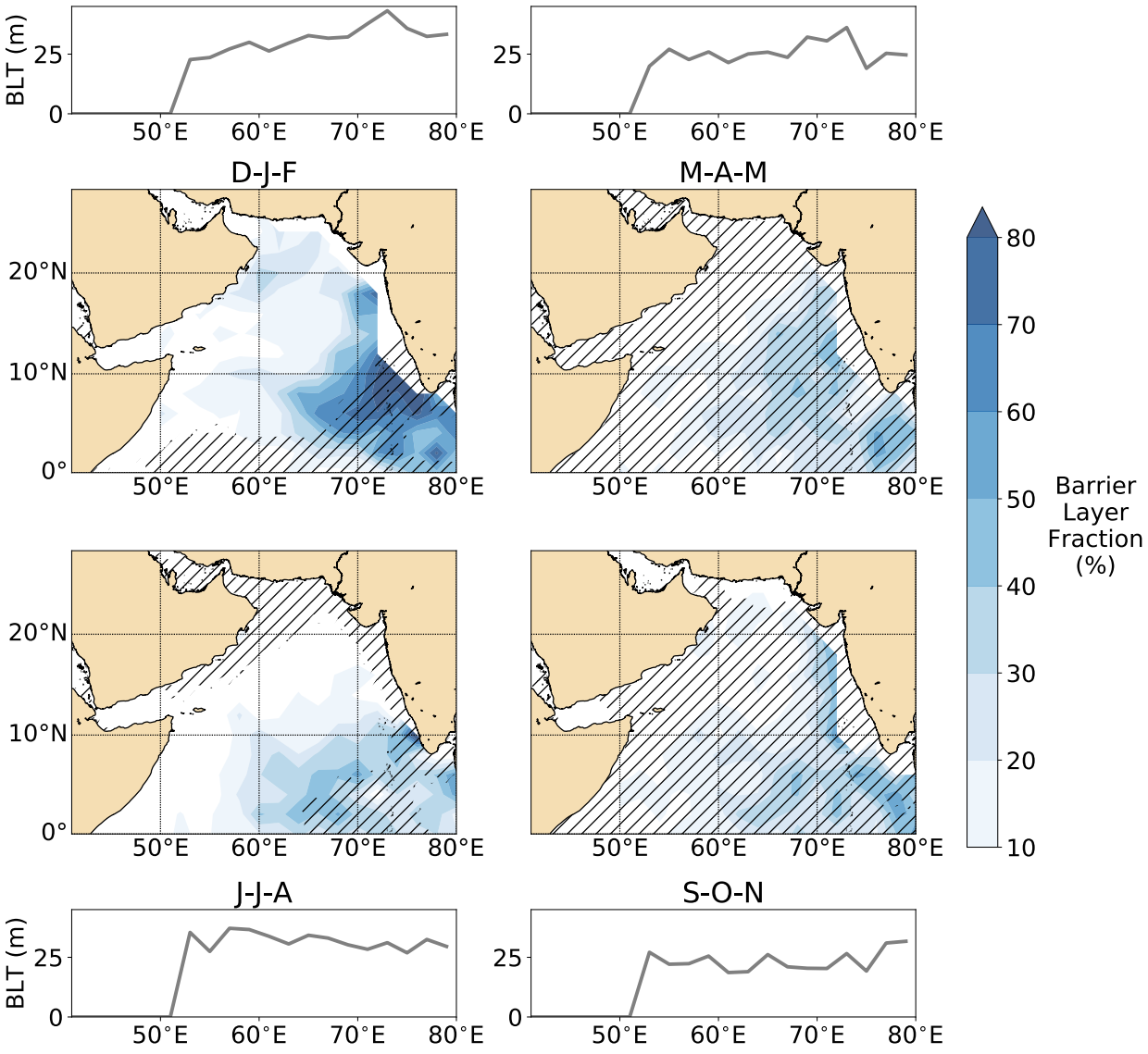

Figure 2.11. Distribution of BLs (blue shading) with regions of stability derived from the Monin-Obukhov parameter (hatch marks). Estimates of stability were derived from NCEP long-term average heat fluxes and wind forcing. Line plots show meridionally averaged barrier layer thickness for each season for the dominant barrier layer regions.

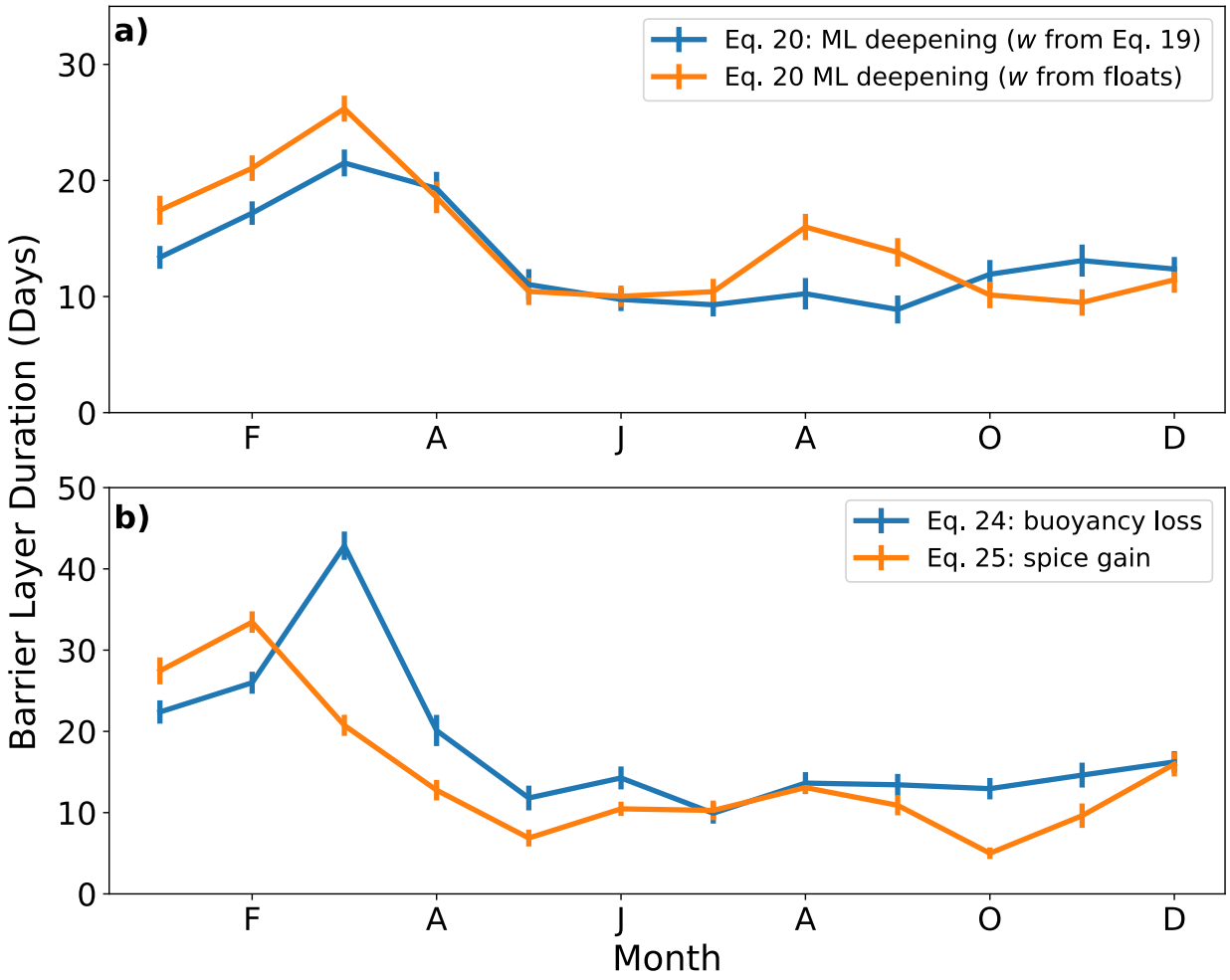

Figure 2.12. Expected duration of BLs in the SEAS due to erosion by (a) ML deepening (Equation 20) with entrainment velocity calculated using Equation 19 (blue line) and directly from floats (orange line) and (b) spice gain (Equation 25) or density changes (Equation 24) near the surface.

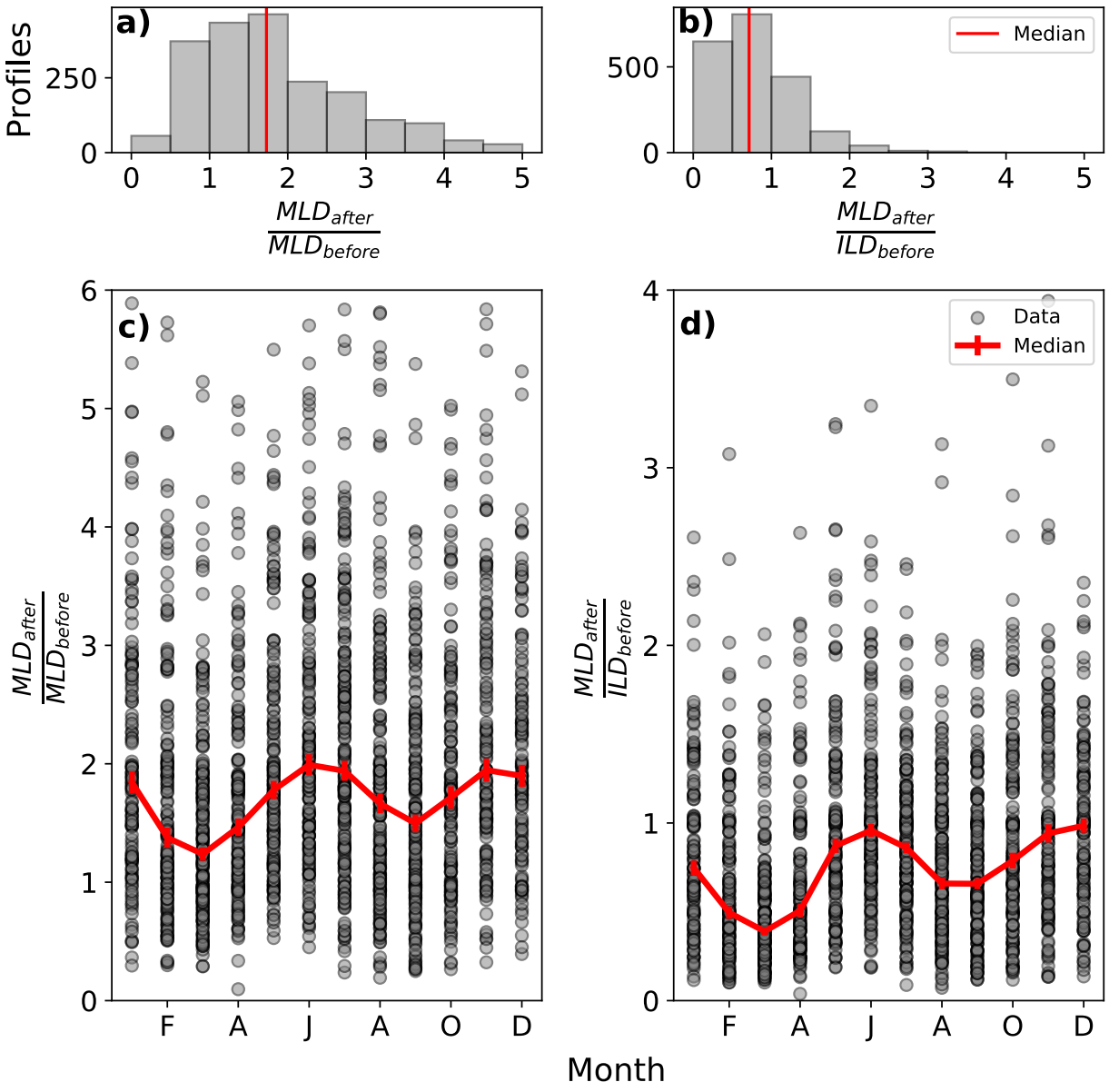

Figure 2.13. Seasonal cycle of mixed layer depths in the SEAS after BL erosion normalized by (a) mixed layer depth during BL presence and (b) isothermal layer depth during BL presence. (c) and (d) show same data distributed by month.

Chapter 3. THE IMPACT OF BARRIER LAYERS ON ARABIAN SEA SURFACE TEMPERATURE VARIABILITY

This work was previously published as: Echols, R., & Riser, S.C. (2020). The Impact of Barrier Layers on Arabian Sea Surface Temperature Variability. Geophysical Research Letters, 47, e2019GL085290. https://doi.org/10.1029/2019GL085290.

3.1 ABSTRACT

Barrier layers (BLs) in the Arabian Sea may lead to increased mixed layer (ML) warming or cooling relative to non-BL conditions, depending on differences in ML properties and atmospheric forcing. We use 13 years of profiling float data from the Arabian Sea to identify when and where each impact is likely based on analysis of the ML temperature tendency. We use a novel method based on the vertical spice profile to identify BL and their characteristics. BL are most likely to increase warming during the northeast monsoon, owing to temperature inversions, preceding the development of the mini-warm pool and onset of the southwest monsoon. This effect is nonuniform and is counteracted in some regions by increased penetration of solar radiation below the shallower ML associated with BL. Relative rates of cooling increase in the presence of BL during the southwest monsoon; intermonsoon periods show little net effect from BL.

3.2 INTRODUCTION

Barrier layers (BL) occur when salinity stratification below the density mixed layer (ML) occurs without temperature stratification, producing a barrier between the ML and the cooler

waters of the thermocline (Figure 1a; Lukas & Lindstrom, 1991; Sprintall & Tomzcak, 1992). Their earliest discovery from high-resolution temperature and salinity profiles in the western Pacific Warm Pool provided density-based ML depths much shallower than those identified based solely on temperature measurements, arising from the local salinity stratification (Lukas & Lindstrom, 1991). They have subsequently been identified throughout the tropics, including the Bay of Bengal and Arabian Sea (e.g. Sprintall &Tomzcak, 1992), where massive transport of freshwater from rivers establishes strong salinity stratification near the surface. Their departure from a thermally stratified ML base raises questions about their impact on sea surface temperatures (SST).

In the Southeast Arabian Sea (SEAS), where high sea surface temperatures (SST) have been observationally linked to the southwest monsoon onset (Rao & Sivakumar, 1999; Shenoi et al., 1999), models show evidence of enhanced warming associated with BLs during the spring intermonsoon prior to monsoon onset (Masson et al., 2005). Shallower MLs may support increased warming and momentum transfer due to heat fluxes and wind stress being absorbed in a thinner layer (Cronin & McPhaden, 2002; Cronin & Kessler, 2009); modification of the stratification below the ML can alter wind-driven mixing (Niiler & Kraus, 1977); and the separation between the density ML and the thermocline may alter the consequences of mixing. Strong salinity stratification allows the upper ocean to support substantial temperature inversions without convective overturn (Kurian & Vinayachandran, 2006; Shankar et al., 2004; Vinayachandran et al., 2007).

Yet it should not be a foregone conclusion that BL presence results in increased ML warming. Vinayachandran et al. (2007) proposed that shallow MLs subjected to net heat flux from ocean to atmosphere may in fact lead to enhanced cooling, although at the time there was

insufficient data to fully investigate this hypothesis. The Arabian Sea is a complex region, with seasonally reversing currents (Schott & McCreary, 2001), upwelling, downwelling, and widely varying seasonal and regional heat flux, making a simple statement about BL impact impossible. In addition, modeling studies sometimes overestimate ML depth associated with BL (Masson et al., 2005) and often have lower vertical resolution than is available with modern observational techniques. Given the range of possible BL impacts, nuances in the differences between BL and non-BL profiles may be significant to their impact.

Here, we use thirteen years of Argo float data to examine the impact of BL in the SEAS by comparing the characteristics of BL and non-BL profiles, and examining how differences in these characteristics impact relevant terms of the ML temperature tendency equation. The study benefits from greater data density and wider data distribution than previous observational studies, a method for identifying BL that more precisely captures ML and BL characteristics, and greater vertical resolution in the data than previous modeling studies.

3.3 DATA

This study uses Argo-type float data spanning the years 2004-2016, starting when the float distribution reached approximately full spatial and temporal coverage in the region and ending when this study began. Floats typically spend ~9 days at a parking depth near 1000m, before descending to 2000m and profiling upward to the surface, achieving a vertical measurement resolution of ~2m near the surface (Riser et al, 2016). We focus on the SEAS, here considered to be the region bounded by the Equator and 15° N, spanning longitudes of 60°E and 80°E (Figure 2, red outlines). Initially, all available temperature (T) and salinity (S) profiles were considered. Any profile with missing temperature or salinity data in the top 10 meters was

discarded, as well as profiles with values outside a reasonable range for this location ($0<T<35°C$; $20<S<40$). Profiles whose maximum pressure was less than 100 meters, or with large gaps in the pressure data, were also excluded. After these quality control measures, there were a total of 13,923 useable profiles of temperature and salinity (~66% of the original total). In contrast with previous observational studies (e.g. Thadathil et al., 2008, de Boyer Montégut et al., 2014), data coverage is better across the eastern and southern portions of the region (>50 profiles seasonally per 2º×2º box outside of coastal and shallow regions). The monthly average number of total float profiles is roughly even between months, although there is some interannual variability (Figure 2).

Daily averaged heat fluxes and wind fields from the National Centers for Environmental Prediction/Department of Energy (NCEP-DOE) Reanalysis (version 2) were used to calculate approximate fluxes associated with individual profiles (Kanamitsu et al, 2002). While the lack of collocated measurements (such as those from shipboard measurements) poses a challenge for making precise calculations associated with an individual profile, the goals of this study are to understand bulk regional impacts of BLs in the SEAS, and these data provide a reasonable starting point.

3.4 METHODS

As defined by Lukas and Lindstrom (1991), the BL is the layer between the base of the density ML and the top of the thermocline that exists due to strong salinity stratification (Figure 1a). Rather than consider a temperature threshold to define the end of the BL, we consider diapycnal changes in spice, a state variable that allows simultaneous consideration of

temperature and salinity (Veronis, 1972; Flament, 2002; Jackett & McDougall, 1985; Shcherbina et al., 2009). An approximate differential form of spice is

$$d\tau = \rho(\alpha d\theta + \beta dS) \quad (1)$$

where τ denotes spice, ρ is the reference density, θ is potential temperature, $\alpha = -\frac{1}{\rho}(\frac{\partial \rho}{\partial \theta})$, and $\beta = \frac{1}{\rho}\left(\frac{\partial \rho}{\partial S}\right)$ (Shcherbina et al., 2009). BLs exist when lower density, fresh (and sometimes cooler) water (low spice) overlays higher density salty (and sometimes warmer) water (high spice) (Echols & Riser, 2020). The top of the thermocline, and base of the BL, coincides with the reverse transition in spice. In this framework, the base of the ML is defined as the depth of the first diapycnal transition from lower to higher spice (beginning of salinity-controlled stratification) and the base of the BL is defined as the diapycnal transition from higher to lower spice (temperature-controlled stratification) (Figure 1b; Echols & Riser, 2020). We use the shallow-water definition of spiciness described in McDougall and Krzysik (2015) to quantitatively determine spice. This definition accommodates variations in the strength of temperature and salinity stratification without selecting a different temperature threshold, and identifies the depth of the appropriate transition rather than the depth where a threshold has been exceeded. Figure 2 shows the seasonal BL frequency determined from this method. BL frequency is the value R used by Mignot et al (2009) to calculate porosity (1-R), although without the same restrictions on BL thickness.

To investigate the evolution of SST through interactions with the atmosphere, with and without BLs present, we focus on the particular terms in a ML temperature tendency equation (Foltz et al., 2003; McPhaden & Hayes, 1991; Stevenson & Niiler, 1983) that may be affected by the changes to the ML associated with BLs. Assuming that temperature and velocity are both constant over the ML depth, we use

$$\underset{(a)}{\frac{\partial T}{\partial t}} + \underset{(b)}{(\mathbf{v} \cdot \nabla T)} + \underset{(c)}{\frac{(T_{ML} - T_{-h}) w_e}{h}} = \underset{(d)}{\frac{q_0 - q_{-h}}{h \rho c_p}} \quad (2)$$

where **v** represents the horizontal velocity within the ML, T_{ML} is the vertically averaged ML temperature, T_{-h} is the temperature just below the base of the ML (here taken as 10m below the bottom of the ML), h is the ML depth, and w_e is the entrainment velocity. Term (*a*) represents the temperature tendency; (*b*) advection by the mean flow; (*c*) entrainment of water below the ML; and (*d*) heat flux into the ML, in which the total surface heat flux (q_0) is decreased by the heat flux due to shortwave radiation penetrating downwards through the base of the ML and turbulent mixing (q_{-h}). For the purposes of this study, we focus solely on the penetrative radiation component of q_{-h} and incorporate turbulent mixing into the calculation of w_e (see Equation 4). Given the data available for this study, which does not include horizontal velocities tied to specific float profiles, and previous work showing that advection was not a dominant term in the heat budget (Durand et al., 2004; Rao & Sivakumar, 1999), we will only evaluate terms (*a*), (*c*), and (*d*) in considering the impacts of BLs.

For shallow MLs, and typical tropical heat fluxes, the radiation penetrating the base of the ML in term (*d*), q_{-h}, could significantly reduce the heat transfer to the ML. In an ideal BL, the ML resides within a completely isothermal layer, so term *(c)* would be identically zero, unlike the cooling we might expect with a temperature-stratified ML. Figure 1a shows an example profile where T_{ML} does not differ significantly from T_{-h}; even for BL profiles that deviate from this model and have slight temperature stratification, we expect this term to be much smaller than in a normal ML and likely negligible if T_{-h} is very close to T_{ML} (Vialard & Delecluse, 1998). These effects may exert opposing influences on SST.

Temperature inversions, where the temperature at depth exceeds SST by at least 0.2°C, can also develop during the lifetime of the BL (Kurian & Vinayachandran, 2006) due to surface

cooling under negative heat flux conditions (term (*d*) in Equation 2 is negative, implying net heat flux out of the ML). Significant temperature inversions are only possible if the salinity stratification is strong enough to compensate for the density changes associated with increased temperature, i.e. if $\beta dS > \alpha dT$ (where $\alpha = -\frac{1}{\rho}(\frac{\partial \rho}{\partial T})$ and $\beta = \frac{1}{\rho}(\frac{\partial \rho}{\partial S})$); otherwise, convective overturn would lead to BL erosion. Temperature inversions have been observed in the SEAS during the northeast monsoon (Durand et al., 2004; Shankar et al., 2004; Vinayachandran et al., 2007). Therefore, term *(c)* could also contribute a positive tendency to the SST change when $T_{-h} > T_{ML}$ (Durand et al., 2004; Shankar et al., 2004; Smyth et al., 1996). Differences in ML depth, *h*, could further alter the impact of terms (*c*) and (*d*) owing to its position in the denominator.

Several models exist for determining the penetrative radiation at the base of the ML. For the Arabian Sea, BLs are most commonly found in the low chlorophyll southeastern regions (Sweeney et al., 2003; Figure 2) and profiling floats rarely enter the near-coastal waters of western India where productivity (and thus chlorophyll) can be elevated. Therefore, here we use a simple parameterization with a constant extinction coefficient of $\gamma = 0.04\ \mathrm{m}^{-1}$ corresponding to a 25m e-folding depth (Foltz et al., 2003; Niiler & Kraus, 1977; Wang & McPhaden, 1999) to estimate q_{-h} using

$$q_{-h} = 0.45 q_{short} e^{-\gamma h} \tag{3}$$

where q_{short} is the net surface shortwave heat flux. Values of the coefficient, which describes the fraction of light penetrating past the uppermost few meters, range between 0.42 and 0.47 across studies of tropical regions, so we follow Wang and McPhaden (1999) in using 0.45.

Lacking direct measurements of the vertical velocity, estimating the entrainment velocity, w_e, is difficult. Fully assessing w_e from the turbulent kinetic energy balance requires

consideration of free convection, shear flow, vertical turbulent energy flux, and dissipation (Niiler & Kraus, 1977). Data required for investigation of shear effects or dissipation are not available for this region; to approximate w_e, we instead use an expression to estimate the effects of wind work without any buoyancy effects, after McPhaden and Hayes (1991),

$$w_e = \frac{2m\left(\frac{\rho_a C_D}{\rho_0}\right)^{3/2}|U|^3}{-g\left(\frac{\Delta\rho}{\rho_0}\right)h} \quad (4)$$

where m is an efficiency factor (Niiler & Kraus, 1977), ρ_0 is the ML density, $\Delta\rho$ is the jump in density across the base of the ML, g is gravity, and ρ_a, C_D, and U are the air density, drag coefficient, and wind velocity respectively. The presence of a BL could modify this term through changes to $\Delta\rho$ or ML depth. Entrainment is only relevant for cases of ML deepening, where water below the ML might be incorporated into the ML; for values that might suggest ML shoaling we neglect it. We also examine changes to the ML depth, $\frac{\partial h}{\partial t}$, estimated from the change in ML between float profiles divided by the time between the profiles, as well as the formulation used by Lee et al. (2000), which includes buoyancy effects. Because floats sometimes move between water masses, the former calculation is not completely reliable for determining ML deepening. However, all three methods show comparable values and seasonal cycles.

3.5 Results

Basin-wide, the presence of BLs is associated with differences in the depth of the ML and the stratification immediately below it; the greatest difference between BL and non-BL conditions is observed during the southwest and northeast monsoons (Figure 3). In comparison

with MLs determined by temperature stratification, BLs are typically associated with lower stratification below the ML (Figure 3a), and shallower MLs (Figure 3b). This produces greater rates of ML deepening (Figure 3c) because there is lower resistance to wind work (Equation (4)). Regions with strong BL presence (as in Figure 2) are typically associated with a 15-25 W m^{-2} increase in average penetrative radiation, q_{-h} (Figure 3d), implying less heat flux into the ML and more downward heat flux below the ML. This effect is most pronounced during the northeast monsoon where BL are most frequent (not shown).

Figure 3e shows the impact of BLs on term (c) (temperature change due to entrainment) in Equation (1). Entrainment cooling is greatly reduced in the presence of a BL. During the northeast monsoon (December and January) the entrainment term tends positive for BL profiles. The magnitude of the effect depends slightly on the depth below the ML used for T_{-h} and the method used to calculate w_e. However, all methods produced comparable estimates. The small positive values observed in December and January in Figure 3e for the BL line are consistent with the occurrence of temperature inversions during the winter monsoon (Durand et al., 2004). The temperature inversions observed here average $\Delta T \approx 0.4$℃ above T_{ML} during the northeast monsoon in the SEAS, accounting for roughly 23% of all profiles (34% if the restricted SEAS definition employed by previous studies is used (Durand et al., 2004)). Although most temperature inversions fall in the range $0.2℃ \leq \Delta T \leq 0.5℃$, occasionally BL are observed with $\Delta T \sim 1.0℃$. This is consistent with work showing negative atmospheric heat fluxes in some parts of the region where BLs are formed (Shankar et al., 2004; Thadathil & Gosh, 1992, Thompson et al., 2006). Float data indicate that temperature inversions can persist for two or more consecutive profiles during the northeast monsoon (> 20 days), emphasizing the importance of considering this term on seasonal time scales. The values in Figure 3e may overestimate the amount of

cooling given that the monthly values are determined by assuming that the ML deepening persists for an entire month. However, qualitatively these terms confirm that the presence of a BL significantly reduces entrainment cooling and could even contribute to ML warming instead of cooling in the northeast monsoon season, potentially pre-conditioning the ML for the subsequent warming during the intermonsoon period (March-May) (Durand et al., 2004).

Figure 3f depicts the impact of BLs on the heat flux (term (*d*)). Consistent with the shallower ML and therefore increased penetrative radiation, BL tend to have a cooling effect relative to non-BL when considering the heat flux. However, the extent of the difference varies by season. The spring intermonsoon shows the smallest difference; combined with the decrease in entrainment cooling, this suggests increased warming as a result of the presence of BL, potentially significant leading up to the onset of the southwest monsoon. The greatest difference occurs during the southwest monsoon and into the fall, where the presence of BL is associated with a shift from a net warming to net cooling. BL are less common during the fall intermonsoon than either the northeast or southwest monsoon, so their presence is unlikely to lead to cooling; however, their presence may damp the heating otherwise expected during this season.

Figure 4 shows the difference between the temperature tendency resulting from terms (*c*) and (*d*) for BL and non-BL cases in 2º×2º boxes. Positive (negative) values indicate increased (decreased) warming or decreased (increased) cooling associated with BL compared with non-BL. Figure 4 indicates that the effects summarized by Figure 3 are localized and not uniform throughout the SEAS. In agreement with Durand et al. (2004), we find a warming signal off the west coast of India, coincident with where BL are most frequent (Figure 2) and temperature inversions the strongest. The locations of significant relative warming values coincide with a

positive sum of terms (*c*) and (*d*) for the BL cases (not shown), and indicate increased warming on the order of > 0.5ºC month^{-1}.

Cooling is more likely during the southwest monsoon, which reflects the large difference in ML depth between BL and non-BL profiles (Figure 3b). The pattern of cooling, while not significant everywhere, is distributed throughout the region where BL typically constitute more than 40% of the profiles and is on the order of 0.5ºC month^{-1}. Despite strong positive atmospheric heat flux during the spring and fall intermonsoon seasons, we do not encounter a significant increase in the expected warming associated with the presence of BL, possibly because the decreased entrainment cooling is offset by decreased absorption of heat within the ML for most of the basin.

3.6 Discussion

Several hypotheses relating to the impact of BL on SST have been previously proposed: warming (e.g. Lukas & Lindstrom, 1991; Durand et al., 2004; Masson et al., 2005); cooling (Vinayachandran et al., 2007); and insulation of the ML from atmospheric forcing (Vialard & Delecluse, 1998). This work illustrates that all of these effects may occur within the same region depending on the atmospheric fluxes and characteristics of the ML. Increased warming or decreased cooling may occur in the presence of BL during the northeast monsoon (where temperature inversions are most common) and spring intermonsoon (where differences in the impact of the heat flux are minimum), but increased cooling is likely during the southwest monsoon when ML depth differences are maximized. However, in many cases, there is not a statistically significant difference between the BL and non-BL cases (gray shading, Figure 4), despite significant differences in the individual ML characteristics and terms in the temperature

tendency equation (Figure 3). The magnitude of the increased warming is smaller than that determined from modeling studies (e.g. Durand et al., 2004), perhaps due to the lack of near-coast data (where temperature inversions are most frequent) or differences in observed ML depth.

BLs are conducive to enhanced nighttime cooling of the ML during transient negative surface heat flux, and also longer-term cooling due to the strong, sustained salinity stratification. Changes to the ML temperature are a significant factor in determining the magnitude of temperature inversions, where cooling at the surface traps warmer water below. In the long term, this may negate some of the additional cooling allowed by the BL at the surface through vertical entrainment, but is unlikely to fully compensate the effects of large net heat loss at the surface when BL are present. The statistically insignificant difference in BL and non-BL scenarios illustrated in Figure 4 suggests that the warming and cooling effects of BL may be counteracting one another, as suggested by Vialard and Delecluse (1998).

Penetrative radiation in this region is large in comparison with the net surface heat flux. This plays a critical role in determining differences between BL and non-BL scenarios. Other studies in this area identified comparable values of q_{-h} (e.g. Sengupta et al., 2008), but its contribution to the overall temperature tendency warrants additional investigation. The use of a constant extinction coefficient in Equation 3 may lead to overestimates of q_{-h}, likely underestimating any increase in warming associated with BL. Furthermore, while advective processes certainly play a role in forming BL in this region (e.g., Durand et al, 2007), it remains to be seen how the presence of BL may influence advective changes to SST. While each of these may be worth investigating, they are likely to be of lower significance than the first order approximation calculated here.

Finally, the work here has focused on seasonal patterns based on temporal and spatial averages. However, 1-dimensional modeling studies (e.g. Shinoda et al., 2005) have demonstrated the importance of diurnal cycles in affecting intraseasonal variations in SST. These studies show differences in SST when hourly data are used to calculated temperature changes rather than daily averaged (as used here). However, the majority of these studies have been conducted on thermally stratified ML, and thus show ML deepening at night that is unlikely to occur in the presence of BL. Transient temperature inversions associated with night-time cooling are possible in the presence of BL. This suggests that diurnal cycles (and their impacts on intraseasonal SST variation) may differ in the presence of BL; for a study of those processes, a different dataset would be required.

3.7 Acknowledgments

We thank K. Drushka, M. Cronin, and three anonymous reviewers for their feedback and suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also gratefully acknowledge the generous support of Office of Naval Research through NASCar grant N00014-15-1-2254 and NOAA Grant NA15OAR4320063 to the University of Washington, through the Joint Institute for the Study of the Ocean and the Atmosphere. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. *NCEP/DOE 2 Reanalysis data provided by the*

NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at

https://www.esrl.noaa.gov/psd/.

References

Agarwal, N., R. Sharma, A. Parekh, S. Basu, A. Sarkar, & Agarwal, V.K. (2012). Argo observations of barrier layer in the tropical Indian Ocean, *Advances in Space Research*, *50*, 642-654. http://dx.doi.org/10.1016/j.asr.2012.05.021.

Balaguru, K., Chang, P., Saravanan, R., Leung, L. R., Xu, Z., Li, M., & Hsieh, J.-S. (2012). Ocean barrier layers' effect on tropical cyclone intensification. *PNAS*, *109*(36), 14343-14347. https://doi.org/10.1073/pnas.1201364109

de Boyer Montégut, C., Durand, F., Bourdallé-Badie, R., & Blanke, B. (2014). Role of fronts in the formation of Arabian Sea barrier layers during summer monsoon. *Ocean Dynamics*, *64*, 809-822. https://doi.org/10.1007/s10236-014-0716-7

Cronin, M. F., & Kessler, W. S. (2007). Near-surface Shear Flow in the Tropical Pacific Cold Tongue Front. *Journal of Physical Oceanography, 39,* 1200-1215. https://doi.org/10.1175/2008JPO4064.1

Cronin, M. F., & McPhaden, M. J. (2002). Barrier layer formation during westerly wind bursts. *Journal of Geophysical Research*, *107*(C12), 8020. https://doi.org/10.1029/2001JC001171

Durand, F., Shankar, D., de Boyer Montégut, C., Shenoi, S. S. C., Blanke, B., & Madec, G. (2007). Modeling the Barrier Layer Formation in the Southeastern Arabian Sea. *Journal of Climate*, *20*, 2109-2120. https://doi.org/10.1175/JCLI4112.1

Durand, F., Shetye, S. R., Vialard, J., Shankar, D., Shenoi, S. S. C., Ethe, C., & Madec, G. (2004). Impact of temperature inversions on SST evolution in the South-Eastern Arabian Sea during the pre-summer monsoon season. *Geophysical Research Letters*, *31*, L01305. https://doi.org/10.1029/2003GL018906

Echols, R., & Riser, S.C. (2020). Spice and Barrier Layers: an Arabian Sea Case Study. *Journal of Physical Oceanography, in press.* https://doi.org/10.1175/JPO-D-19-0215.1

Ffield, A. (2007). Amazon and Orinoco river plumes and NBC rings: Bystanders or participants in hurricane events? *Journal of Climate*, *20*, 316-333. https://doi.org/10.1175/JCLI3985.1

Flament, P. (2002). A state variable for characterizing water masses and their diffusive stability: spiciness, *Progress in Oceanography*, *54*, 493-501.

Foltz, G. R., Grodsky, S. A., Carton, J. A., & McPhaden, M. J. (2003). Seasonal mixed layer heat budget of the tropical Atlantic Ocean. *Journal of Geophysical Research*, *108*(C5), 3146. https://doi.org/10.1029/2002JC001584

Jackett, D.R. and McDougall, T.J. (1985). An oceanographic variable for the characterization of intrusions and water masses, *Deep-Sea Research.*, *32*(10) 1195-1207.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S-K, Hnilo, J.J., Fiorino, M., and Potter G. L. (2002). NCEP-DOE AMIP-II Reanalysis (R-2), *Bulletin of the American Meteorological Society*, *83*, 1631-1643, https://doi.org/10.1175/BAMS-83-11-1631

Kurian, J., & Vinayachandran, P. N. (2006). Formation mechanisms of temperature inversions in the southeast Arabian Sea. *Geophysical Research Letters*, *33*, L17611. https://doi.org/10.1029/2006GL027280

Lee, C.M., Jones, B.H., Brink, K.H., & Fischer, A.S. (2000). The upper-ocean response to monsoonal forcing in the Arabian Sea: seasonal and spatial variability. *Deep-Sea Researh II, 47*, 1177-1226.

Lukas, R. & Lindstrom, E. (1991). The Mixed Layer of the Western Equatorial Pacific Ocean. *Journal of Geophysical Research*, *96*(Supplement), 3343-3357.

Masson, S., Luo, J.-J., Madec, G., Vialard, J., Durand, F., Gualdi, S., et al. (2005). Impact of barrier layer on winter-spring variability of the southeastern Arabian Sea. *Geophysical Research Letters*, *32*, L07703. https://doi.org/10.1029/2004GL021980.

McDougall, T.J., & O.A. Krzysik, 2015: Spiciness, *Journal of Marine Research,73*, 141-152. https://doi.org/10.1357/002224015816665589.

McPhaden, M.J., & Hayes, S.P. (1991). On the variability of winds, sea surface temperature, and surface layer heat content in the Western Equatorial Pacific. *Journal of Geophysical Research*, *96*(Supplement), 3331-3342.

Mignot, J., C. de Boyer Montégut, & Tomczak, M. (2009). On the porosity of barrier layers, *Ocean Science*, *5*, 379-387.

Niiler, P. P., & Kraus, E. B. (1977). One-dimensional models of the upper ocean. In E. B. Kraus (Ed), *Modelling and Prediction of the Upper Layers of the Ocean,* (pp. 143-172). New York : Pergamon.

Pailler, K., Bourlès, B., & Gouriou, Y. (1999). The barrier layer in the Western Tropical Atlantic Ocean. *Geophysical Research Letters*, *26*(14), 2069-2072.

Rao, R. R. & Sivakumar, R. (1999). On the possible mechanisms of the evolution of a mini-warm pool during the pre-summer monsoon season and the genesis of onset vortex in the south-east Arabian Sea. *Quarterly Journal of the Royal Meteorological Society*, *125*, 787-809.

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M. (2016). Fifteen years of ocean observations with the global Argo array. *Nature Climate Change*. Nature Publishing Group. https://doi.org/10.1038/nclimate2872

Schott, F.A., and McCreary, J.P. (2001). The monsoon circulation of the Indian Ocean, *Progress in Oceanography*, *51*, 1-123.

Sengupta, D., Parampil, S. R., Bhat, G. S., Murty, V. S. N., Ramesh Babu, V., Sudhakar, T. et al. (2008). Warm pool thermodynamics from the Arabian Sea Monsoon Experiment (ARMEX), *Journal of Geophysical Research*, *113*, C10008, https://doi.org/10.1029/2007JC004623

Shankar, D., Gopalakrishna, V. V., Shenoi, S. S. C., Durand, F., Shetye, S. R., Rajan, C. K., et al. (2004). Observational evidence for westward propagation of temperature inversions in the southeastern Arabian Sea. *Geophysical Research Letters*, *31*, L08305. https://doi.org/10.1029/2004GL019652

Shcherbina, A. Y., Gregg, M. C., Alford, M. H., & Harcourt, R. R. (2009). Characterizing thermohaline intrusions in the North Pacific Subtropical Frontal Zone. *Journal of Physical Oceanography*, *39*, 2735-2756. https://doi.org/10.1175/2009JPO4190.1

Shenoi, S. S. C, Shankar, D., & Shetye, S. R. (1999). On the sea surface temperature high in the Lakshadweep Sea before the onset of the southwest monsoon. *Journal of Geophysical Research*, *14*(C7), 15703-15712.

Shenoi, S. S. C, Shankar, D., & Shetye, S. R. (2002). Differences in heat budgets of the near-surface Arabian Sea and Bay of Bengal: Implications for the summer monsoon. *Journal of Geophysical Research*, *107*(C6), 5-1-5-14, https://doi.org/10.1029/2000JC000679.

Smyth, W.D., Hebert, D., & Moum, J.N. (1996). Local ocean response to a multiphase westerly wind burst: 2. Thermal and freshwater responses. *Journal of Geophysical Research: Oceans, 101*(C10), 22,513-22,533. https://doi.org/10.1029/96JC02006

Sprintall, J. & Tomczak, M. (1992). Evidence of the barrier layer in the surface layer of the tropics. *Journal of Geophysical Research*, *97*(C5), 7305-7316.

Stevenson, J.W. & Niiler, P.P. (1983). Upper Ocean Heat Budget During the Hawaii-to-Tahiti Shuttle Experiment. *Journal of Physical Oceanography*, *13,* 1894-1907.

Sweeney, C., Gnanadesikan, A., Griffies, S. M., Harrison, M. J., Rosati, A. J., & Samuels, B. L. (2005). Impacts of shortwave penetration depth on large-scale ocean circulation and heat transport. *Journal of Physical Oceanography*, *35*, 1103-1119.

Thadathil, P. & Gosh, A. K. (1992). Surface layer temperature inversions in the Arabian Sea during winter. *Journal of Oceanography*, *48*(3), 293-304. https://doi.org/10.1007/BF02233989

Thadathil, P., Thoppil, P., Rao, R. R., Muraleedharan, P. M., Somayajulu, Y. K., Gopalakrishna, V. V., et al. (2008). Seasonal variability of the observed barrier layer in the Arabian Sea. *Journal of Physical Oceanography*, *38*, 624-638. https://doi.org/10.1175/2007JPO3798.1
Thompson, B., Gnanaseelan, C., & Salvekar, P.S. (2006). Seasonal evolution of temperature inversions in the north Indian Ocean, *Current Science, 90*(5) 697-704.

Veronis, G. (1972). On properties of seawater defined by temperature, salinity, and pressure, *Journal of Marine Research*, *30*(2) 227-255.

Vialard, J. & Delecluse, P. (1998). An OGCM study for the TOGA decade. Part 1: Role of salinity in the physics of the Western Pacific Fresh Pool, *Journal of Physical Oceanography*, *28*, 1071-1088.

Vinayachandran,P., Shankar, D., Kurian, J., Durand, F., & Shenoi, S. S. C. (2007). Arabian Sea mini warm pool and the monsoon onset vortex. *Current Science*, *93*(2), 203-214.

Wang, W. & McPhaden, M. J. (1999). The surface-layer heat balance in the Equatorial Pacific Ocean. Part 1: Mean Seasonal Cycle. *Journal of Physical Oceanography*, *29*, 1812-1831.

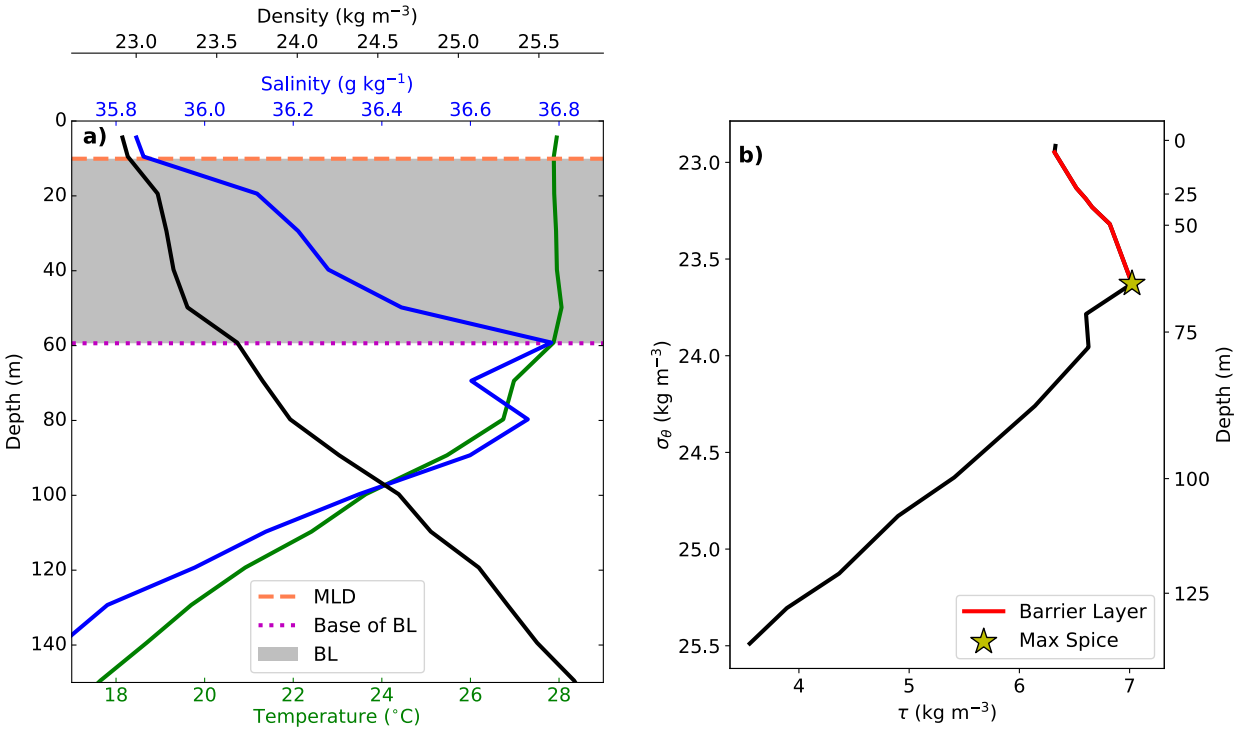

Figure 3.1. (a) Temperature (green), salinity (blue), and density (black) profiles indicating presence of a BL (gray shading). Base of ML (orange) and BL (magenta) indicated by dashed lines. (b) Plot of density vs. spice for profile shown in (a). Red line indicates the BL.

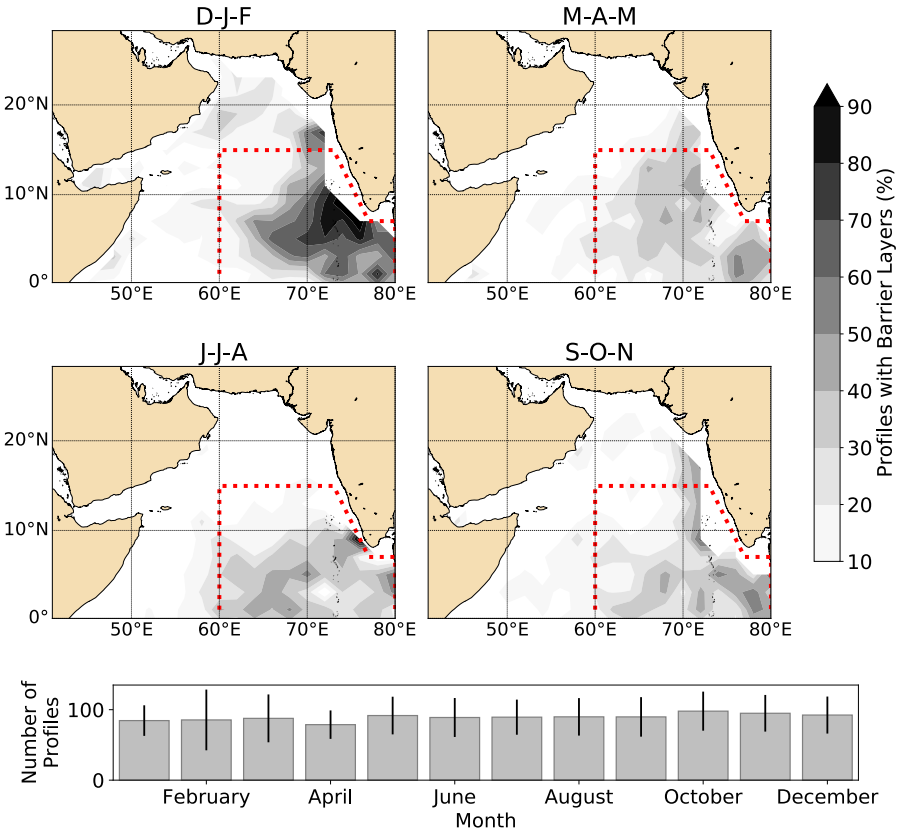

Figure 3.2. BL frequency by season. Red-dashed area indicates SEAS-definition used in this study. D-J-F is the northeast monsoon, December-February; M-A-M is the spring intermonsoon, March-May; J-J-A is the southwest monsoon, July-August; and S-O-N is the fall intermonsoon, September-November. Bottom panel shows average monthly number of profiles in study area; error bars indicate standard deviation across all years.

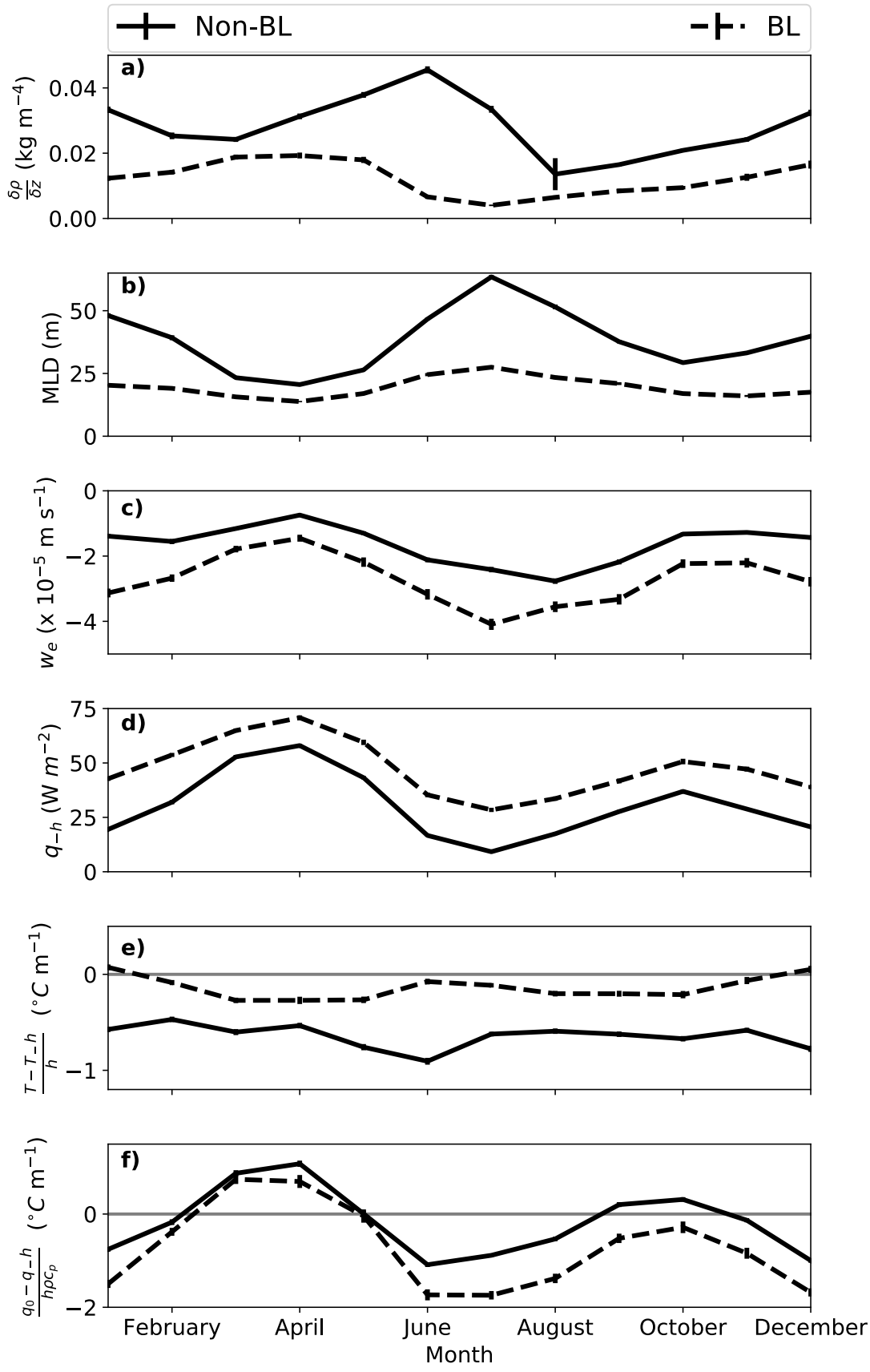
Non-BL
BL
a)
b)
c)
d)
e)
f)
$\frac{\delta\rho}{\delta z}$ (kg m^{-4})
MLD (m)
w_e (x 10^{-5} m s^{-1})
q_{-h} (W m^{-2})
$\frac{T-T_{-h}}{h}$ (°C m^{-1})
$\frac{q_0-q_{-h}}{h\rho c_p}$ (°C m^{-1})
February
April
June
August
October
December
Month

Figure 3.3. Comparison of ML characteristics in the SEAS between BL and non-BL profiles: (a) Stratification below the ML; (b) ML depth; (c) entrainment velocity (Equation 3); (d) penetrative radiation; (e) monthly change in temperature due to entrainment mixing; (f) monthly change in temperature due to heat flux.

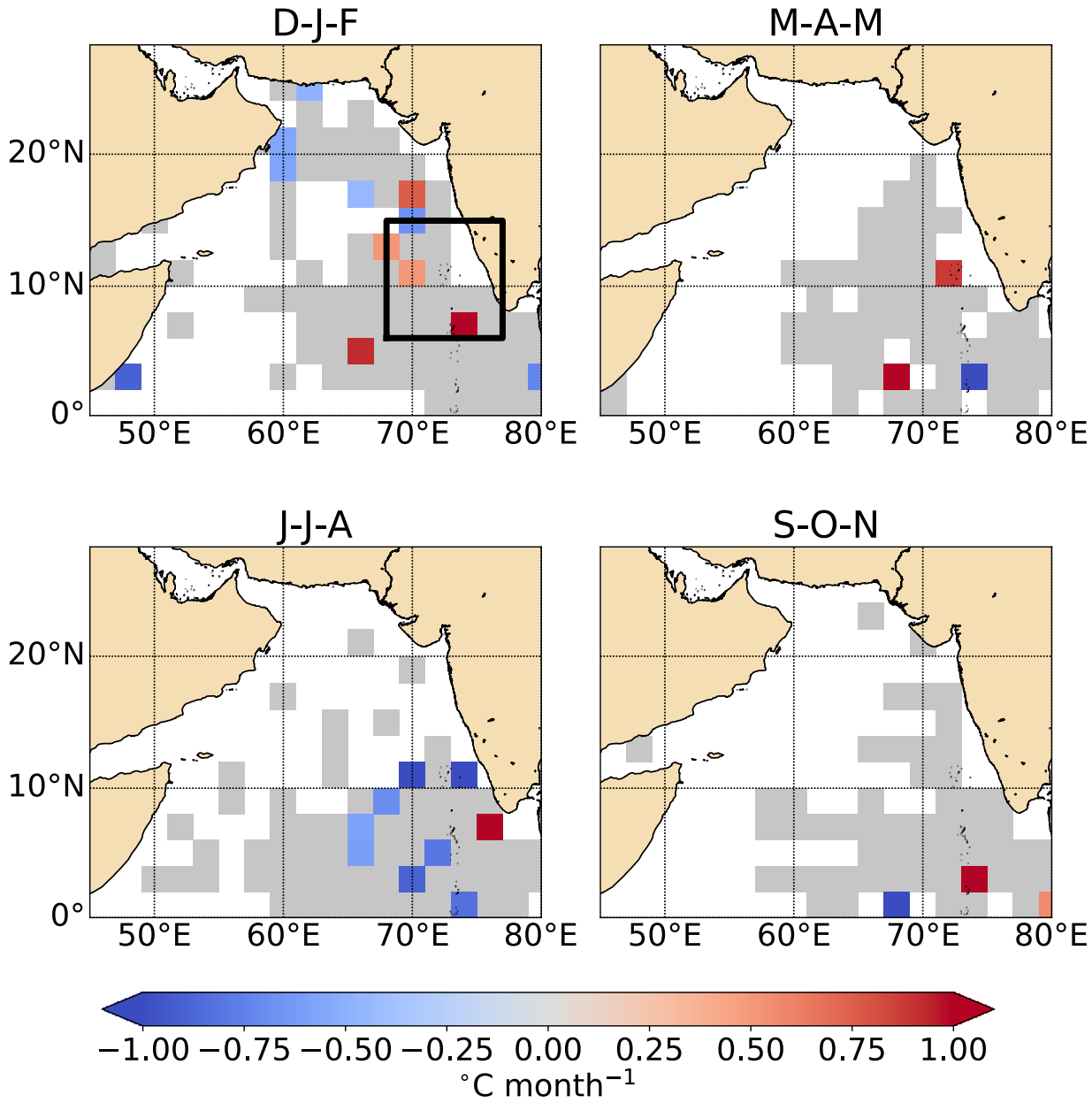

Figure 3.4. Difference in monthly net heating rate between BL and non-BL profiles: $\frac{\partial T}{\partial t}_{BL} - \frac{\partial T}{\partial t}_{non-BL}$. Red (blue) shading indicates increased (decreased) warming or decreased (increased) cooling associated with presence of BL. Grey regions indicate areas where BL are common but difference between BL and non-BL is not statistically significant (p>0.05).

Chapter 4. GLOBAL PATTERNS IN VERTICAL CHLOROPHYLL PROFILES REVEALED BY BIOGEOCHEMICAL FLOATS AND UNSUPERVISED MACHINE LEARNING

4.1 ABSTRACT

Satellite data shows that chlorophyll concentrations can vary over several orders of magnitude at the sea surface, but this is also true in subsurface layers beyond the reach of satellite data. This vertical variation has implications for many topics of interest, but has not been adequately characterized due to a sparsity of data, both in quantity and geographic distribution. Biogeochemical Argo floats have greatly expanded the quantity and spatial distribution of data available for this type of work. Here, we use an unsupervised machine learning clustering technique to identify clusters of chlorophyll profiles with similar characteristics, and then examine the spatial and temporal distribution of the clusters. We find many similarities with previous studies that have examined related problems, but with additional information concerning previously understudied regions and improved resolution of seasonal variability. We find 3 broad classes of profile shapes (deep subsurface chlorophyll maxima, shallow subsurface chlorophyll maxima, and well-mixed profiles), each of which contains a spectrum of profiles. We find strong seasonal and spatial distinctions between these subgroups, particularly in regions with strong seasonal differences in physical forcing and irradiance. In particular, the Southern Ocean shows a strong seasonal cycle in the dominant subgroups. Well-mixed profiles constituting more than 80% of the profiles during winter months, comprising the majority of profiles everywhere except near the coast of Antarctica and Drake Passage. Summer months see a more even distribution between well-mixed and shallow subsurface chlorophyll maximum

profiles, with the latter again being geographically concentrated in a few regions. Finally, we make recommendations for future applications of this work to the fields of biogeography and upper ocean heating.

4.2 INTRODUCTION

The distribution of chlorophyll in the upper layer of the ocean has significant implications for light attenuation and euphotic zone depth (Morel, 1988; Morel and Antoine, 1993; Ohlmann et al, 1996; Sathyendranath and Platt, 1988), upper ocean heating (Manizza et al., 2005; Gnandesikan and Anderson, 2009; Park et al, 2015), and phytoplankton biomass (Arteaga et al, 2020). This distribution varies widely within the top 200m of the ocean (Morel and Berthon, 1989; Uitz et al, 2006; Mignot et al, 2011; Carranza et al, 2018), including variations across several orders of magnitude near the sea surface, with no reliable reference climatologies in existence (Claustre et al, 2020). Satellites can provide a global distribution of the concentration of chlorophyll-a in the surface ocean, and much of what we know about bloom timing (e.g. Levy et al, 2007), relative importance of different parameterizations of shortwave radiation (Manizza et al, 2005), and biogeography (Ardyna et al, 2016; Hardman-Mountford et al, 2008; Reygondeau et al 2013) relies on this information. However, the value of chlorophyll assessed by satellites is an average over the first optical depth (Smith, 1981; Gordon and McCluney, 1975) and therefore is not always reliable for producing estimates of subsurface distributions (e.g. Arrigo et al, 2011; Cornec et al, 2021). This can be problematic in studies of upper ocean heating (Manizza et al, 2005) and phytoplankton biomass (Arrigo et al, 2011). Although biogeochemical and earth system models can provide valuable insight into these processes when coupled to physical models, they also can exhibit significant biases (Lim et al, 2018), making the realistic

assessments of the impact of the chlorophyll difficult. Surface chlorophyll values from satellites are also often insufficient for model validation and tuning (Wang et al, 2020). Therefore, improved representation of the global vertical distribution of chlorophyll is essential for improving our understanding of a whole host of physical and biogeochemical processes.

Subsurface and deep chlorophyll maxima (SCM and DCM) are common features in the world ocean (Cullen, 2015, Baldry et al, 2020). In tropical and subtropical gyre regions, they typically form in response to nutrient-light colimitation (Cullen et al, 2015; Baldry et al 2020), whereas in the Southern Ocean they are more likely to form due to diatom aggregation, sea-ice retreat, eddies, subduction events, and photoacclimation (Baldry et al, 2020); therefore profiles with SCM or DCM may exhibit different characteristics relative to the physical profile in different regions. These subsurface chlorophyll features evolve with time in response to subsurface physical processes such as motion of the thermocline (Cullen, 2015; Ravichandran et al, 2012; Baldry et al, 2020). However, in some regions, the distribution of chlorophyll within the upper ocean mirrors the well-mixed temperature and salinity profiles in the surface layer of the ocean (Carranza et al, 2018; Uitz et al, 2006). Furthermore, the distribution of phytoplankton can change seasonally with availability of nutrients provided to the surface ocean through mixing, restratification of the mixed layer, storm mixing, and light availability (e.g. Carranza et al, 2018), but few datasets are well-suited for investigating such seasonal variability on a global scale.

Several studies have developed globally representative sets of chlorophyll profiles (Morel and Berthon, 1989; Uitz et al, 2006; Mignot et al, 2011). Both Morel and Berthon (1989) and Uitz et al (2006) relied on high performance liquid chromatography (HPLC) data from cruises, while Mignot et al (2011) paired chlorophyll-a fluorescence profiles with HPLC. In both cases,

the authors sought to form an explicit relationship between surface chlorophyll concentration and profile shape. For Morel and Berthon (1989) and Uitz et al (2006), the representative profiles took the form of Gaussian distributions superimposed on a background chlorophyll concentration, whereas Mignot et al (2011) also incorporated sigmoid ("well-mixed") profile shapes into their analysis. These studies encapsulate their datasets extremely well, and in the case of Uitz et al (2006) extend the analysis to other pigments, allowing for the determination of the contributions of different phytoplankton size classes to the total. However, these studies were working with relatively limited datasets: ~3800 profiles in Morel and Berthon (1989) derived from cruise reports and databases; 2419 HPLC profiles in Uitz et al (2006); and 700 paired profiles in Mignot et al (2011). The global distribution included very few profiles in the Indian or Southern Oceans, both of which experience unique and varied conditions, and the reliance on cruise data means that not all seasons are equally well represented and that repeated sampling at a single site, such as one gets from satellites, is unlikely. Also, other more recent studies have shown that variations in the subsurface chlorophyll maximum depth and magnitude are tied to subsurface processes (e.g., Ravichandran et al, 2012), that mixed layer chlorophyll concentration and vertical distribution can change within a mixed layer on time scales different from physical mixing processes (e.g. Carranza et al, 2018), and that the amplitude of deep chlorophyll maxima can vary without changes to the surface chlorophyll expression (e.g., Thushara et al, 2019). The goals of this study are therefore to explore whether a globally representative set of profiles can be developed that captures a broader cross-section of the ocean, allows for exploration of seasonal variability, and does not explicitly tie profile shape to surface concentration or mixed layer characteristics, at least initially.

Autonomous floats equipped with biogeochemical sensors are now globally distributed in the open ocean, gathering not only temperature and salinity data for the upper 2000m of the ocean, but also oxygen, nitrate, chlorophyll fluorescence, and other quantities of interest (Johnson et al, 2017; Claustre et al, 2020). One of the many advantages of these floats is their ability to gather data in regions and at times that are largely inaccessible by ships, such as the Southern Ocean in winter. The data gathered by these floats has already formed the basis for a number of regional studies of biogeochemical processes, including the Arabian Sea (Ravichandran et al, 2012; Prasanth et al, 2021), Mediterranean Sea (Organelli et al, 2017), and Southern Ocean (Carranza et al, 2018; Arteaga et al, 2020; Gu et al, 2020). However, thus far, only one global assessment in patterns in vertical chlorophyll distribution has been carried out using this more comprehensive and widely distributed dataset (Cornec et al, 2021), which focused explicitly on evaluating the varying characteristics of deep fluorescence maxima (DFM) across the global ocean, rather than categorizing profiles into a set of representative shapes (not necessarily DFM) and exploring their seasonal variability. Allowing for the full spectrum of possible profile shapes is important for this work, in order to support future studies of how chlorophyll concentration affects heating, biogeography, and other seasonal patterns in chlorophyll concentration and bloom structure and timing.

The goals of this study require a method for classification that can be used for a very large dataset, can be applied globally, and is agnostic concerning the relationship between surface chlorophyll values and the vertical distribution of chlorophyll. Previous studies have explored the viability of pairing unsupervised machine learning methods with profiles of physical variables gathered from autonomous floats to explore regional patterns in the North Atlantic (Maze et al, 2017) and Southern Ocean (Jones et al, 2019, Rosso et al 2020) and

temporal variability in the Pacific Ocean in relationship with ENSO (Houghton and Wilson, 2020). Common among each of these works, regardless of the focus, was the use of a method for classifying individual profiles into groups that did not rely on predefining the characteristics of each group. Each of these studies also revealed a physically interpretable classification scheme that clearly partitioned profiles into distinct hydrographic regimes, even without explicit input of location when creating the model.

In this study we explore the utility and applications of an unsupervised clustering algorithm to a global dataset of chlorophyll fluorescence profiles and compare the results with previous efforts at defining globally applicable profile classes, biogeographical regions, and defining the characteristics of and mechanisms behind subsurface chlorophyll maxima (SCM). Section 4.3 discusses the dataset used. Section 4.4 discusses the details of the machine learning algorithm and how it has been modified from previous studies as well as other methods used in this study. Section 4.5 discusses the results, including a detailed regional study of the Southern Ocean to demonstrate the interpretability of the method in regions with relatively dense data coverage. Section 4.6 is the discussion, including comparison with previous work and suggestions for further studies.

4.3 DATA

This study uses data from Argo biogeochemical (BGC-Argo: https://biogeochemical-argo.org) floats. BGC-Argo is a globally distributed array of autonomous profiling floats (Argo, 2000; Claustre et al, 2020) carrying both a conductivity-temperature-depth (CTD) sensor for measuring temperature and salinity as well as an array of biogeochemical sensors. Each float typically spends ~9 days at a parking depth around 1000m, then descends to 2000m before profiling upward to the surface with a vertical resolution of ~2.5-5m near the surface (Johnson et

al, 2017; Claustre et al, 2020). Data are immediately transmitted to data collection centers, undergoing real time quality control measures (Argo, 2000; Baldry, 2021; Schmechtig et al, 2014), and are typically available for use within 24 hours. For this study, we used data from 546 floats carrying bio-optical sensors to measure chlorophyll-a fluorescence, spanning the years 2006-2020. Floats were equipped with a WET Labs FLBB, ECO-Triplet, or MCOMS chlorophyll a fluorometer (λ_{ex} = 470nm, λ_{em} = 695 nm for all) (Johnson et al, 2017, Claustre et al, 2020; Cornec et al, 2021). We employed the synthetic vertical profiles in this study, which are an amalgamation of the vertical profiles from each of the individual sensors to the same pressure grid, with data not measured by a particular sensor at a particular pressure indicated with a 'nan'. We used both real-time data and those data that have undergone more extensive manual quality control measures (delayed mode).

Due to inconsistencies between quality control flags for individual data points and overall profiles, we initially retained all available chlorophyll profiles (81,973) and implemented a set of manual quality control checks. Given the focus of this work on upper-ocean processes, we discarded any profile that did not have data in the upper 10m, profiles whose maximum pressure was less than 200m, and profiles with a pressure difference larger than 20m in the top 200m. We also discarded any profile with fewer than 10 non-'nan' data points (in the synthetic BGC-Argo files, some 'nan' values indicate the merging of the physical and biogeochemical sensor data, and should therefore be ignored). Profiles without a valid latitude and longitude were also discarded, as well as profiles whose chlorophyll concentration below 200m was greater than 0.3 mg/m^3, and profiles with fewer than 10 non-negative real values. Of these manual QC checks, the pressure issues were by far the most frequent problem, affecting ~5800 profiles, largely profiles classified as "good" by the real-time quality control checks.

After automatic and manual quality control, 71,272 profiles remained. Of these, 8,639 had 'C'-'F' QC classifications (12%) and 11,581 had no QC tag at all (16%). The profiles are distributed roughly evenly by month, but they exhibit an uneven geographical distribution (**Figure 4.1**), largely due to the efforts of the Southern Ocean Carbon and Climate Observations and Modeling project (http://soccom.princeton.edu) to deploy Argo floats in the Southern Ocean (Johnson et al, 2017).

Raw chlorophyll fluorescence values are susceptible to several potential issues, namely bias in the calibration and non-photochemical quenching during daytime conditions (Johnson et al, 2017). Roesler et al (2017) identified a global averaged factor of 2 bias in the WET Labs ECO sensor calibration, which is accounted for in the quality controlled data (Johnson et al, 2017). However, regional differences in this factor, which ranges from <1 for the Black Sea to ~6 for the Southern Ocean (Johnson et al, 2017), are not accounted for. Therefore any absolute values discussed here may be subject to influences from the regional variability in this bias. Non-photochemical quenching primarily affects near-surface values, as it is a photo-protection mechanism by which the fluorescence-to-chlorophyll-a ratio decreases without an actual change in the chlorophyll concentration (Xing et al 2012; Thomalla et al, 2018). This can introduce subsurface chlorophyll maxima into the data where none exist or significantly bias the near-surface values reported (Baldry et al, 2020). Some float profiles have been corrected for this effect, where viable (Schmectig et al, 2014), but we explore the effects of this phenomenon on the clustering as well.

4.4 METHODS

4.4.1 Data pre-processing

Prior to performing any clustering, the data were cleaned and pre-processed, which is critical for producing interpretable and useful results with machine learning clustering methods. We applied a 5-point median filter to each profile, which reduces spikiness and eliminates erroneous single point maxima (particularly at the surface) without necessitating discarding the profile. Some previous studies working with chlorophyll fluorescence profiles have used a 3-point median filter (Carranza et al, 2018), but we found that this did not adequately a suppress single point maxima in all cases. In some cases, using a 5-point filter decreases the magnitude of a subsurface chlorophyll maximum (SCM), but other characteristics of the shape (depth of SCM, width of SCM) remain the same, suggesting suitability for the purposes of clustering. Following the filtering, we interpolated to a 5-meter grid from the surface to 250 meters to create a uniform grid for all profiles (results were similar with interpolation to a 1-m grid). Finally, following Carranza et al (2018), we subtracted the smallest "deep sea" value to account for potential offsets in the calibration (except in the Black Sea, where fluorescence values increase at depths shallower than 250m due to deep sea red fluorescence (Xing et al, 2016 and references therein); note that this phenomenon also occurs in other basins but often at depths greater than those used for this study).

In order to reduce dimensionality and prioritize a focus on overall profile shape rather than the absolute chlorophyll value at each depth, we perform principal component analysis (PCA) on the smoothed, interpolated profiles (Jones et al, 2019, Maze et al 2017, Rosso et al 2020, Houghton and Wilson, 2020). PCA uses singular value decomposition to project the data onto a set of orthogonal basis functions. Applying PCA to our dataset allows us to represent it as a linear combination of orthogonal principal components (PC) rather than retaining data at every individual depth. The PCs are calculated using the transformation:

$$X(z) = \sum_{n=1}^{N} P(z,n)Y(n) \quad (1)$$

where *N* is the total number of Principal Components, *z* is the pressure level, *Y(n)* is the nth principal component, and *P(z,n)* is the transformation matrix between the original pressure-level space and the orthogonal basis space. Details on the execution of PCA and other methods detailed here are in Appendix A.

Chlorophyll values skew strongly toward low values with a very long tail towards higher values (Figure S1). Therefore, before performing the PCA, we log-normalize the data at every depth. This produces a close-to-normal distribution at each depth. To compensate for negative chlorophyll concentrations in some profiles we subtract the smallest (most-negative) value across the whole dataset from every profile prior to log-normalizing. This maintains the correct relative magnitude between all profiles. Negative values for chlorophyll concentration are physically unrealistic in absolute terms, but can be correct relative to the rest of their profiles and negative due to a combination of calibration and interpolation effects. No further normalization was performed prior to implementing PCA, in part because we expect all values to approach 0 in the deep ocean and do not wish to introduce scaled variance at a depth where none should exist.

Unlike the previous studies that have focused on temperature data from more constrained regions (Jones et al, 2019, Maze et al 2017, Houghton and Wilson, 2020), the first principal component here explains only about 45% of the variance. This is to be expected from a global study of a variable that can change by several orders of magnitude across seasonal time scales. Explaining 99% of the variance in the dataset requires 17 PCs, whereas In order to explain 99.9% of the variance, as has been done in previous studies, 45 components would have been

required, at which point there would be negligible dimensionality reduction. 17 components represents a factor of 3 reduction in dimensionality for this dataset (5-m grid, 0-250m), and produces significantly better results than without PCA (repeatability, similarity between clustered profiles), and so was considered satisfactory. The characteristics of this decomposition (percent variance explained by each PC, number of components required to explain 99% of the variance) may change as the number and distribution of float profiles change in the future. The 17 PC used are shown in Figure S2.

4.4.2 Gaussian Mixture Model

Clustering methods fall into two categories: supervised and unsupervised. Supervised methods rely on datasets with known categorization to train a model, which is then deployed on unlabeled data. Unsupervised methods use unlabeled data and sort it into clusters in ways that minimize some statistical characteristics, such as in-group mean values. Unsupervised methods are preferred here to allow for feature discovery. In either case, the number of clusters into which individual items can be sorted must be specified ahead of time. This is a non-obvious and non-trivial aspect of the clustering problem: how do we know how many clusters to create, if the whole goal of the study is to identify potentially unknown or understudied characteristics in chlorophyll profiles? Previous studies have used the Bayesian Information Criterion (BIC) with great success to determine a range of optimal values for the number of clusters (Jones et al, 2019; Maze et al 2017). The BIC score provides a trade-off between quality of fit and number of clusters. If we select the number of clusters to be equal to the number of profiles, the fit will be perfect but the penalty for having that so many clusters will be enormous (and we also will not be able to identify any broader patterns). If we retain a very small number of clusters, the fit of

the model will be very poor, but that will be offset to some degree by the efficiency of having a small number of clusters (at the extreme, we are again unlikely to learn much because many profiles will be lumped together that we might visually classify as dissimilar). The ideal is somewhere in between these two extremes. In this case, we use a randomized subset of the profiles (50 different subsets, randomly selecting 1 profile from each 1×1 degree box with data) and calculate the BIC score for a range of possible cluster numbers between 10 and 30 (this range was determined from trial and error; previous studies focused on temperature data have identified an optimal number of clusters <10). The subsets do not fully compensate for the fact that some regions are more heavily sampled than others, but this does imply that in sparsely sampled regions every available profile is used for the calculation whereas in heavily sampled region each subset only includes some of the values. Each of the 50 subsets includes 9738 profiles, which represents ~13.15% of the total dataset. The BIC scores (mean, standard deviation) are shown in **Figure 4.2**. As with previous studies, there is not a clear single minimum, but a range of minimum values between 17 and 23. For the remainder of this analysis, we will use 21 clusters, but as we discuss below, the general patterns do not change if a slightly smaller or slightly larger number of clusters is used.

The final step is to create the clusters. In selecting the method to use here, we identified several desirable requirements: (1) initial flexibility in number of clusters; (2) no explicit prescription of linkage between surface chlorophyll and cluster assignment; and (3) to be agnostic concerning the expected shapes observed in chlorophyll profiles. Carranza et al (2018) showed that a wide range of vertical chlorophyll profile shapes are possible in just the Southern Ocean alone, and given the quantity of data available we were reluctant to pick a method that might a priori dictate specific features. Following previous work in clustering and classification

using Argo float data (Maze et al, 2017; Jones et al, 2019; Rosso et al, 2020; Houghton et al, 2020) we chose to use an unsupervised machine learning model for clustering, in this case a Gaussian Mixture Model (GMM). A Gaussian Mixture Model represents the data as a linear sum of multidimensional Gaussian distributions (in which the means and standard deviations are initially unknown). Each individual cluster is described by a particular multidimensional Gaussian probability density function (PDF):

$$\mathbb{N}(x; \mu_k, \Sigma_k) = \frac{exp\left[-\frac{1}{2}(x-\boldsymbol{\mu}_k)^T \Sigma_k^{-1}(x-\boldsymbol{\mu}_k)\right]}{\sqrt{(2\pi)^D |\Sigma_k|}} \quad (2)$$

where $\boldsymbol{\mu}_k$ represents the vector of means of the Gaussian distributions, Σ_k is the covariance matrix, and *x* is a single profile taken from the set. The entire dataset is described as a linear weighted sum of these individual multidimensional Gaussian PDFs as

$$p(X) = \sum_{k=1}^{K} \lambda_k \mathrm{N}(X; \mu_k, \Sigma_k) \quad (3)$$

where λ_k is the probability associated with class *k* and *X* is the entire array of profiles (or in this case, principal component weights).

Cluster assignment is performed using an expectation-maximization approach (Maze et al, 2017), minimizing in-cluster variance of the mean and standard deviation. Unlike a method like K-Means, which only attempts to minimize variance of the mean within a cluster, GMM allows for variability within a group (Jones et al, 2019; Maze et al, 2017). In addition, because the data is represented as a PDF, each individual item (a profile, in this case), receives a probability of belonging to each cluster. This last aspect is both physically realistic, in that it

reflects that profiles in the ocean are likely to fall along a continuum of possible behaviors, and is also potentially useful for examining spatial and temporal transitions between clusters. Unlike Jones et al (2019), we do not use a training subset to create the model. We expected significant regional and seasonal variability and did not want to risk excluding data points that were essential for characterizing certain clusters when creating the final model, given the unevenness of the coverage in some areas.

4.4.3 Day/Night calculations

We do not correct for NPQ beyond the corrections included in the data files. In order to gauge the possible impact of this behavior on the clustering results and the average profiles produced from clustering, we determine whether each profile occurred during the day or not. Although quenching effects can extend several hours after sunset, Carranza et al (2018) found that there were no statistically significant differences in the number of profiles with deep fluorescence maximum if they extended the daytime period, so here we have done a simple test for daytime or nighttime in separating profiles. Approximately 36% of the total number of profiles occur during nighttime (38.5% for the Southern Ocean; note that this is different from the result found by Carranza et al (2018), but the number of floats has expanded significantly since that study).

4.4.4 Profile curve-fits

To facilitate comparison with previous studies, we apply mathematical fits to cluster average profiles as well as to individual profiles to develop in-cluster statistics. We test 6 possible fits, drawing on the profile representations presented in Uitz et al (2006), Mignot et al

(2011), and Carranza et al (2018): (1) Gaussian (very low surface chlorophyll with distinct subsurface peak); (2) Sigmoid (well-mixed near-surface layer); (3) Exponential (very elevated surface concentrations rapidly decaying); (4) Gaussian + Line (elevated surface values with pronounced subsurface peak); (5) Gaussian + Sigmoid (well-mixed surface layer with peak at the base, secondary subsurface peak, or more rapid decay below a well-mixed layer than can be captured by a sigmoid alone); (6) Gaussian + Exponential (rapid decay where the maximum chlorophyll is below the surface). For each curve fit, we supplied initial guesses for the fit characteristics, similar to Carranza et al (2018). We apply the curve fitting algorithm to each smoothed, interpolated profile, and retain the Chi-square statistics and the Akaike Information Criteria (AIC) statistic for each fit. We use AIC here rather than BIC because it has a reduced penalty for an increased number of parameters relative to BIC. In order to select the best fit, we identify the minimum Chi-square and AIC values. If they do not agree, we select the fit that has the fewest parameters of the two best choices. There were 1187 profiles where one or more of the curve fits did not converge even with excellent initial guesses. However, there were no cases where none of the curve fits converged, so we simply retained the best option of the fits that did converge. We show two sets of example fits in Figure S3.

4.4.5 Profile shape characteristics

For mixed layer depth calculations, we use a 0.03 kg m^{-3} threshold from 10m density (de Boyer Montegut et al, 2004), which has been shown to be suitable for global datasets, particularly when developing bulk regional statistics. We identify the depth and magnitude of maximum chlorophyll by using a simple maximum over the whole profile. This is very effective for profiles with a pronounced Gaussian shape (Figure S4a,c) although potentially somewhat less

effective for profiles with a sigmoid-like shape (Figure S4b), especially if there is any irregularity in the "well-mixed" region. When comparing features of cluster-average profiles, we only calculate the depth of maximum chlorophyll for profiles that have a pronounced subsurface peak. For the cluster-average profiles that are roughly sigmoid in shape, we focus instead on the depth at which chlorophyll starts to decrease significantly (e.g. the chlorophyll mixed-layer depth) using a threshold value of 0.01 mg m^{-3}. We tested thresholds up to 0.05 mg m^{-3} and the results are qualitatively similar.

4.5 RESULTS

4.5.1 Cluster characteristics

Each individual Argo profile is assigned to the one of the 21 clusters with which it has the highest posterior probability, or the greatest degree of similarity. The distribution between different clusters is shown in **Figure 4.3**. In total, approximately 72% of the profiles have a posterior probability of 90% or higher of association with their cluster group, and 84% with a probability of 75% or higher. This is comparable to the results of Jones et al (2019) for their study of temperature profiles in the Southern Ocean, which found that 74% of their profiles exceeded 90% probability and 86% exceeded 75%. Clusters with particularly high overall posterior probabilities include clusters 7, 10, and 14 (Table 1), which indicates a very high degree of similarity between profiles in these groups. Clusters with relatively low overall posterior probabilities include 16 and 17; this could be an indication of poor cluster match and/or an insufficient number of clusters to describe the variability in the dataset, or it could indicate profiles that are in a transitional region or time between the properties associated with a cluster. We discuss this again later in Section 4.5.2, where we focus on the Southern Ocean. Additional

characteristics for each cluster are also shown in **Table 4-1**, including median, minimum, and maximum surface chlorophyll and seasonal distribution. The values for surface chlorophyll are meant to be descriptive rather than absolute, as Roesler et al (2017) showed regional differences in the accuracy of calibration for WET Labs ECO sensors, which have not been accounted for here beyond quality control applied to the original profiles (Schmectig et al 2014, Baldry, 2021). However, even with this potential ambiguity in mind, there is likely to be significant overlap in the range of surface chlorophyll concentration between clusters, particularly if NPQ is affecting daytime values. In addition, many of the minimum values register as a negative number, but are within the measurement error of 0.04 mg m^{-3} from 0 mg m^{-3}, so we do not consider this cause for concern. Finally, we note that many of the standard deviations exceed the median value, suggesting that values >1 standard deviation below the median would be very negative. In fact, because chlorophyll exhibits a highly skewed distribution, the standard deviation does not characterize the data as neatly as it would for data with a more Gaussian-like distribution.

Weighted average profiles for each cluster are shown in **Figure 4.4**. The weights used are the posterior probability of each profile for that cluster. This average highlights the dominant features of the profile in each cluster; using the median value instead reveals a similar overall shape but with lower chlorophyll concentrations, particularly at depths with the highest concentration. The clusters in **Figure 4.4** are sorted by the surface chlorophyll value of their weighted average profile and fall into three broad categories: clusters with low surface chlorophyll and a prominent SCM (“Deep SCM group”, 1-3, 5-7; 17612 profiles, ~25%); clusters that appear well-mixed near the surface and then exhibit a sharp decrease towards very low concentration (“Well-Mixed Group”, 4,8-10, 12, 13, 15, 17, 19; 30664 profiles, ~43%); and clusters that exhibit some combination of higher surface values and a subsurface chlorophyll

peak (“Shallow SCM group”, 11, 14, 16, 18, 20, 21; 22996 profiles, ~32%). We use these three broad descriptions to create larger sub-groups of clusters, which provides an opportunity for additional model validation beyond just examining the highest posterior probability. The exact number of profiles in each group varies somewhat according to the subjective assignment of profiles to a subgroup and initialization of the model but are fairly consistent. **Figure 4.5**a-c show the clusters sorted into the three main subgroups. We also plot the highest posterior probability versus the second highest posterior probability as a fraction of the total possible probability for each group (**Figure 4.5**d). This allows us to investigate the likely meaning of lower posterior probabilities and/or secondary posterior probabilities. For most clusters, the highest secondary probabilities are associated with clusters in the same general subgroup, suggesting generally similar profile characteristics but with slight variation. For example, the most common secondary probabilities for Cluster 2 are associated with Cluster 1 and Cluster 3, which are also clusters with prominent deep SCM, but with slightly deeper or shallower peaks than in Cluster 2. This is reassuring evidence that the clustering is performing well and that many of the clusters with posterior probabilities lower than 75% are likely equally similar to two well-established clusters.

Previous studies have highlighted the relative location of the maximum chlorophyll and mixed layer depth as a means of distinguishing formation mechanisms for the subsurface chlorophyll maximum. **Figure 4.6**a shows the relative position of the depth of maximum chlorophyll (deep and shallow SCM groups) or the base of the well-mixed chlorophyll layer (well-mixed groups) and the mixed layer depth. A key difference between the deep and shallow SCM groups is the relative position of the base of the mixed layer in comparison with the depth of maximum chlorophyll. Peak chlorophyll occurs significantly deeper than the mixed layer

depth for SCM groups, whereas with shallow SCM groups the depth of maximum chlorophyll is relatively close to the base of the mixed layer, although still typically somewhat deeper. The deep SCM group also occurs almost exclusively in the tropics and low subtropics, whereas the shallow SCM group is more common at higher latitudes (**Figure 4.6**b), suggesting that the clustering into these broad groups may also reflect likely mechanisms. This difference in geographic location for the groups, paired with the uneven data coverage shown in **Figure 4.1**a, likely contributes to the percentage of total profiles each of the subgroups accounts for. The well-mixed clusters show close alignment between the base of the density mixed layer and the base of the well-mixed chlorophyll region, implicating physical forcing as the defining mechanism for producing that profile shape.

Figure 4.7a shows the predominant group number plotted as a function of month and latitude in a modified Hovmoller diagram, allowing us to visualize the spatial and temporal distribution of the various clusters. All profiles for each month are compressed onto a single line (combining all the basins) and seasons are normalized to the north, so that austral and boreal winter can be plotted simultaneously. The groups are sorted from low surface chlorophyll (light yellow, Group 1) to high surface chlorophyll (dark purple, Group 21), as shown in **Figure 4.4**. Recognizable patterns in surface chlorophyll are immediately visible, including the very low surface chlorophyll concentrations expected in the subtropical gyre regions (light yellow bands around 25°N and 25°S latitude), and much higher chlorophyll concentrations during high-latitude summers (dark purple groups). By separating these results out into major basins and excluding data from the Mediterranean and Black Seas (**Figure 4.7**b), it can be seen how some of these global patterns vary by basin. The Northern Indian Ocean (NIO), comprising the Arabian Sea and Bay of Bengal, shows markedly different distribution of clusters from the

Pacific and Atlantic basins at comparable latitude, although coverage between the three basins at those latitudes is quite different. Unlike the subtropical gyre regions in the Atlantic and Pacific which are oligotrophic year round, and thus are almost exclusively characterized by profiles with deep subsurface chlorophyll maxima, the Arabian Sea is only seasonally oligotrophic. During both the northeast and southwest monsoon, coastal upwelling increases nutrient supply allowing for increased surface chlorophyll and a shift towards profiles more likely to exhibit a shallow subsurface chlorophyll maximum (**Figure 4.7**b, Figure S5; Lévy et al, 2007; Schott and McCreary, 2001). Similarly, the Equatorial Pacific (EPO) and Indian Oceans (EIO) are quite distinct. The EPO is dominated by Cluster 16 whereas the EIO is dominated by Cluster 7. The EIO has a strong physical influence from the freshwater flowing out of the Bay of Bengal which blocks mixing between the surface and depth, limiting nutrient availability in the surface ocean. Clusters 7 appears almost exclusively in the northern and equatorial Indian Ocean, while Cluster 3 and 6 appear in those regions as well as sparse distribution in other subtropical regions. This stands in contrast with Clusters 1 and 2, which dominate the subtropical gyres year round, and Mediterranean Sea, primarily in the summer and fall, with only a minor appearance in the northern Bay of Bengal.

One concern with unsupervised clustering methods is that the wrong number of groups might be picked into which to separate our items. We assess this here by comparing the representative average profiles for both a smaller and larger number of clusters within the minimum BIC range shown in **Figure 4.2** (Figure S6) as well as the general spatial and temporal patterns for those classification schemes (Figure S7). Figure S6 shows that the individual average profiles capture the same general features whether the number of clusters is 17, 21, or 24, and that the major profile features characterizing each of the major subgroups does not change if we

select a smaller or larger number of clusters. The approximate number of profiles falling into each of these major subgroups is also fairly similar (not shown). Given that the distinction between these major subgroups is somewhat arbitrary, small shifts in the number of profiles falling into each group are to be expected. The major change observed here with using a smaller or large number of clusters is the degree of separation of features within each category: how many different deep SCM clusters we have into which to separate out profiles with those features (rather than whether or not those features exist). It is thus reassuring that we will likely identify similar overall patterns seasonally and regionally regardless of where along the general minimum for BIC score the cluster number is selected. Figure S7 likewise shows qualitatively similar spatial and seasonal patterns across each of the three cluster numbers. As such, at seems that our choice of optimal cluster numbers for this work revolves around the degree of distinction preferred within a subgroup.

More robust assessment of regional and temporal patterns, including relationships with physical properties, requires high data density, which is only available for some regions. We focus on one region of particular interest below, the Southern Ocean (Section 4.5.2), which has the best overall coverage of the regions in which biogeochemical Argo data is available and also has been the topic of several recent related studies.

4.5.2 Seasonal evolution of the Southern Ocean

The Southern Ocean (here considered to be latitudes < 40° S) contains ~33% of the profiles used for this study (23,622 profiles) and exhibits a much higher concentration of profiles than most other regions. **Figure 4.8**a shows the location of each of the profiles used here, classified by cluster as described above. **Figure 4.8**b shows the seasonal distribution of profiles

within each of the major subgroups. Notably, almost all of the profiles observed in the Southern Ocean fall either within the well-mixed subgroup or the combination subgroup; deep chlorophyll maxima associated with very low surface concentrations are rarely observed and are most often at the northernmost edge of the region. Seasonally, well-mixed clusters dominate during the austral winter season (June-August), comprising >80% of the total profiles during that 3-month window. During the summer months, the distribution is more evenly divided between the well-mixed and combination subgroups. This is similar to the study done by Carranza et al (2018) that found that profiles with a subsurface chlorophyll maximum constituted more than 50% of the summer profiles.

Figure 4.9a shows the annual concentration of profiles in 5×5 degree boxes (the gray shading indicates fewer than 10 profiles). **Figure 4.9**b shows the geographical distribution of profiles in each subgroup (well-mixed or shallow SCM) annually. Well-mixed profiles dominate in the core of the Atlantic Circumpolar Current (ACC), whereas shallow-SCM profiles are more common closer to the Antarctic continent and near the subtropical front and land masses (particularly around Drake Passage and southern New Zealand). **Figure 4.9**c and d show this same information divided into winter (May-September), which has low light, and summer (October-April) half year segments (boxes with fewer than 5 profiles are shaded gray). This highlights which regions are responsible for the seasonal shift in distribution shown in **Figure 4.8**b. The central ACC is dominated by well-mixed profiles in both summer and winter. The biggest seasonal shifts are observed in the near-continental regions, Drake Passage, and southern New Zealand, where shallow SCM groups form the majority of profiles in the summer and in some cases form the majority of profiles observed in the yearly average.

We consider seasonal variability in cluster prevalence and associated physical and biogeochemical variables in the Southern Ocean in **Figure 4.10** and **Figure 4.11**. **Figure 4.10** shows seasonal distribution across the two major subgroups (well-mixed and shallow SCM). **Figure 4.11** elaborates on the distribution of physical and biogeochemical properties within the clusters observed in the Southern Ocean. Because we are only considering clusters from this region, the relative magnitude of chlorophyll concentrations between groups is less likely to be influence by the calibration issues identified by Roesler et al (2017), and therefore more meaningful. The three most dominant clusters in winter are Clusters 8, 9, and 13, which correspond with the clusters with the deepest mixed layers within the well-mixed category, and Cluster 11, which is the shallow SCM cluster with the deepest median mixed layer (**Figure 4.11**a). Although there is significant overlap across groups, the median surface chlorophyll shows clear patterns corresponding to the seasonal dominance of each cluster (**Figure 4.11**b). During the summer, the contributions from the well-mixed group likewise shift to clusters with shallow mixed layers and higher chlorophyll contributions (most notably SCM group Cluster 16), with negligible contributions from those clusters that were most dominant during the winter. Although we do not break down the physical and biogeochemical properties by season, the shift in group prevalence suggests that the cluster characteristics most likely reflect their seasonal occurrence. The alignment between the physical and biological characteristics of the clusters and the season in which they occur supports the effectiveness of the clustering. Plots of surface dissolved oxygen (**Figure 4.11**c) and SST (**Figure 4.11**d) do not provide a clear distinction between groups, as there is significant overlap across all groups, except to the extent that the median dissolved oxygen concentration of the group with the lowest median surface chlorophyll

concentration (group 4) is lower than the median dissolved oxygen concentration of that with the highest surface chlorophyll (group 21).

Although there are relatively few profiles with posterior probabilities below 50%, it is worthwhile to consider their location, distribution, and possible explanation. **Figure 4.12**a shows the posterior probabilities for each profile plotted seasonally. Profiles with a posterior probability less than 50% are shown in red. Lower probability profiles seem to be somewhat concentrated downstream of the Drake Passage and Kerguelen plateau, and otherwise distributed randomly. Seasonally, these low probability profiles are most often associated with clusters 16, 17, and 19 (**Figure 4.12**b) in the Southern Ocean, during the summer and fall (**Figure 4.12**c), which are the seasons with generally higher chlorophyll concentrations in the Southern Ocean. Profile 16 is part of the shallow SCM group, whereas profiles 17 and 19 are part of the well-mixed group. Examples of some lower-probability profiles are shown in **Figure 4.12**d. Given that the well-mixed profiles in particular mostly present a similar shape but with a range of depths associated with the base of the well-mixed portion of the profile, it is not surprising that some of these profiles are somewhat ambiguous in terms of the best match of cluster. In some cases, these lower probability profiles match the general shape of the cluster average but have a significantly different amplitude from the numerical value of the cluster average. In isolated cases, a distinct subsurface chlorophyll maximum appears in a cluster of profiles that is collectively part of the well-mixed group, which suggest misclassification due to other feature similarities with that group. As a whole, low-probability profiles represent a small percentage of the overall profiles, and actual potential misclassifications where there is no meaningful overlap with the other cluster profiles are not a large portion of the total.

Previous work has considered the potential influence of non-photochemical quenching (NPQ) on the observed profile shapes and measured chlorophyll values, both in the Southern Ocean (Carranza et al, 2018) and elsewhere (e.g. the Arabian Sea: Prasanth et al, 2021). To assess the potential impact of NPQ on the clustering, including whether the shallow SCM profiles are real or an artifact of NPQ on the near surface values, we divide the profiles into daytime and nighttime bins and compare the reported weighted average surface values and vertical profiles for each group. Again, by only focusing on the Southern Ocean profiles for this analysis, we hope to avoid any potential influence of regional differences in the appropriate calibration of fluorescence:chlorophyll (Roesler et al, 2017). To determine whether the surface values in each group are statistically different, we use a non-parametric t-test (Mann-Whitney) for the day and night bins for each cluster. These results are shown in **Figure 4.13**a. With the exception of Group 12, every cluster shows a statistically significant difference between the day and night surface values, with nighttime values being higher. In general, clusters with a higher median surface chlorophyll also show a larger day-night difference: clusters 8, 9 and 10 have an average percentage difference between daytime and nighttime of ~11%, whereas for clusters 19 and 21, the daytime surface values are only ~36% of the nighttime surface values. **Figure 4.13**b shows a representative subset of the weighted average profiles for 4 clusters, illustrating the difference between day and night profiles within a particular cluster in the Southern Ocean. The most significant offsets in chlorophyll value are observed at the surface, as expected, and decrease with depth until the profiles eventually match up. Although this does not fully address the question of whether any of the observed SCM features are artificial or not, these profile comparisons show consistency in key features, such as depth of SCM, depth of the well-mixed chlorophyll layer, and general rate of change of chlorophyll at the base of the mixed layer (where

applicable). Furthermore, the original global clustering was repeated using only nighttime profiles and with the exception of fewer PCs required to describe 99% of the variance (vs. 17) and slightly higher chlorophyll concentrations in the weighted average profiles, the patterns identified are very similar.

4.6 Discussion

4.6.1 Profile characteristics

Both Uitz et al (2006) and Mignot et al (2011) found two broad categories of expected profile shapes for the global ocean: those with a deep subsurface chlorophyll maximum (a generalized Gaussian superimposed with a linear decrease in chlorophyll), and those with a well-mixed upper ocean (sigmoid profiles). Carranza et al (2018) found that the Southern Ocean was also largely described by these classes, but with more temporal variability in the exact shape of the profile, allowing for more complex combinations of mathematical descriptions. In this study, we did not dictate any shape characteristics for clusters a priori, but do find many of the same features. Qualitatively, the weighted-average profiles appear to fall into three main classes: a Gaussian (deep SCM clusters); a sigmoid (well-mixed clusters); and a combination of several influences, such as a Gaussian and Sigmoid together (shallow SCM). To pursue this further, we performed curve-fits on both the weighted average profiles for each cluster as well as the individual profiles classified in each cluster to observe patterns in how they are best represented mathematically, and how the results from a curve-fitting approach might differ from the machine learning approach used here. **Table 4-2** shows the distribution of profiles between the 6 possible curve fits allowed (see also Supplemental Figure 7). The most common fit was a Sigmoid superimposed on a Gaussian. These profiles sometimes visually match a pure sigmoid category,

but either have a rate of decrease of chlorophyll below the mixed layer or a secondary chlorophyll maximum that makes this mathematical combination more suitable. Secondary chlorophyll maximum profiles have not been considered at all in previous studies, and are mathematically complex to represent; they also constitute a sufficiently small percentage of our overall dataset that they did not form their own cluster. Although the percentages shown in Table 2 may shift as the data coverage evolves over time, it nonetheless seems clear that exponential profiles are the least common.

The results of the individual curve fits align qualitatively with the weighted-average characteristics of each group (**Figure 4.14**): the deep SCM groups are almost always represented best by either a pure Gaussian or a Gaussian superimposed on a line (agreeing well with the results of Uitz et al (2006), and their assessment of a continuum of Gaussian shapes with elevated surface chlorophyll). However, some of these groups, especially the ones with generally higher surface chlorophyll concentrations (e.g. group 6) also have a significant number of profiles that are well represented by a Sigmoid-Gaussian combination which is necessary in order to achieve both higher surface chlorophyll concentration and rapid decrease at depth. Similarly, the weighted average profiles that appear sigmoid in shape are almost always best represented by either a pure Sigmoid or a Sigmoid-Gaussian combination. The overlap in the best choice for curve fit for these two broad subgroups that, at least visually, appear extremely different, suggests that when taking a global approach to describing profiles mathematically, some caution is warranted in translating directly between the optimized curve fit and the expected profile shape and any associated physical mechanisms. Furthermore, the distribution across different curve fits remains similar if we separate out the Southern Ocean, so the breakdown is not the result of including multiple ocean regimes in one cluster. Finally, if we

separate out each cluster into daytime and nighttime profiles, the breakdown between different optimal curve fits within each group remains roughly the same, both for a global ocean assessment and the case for the Southern Ocean only. There is a slight increase in pure Gaussian fits for daytime profiles in clusters with a subsurface chlorophyll maximum (deep or shallow), which may be consistent with depressed daytime surface fluorescence resulting from NPQ.

This also provides some insight into interpreting different methods for characterizing chlorophyll profiles. With the clustering approach, profiles are classified with other profiles with which they have the most overlap in terms of general features (i.e. the profiles with the most similar principal component weights). As such, profiles with deep chlorophyll maxima at the same approximate depth are likely to be grouped together, even if their surface chlorophyll concentrations are different enough that they would be best characterized mathematically in different ways or would land in different groups if surface chlorophyll were used to delineate groups. Similarly, because a linear combination of mathematical shapes can have the same components but dramatically different weighting for each component, attributing a set of profile shapes to a particular combination of mathematical forms may not be as effective for a global dataset as it is in a regional study, such as in Carranza et al (2018).

Finally, a longstanding issue in studies of phytoplankton and chlorophyll distribution in the water column, and hence biomass, is that of relating surface chlorophyll concentration to a particular subsurface distribution. Uitz et al (2006) explicitly related their individual profile shapes to a set of non-overlapping surface chlorophyll values, and yet as we see here, profiles with distinct subsurface features often overlap in their surface chlorophyll ranges. Even considering the realistic possibility that the absolute chlorophyll concentrations reported by the Argo floats may require additional fine tuning (Roesler et al, 2017), clusters that occur within the

same general region, such as the well-mixed profiles in the Southern Ocean, show considerable overlap. In these cases, it is likely that the base of the well-mixed layer is set by the physical processes that produce a particular mixed layer depth, and that many different chlorophyll concentrations can exist within that. As a result, we think that there is considerable merit to an approach such as this one that classifies profiles based on their shape characteristics rather than their surface concentration, but recognize that the challenge remains of discerning subsurface characteristics from surface values. As the coverage of floats increases and the absolute chlorophyll concentrations become better established and calibrated with satellite values, we hope to provide similar tools for assessing subsurface profile characteristics from above.

4.6.2 Comparison with previous studies

A challenge with any machine learning approach is the question of whether it returns physically interpretable results from which we can genuinely learn something: the results presented here suggest that this analysis can provide a useful way to distinguish both between broad categories of behavior and more nuanced shifts in dominant physical forcing mechanisms (upper ocean mixing, light availability) and possible bioegeochemical influences (iron fertilization, nutrient availability), particularly when the results can be paired with additional biogeochemical and physical data gathered by the floats. **Figure 4.7** in particular accentuates global patterns where similar profile shapes are found. Although this study focused on unsupervised clustering, it also produces some similar patterns in dominant features to the supervised classification and grouping performed by Cornec et al (2021) in their study characterizing deep fluorescence maxima (DFM) throughout the global ocean. For example, at high latitudes, deep chlorophyll maxima comprise a relatively small fraction of the profiles,

whereas at low latitudes they are the dominant feature, although this study identified a larger percentage of SCM-type profiles in the Southern Ocean during the summer months. Furthermore, their clustering based on characteristics of the DFM identifies the same separation in dominant clusters that we find here: the profiles with subsurface chlorophyll maxima in the tropics and subtropics exhibit a different set of profile characteristics and also relationship with the physical profile. Profiles with deep, relatively small, chlorophyll maxima dominate in subtropical regions, and profiles with shallower subsurface chlorophyll maxima, closer to the depth of the actual mixed layer, dominate in higher latitude regions. This is consistent both with Cornec et al (2021) and previous work identifying the dominant mechanisms in forming SCMs at different latitudes (Cullen, 2015; Baldry et al, 2020). This is a promising result for future efforts that might attempt to delineate regions that likely have different influences on the vertical chlorophyll structure using unsupervised methods.

In the work presented here, the seasonal evolution in profile characteristics results in a transition between clusters with distinct characteristics, whereas Cornec et al (2021) looked at largescale regions that had broadly similar characteristics and then tracked the seasonal evolution within those groups. For example, in subtropical regions, they identified a deep acclimation zone, (DAZ) and found shifts in mixed layer depth, surface chlorophyll, and other properties influencing the structure of the vertical chlorophyll profile. Although direct comparison is somewhat complicated, we can see that their results showed that profiles in the DAZ had the shallowest mixed layers and lowest satellite chlorophyll during the summer months (their Figure 13); we find, similarly, that the summer months in the subtropics are dominated by clusters with the lowest surface chlorophyll and relatively shallow mixed layers, and that there is a significant shift in fall and winter (**Figure 4.7** and **Table 4-1**). One focus of their work that we do not

address here is whether a deep fluorescence maximum is due to photoacclimation at the surface or an actual biomass maximum. In looking at the geographic locations of their four main groups and the mechanisms they identify, it is likely that some of the clusters here lump together profiles that in their work are separated out into the deep acclimation zone (DAZ) and deep biomass zone (DBZ) (their Figure 8). Each of these approaches is valuable and depends on the future applications envisioned for the work: quantification of biomass vs incorporation of actual chlorophyll characteristics into models, for example.

Baldry et al (2020) performed a detailed review of the formation mechanisms and observed locations of SCM and diatom SCM (dSCM) in the Southern Ocean. **Figure 4.9** highlighted regions in the Southern Ocean that have a year-round high concentration of profiles with shallow subsurface chlorophyll maximum, and these align well with regions that experience iron fertilization by land masses (Baldry et al, 2020; Ardyna et al, 2016). This is certainly not conclusive supporting evidence, but the consistency between this distribution and the results of previous studies is promising for future applications of this approach, particularly when paired the geographic separation between clusters with distinct characteristics noted earlier.

The delineation between groups complements previous studies that have used surface chlorophyll only to explore biogeography and phytoplankton phenology. For example, we see three clusters that represent most of the profiles in the northern Indian Ocean (clusters 3, 6, and 7) with some seasonal patterns as to which cluster is most prevalent (Figure S5). Initial studies of global biogeography indicated that much of the region constituted a single biogeographical province (“MONS”, Longhurst, 2007), but subsequent work showed there are significant seasonal shifts, often associated the monsoon upwelling along the borders of the Arabian Sea and riverine outflow in the northern Bay of Bengal (Reygondeau et al, 2013), which shift the nature

of the region from oligotrophic to eutrophic in certain areas. The coverage of this region by BGC Argo floats is not as comprehensive as that provided by satellite data, although not prone to some of the surface visibility issues associated with cloud cover during the southwest Monsoon, and therefore precise comparison of results is difficult. However, we also see seasonal shifts in the most prevalent clusters in the Arabian Sea, and less so in the equatorial region and the Bay of Bengal. For example, the prominence of the highest surface chlorophyll groups in the Arabian Sea during winter match the expansion of the "ARAB" biogeographical province during those months (Reygondeau et al, 2013). The Bay of Bengal is dominated by the same clusters year round, but is not well covered in the far norther where a separate biogeographical province shows more seasonal variation. Similar arguments can be applied to the Southern Ocean, where Reygondeau et al (2013) identified year round biogeographical provinces downstream of the Drake Passage and New Zealand (similar to the patterns shown in Figure 9), with seasonally shifting biogeographical provinces at various distances from the Antarctic coast. Furthermore, unlike studies of biogeography and seasonal evolution of phytoplankton blooms in the Indian Ocean (Lévy et al, 2007), the patterns highlighted here also include information about the vertical distribution of chlorophyll. As the dataset expands, this approach may prove useful in refining our understanding of the seasonal shifts not only in surface chlorophyll but also in distribution and allow us to explore interannual and long-term trends, particularly if coupled with physical or other biogeochemical variables.

4.6.3 Limitations of the study

As noted several times already, Argo biogeochemical floats are unevenly distributed, and therefore any statistics about the overall prevalence of a particular profile type should only be

considered for region-specific studies where there is enough data that we are likely to be resolving seasonal variability across the whole region (i.e. the Southern Ocean). Many of the patterns identified here are consistent with previous work, both in terms of profile shape (Uitz et al 2006; Mignot et al, 2011; Carranza et al, 2018; Prasanth et al, 2021; Gu et al, 2020) and seasonal variability (Carranza et al, 2018; Prasanth et al, 2021; Gu et al 2020), but it is not yet clear that we are adequately capturing the full spectrum of possible behaviors or their relative occurrence. For example, we currently have no clusters with a characteristic secondary chlorophyll maximum, a feature observed in oxygen minimum zones (Tiano et al, 2014) and the Black Sea (Churilova et al, 2019). In the Eastern Tropical North Pacific, for example, we observe secondary chlorophyll maxima in every profile, but the profiles are classified in a range of different groups exhibiting deep and shallow primary chlorophyll maximum. These profiles appear to be categorized according to the characteristics of this primary maximum, with the largest number of profiles assigned to Group 6 (see **Figure 4.4**). We likely presently lack adequate data in regions with those features. As the dataset evolves, we might expect both the range of optimal clusters and their precise features to adjust somewhat. However, despite these limitations, the current posterior probabilities suggest that many of the available profiles are very well described by the current clusters.

A second limitation is that this study only addresses open ocean regimes, since profiling floats rarely enter coastal waters where the depth is less than 2000m. In such coastal regions we would likely expect much higher overall surface chlorophyll concentrations, and possibly a departure from Type 1 water (Jerlov, 1977). The west coast of South America near Peru, the coast of the Somali and Arabian Peninsulas, and the Eastern Coast of Australia have received some coverage, but this constitutes a very small overall fraction of the dataset. Because this

particular dataset is ever unlikely to achieve excellent coverage of coastal regions, a more comprehensive study of those regions would require pairing this with a different dataset.

Finally, given the regional variability of the relationship between calibrated chlorophyll values and the actual chlorophyll concentration (Roesler et al, 2017), it is beyond the scope of this study to propose explicit relationships between a surface chlorophyll concentration and the appropriate profile shape as other studies have done. While we find strong patterns in seasonal evolution of profile shape and characteristics, many of the profile groups span a wide range of surface chlorophyll concentrations, even when only looking at a single region (e.g. the Southern Ocean), further complicating any efforts to tie surface chlorophyll concentration to profile shape. As a result, it is believed that the results of this study are best suited to providing insight into regional patterns in vertical chlorophyll behavior, while not attempting to extrapolate to a vertical profile shape that could be related to values derive from satellites without further work to refine the Argo calibration scheme.

4.6.4 Future work

Satellite data has been used very successfully to define biogeographical regimes (Reygondeau et al, 2013; Hardman-Mountford et al 2008; Ardyna et al, 2016) and characterize their seasonal evolution, but studies rarely include data about the vertical distribution of chlorophyll as well. Rosso et al (2020) showed the potential for applying a similar machine learning approach as we have used here to a set of multiple hydrographic variables (temperature and salinity). Given the success here of characterizing the seasonal evolution of broadly classified chlorophyll regimes, it is suggested that expanding this work to include other variables such as temperature, salinity, and oxygen, could provide valuable insight into more fully

characterizing biogeographical regimes. It would be of particular interest to find whether such a study including vertical data aligned well with existing regional (Ardyna et al, 2016) and global biogeographical studies (Reygondeau et al, 2013).

Finally, the set of characteristic vertical chlorophyll profiles outlined here provides a valuable tool for exploring the impact of chlorophyll distribution on upper ocean heating and expanding on previous work studying self-shading (e.g. Manizza et al, 2005). These well-defined profiles, their standard deviations, and their observed seasonal variability), provide a pathway to quantifying the magnitude of impact associated with different profile characteristics in combination with different physical scenarios. This work is currently underway.

4.7 ACKNOWLEDGMENTS

Many thanks to the numerous scientists who provided feedback on this work as it was being developed, including the members of the Riser/Gray research group, Kimberlee Baldry, and Isa Rosso. The Argo data used to create this climatology were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. doi.org/10.17882/42182#56126.

References

Ardyna, M., Claustre, H., Sallée, J.-B., D'Ovidio, F., Gentili, B., van Dijken, G., D'Ortenzio, F. & Arrigo, K. R. (2017). Delineating environmental control of phytoplankton biomass and phenology in the Southern Ocean, Geophysical Research Letters, 44, 5016–5024, doi:10.1002/2016GL072428.

Argo (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE. https://doi.org/10.17882/42182 (data downloaded 1/1/2021)

Arrigo, K. R., Matrai, P. A., & van Dijken, G. L. (2011). Primary productivity in the Arctic Ocean: Impacts of complex optical properties and subsurface chlorophyll maxima on large-scale estimates, Journal of Geophysical Research, 116, C11022, doi:10.1029/2011JC007273.

Arteaga, L., Boss, E., Behrenfeld, M., Westberry, T., Sarmiento, J (2020). Ecological drivers of phytoplankton bloom cycles in the Southern Ocean. Earth and Space Science Open Archive, doi: 10.1002/essoar.10501279.1

Baldry, K. (ed.) (2021) Biogeochemical Argo Cheat Sheets: Data distribution; Quality control and GDAC; Chlorophyll-a; Optical backscatter; pH; Irradiance; Oxygen; Nitrate. Hobart, Tasmania, Institute of Marine and Antarctic Studies, 8pp. doi: 10.25607/OBP-981.

Baldry K., Strutton, P.G., Hill, N.A. & Boyd, P.W. (2020). Subsurface Chlorophyll-a Maxima in the Southern Ocean. Frontiers in Marine Science, 7:671, doi: 10.3389/fmars.2020.00671

Carranza, M. M., Gille, S. T., Franks, P. J., Johnson, K. S., Pinkel, R., & Girton, J. B. (2018). When mixed layers are not mixed. Storm-driven mixing and bio-optical vertical gradients in mixed layers of the Southern Ocean. Journal of Geophysical Research: Oceans, 123, 7264–7289. https://doi.org/10.1029/2018JC014416.

Churilova, T., Suslin, V., Sosik, H.M., Efimova, T., Moiseeva, N., Moncheva, S., Mukhanov, V., Rylkova, O.& Krivenko, O. (2019). Phytoplankton light absorption in the deep chlorophyll maximum layer of the Black Sea, European Journal of Remote Sensing, 52:sup1, 123-136, doi:10.1080/22797254.2018.1533389

Claustre, H., Johnson, K, & Takeshita, Y. (2020). Observing the Global Ocean with Biogeochemical Argo. Annual Review of Marine Science, 12: 23-48.

Cornec, M., Claustre, H., Mignot, A., Guidi, L., Lacour, L., Poteau, A., et al. (2021). Deep chlorophyll maxima in the global ocean: Occurrences, drivers and characteristics. Global Biogeochemical Cycles, 35, e2020GB006759. https://doi.org/10.1029/2020GB006759

Cullen, J. J. (2015). Subsurface chlorophyll maximum layers: enduring enigma or mystery solved? Annual Review of Marine Science, 7:207-239.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of Geophysical Research, 109, C12003, doi:10.1029/2004JC002378.

Gnanadesikan, A. & Anderson, W.G. (2009). Ocean Water Clarity and the Ocean General Circulation in a Coupled Climate Model. Journal of Physical Oceanography, 39, 314-332.

Gordon, H.R. & McCluney, W.R. (1975). Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing. Applied Optics, 14, 413-416.

Gu, Y., Cheng, X., Qi, Y., & Wang, G. (2020). Characterizing the seasonality of vertical chlorophyll-a profiles in the Southwest Indian Ocean from the Bio-Argo floats. Journal of Marine Systems, 212.

Hardman-Mountford, N.J., Hirata, T., Richardson, K.A., & Aiken, J. (2008). An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. Remote Sensing of Environment, 112: 3341-3352.

Houghton, I. A., & Wilson, J. D. (2020). El Niño detection via unsupervised clustering of Argo temperature profiles. Journal of Geophysical Research: Oceans, 125, e2019JC015947. doi:10.1029/2019JC015947.

Jerlov, N.G. (1977). Classification of sea water in terms of quanta irradiance. *ICES Journal of Marine Science,* 37(3), 281-287. https://doi.org/10.1093/icesjms/37.3.281Johnson et al, 2017

Johnson, K. S., et al. (2017), Biogeochemical sensor performance in the SOCCOM profiling float array, Journal of Geophysical Research: Oceans, 122, 6416–6436, doi:10.1002/2017JC012838.

Jones, D.C., Holt, H. J., Meijers, A. J. S., & Shuckburgh, E. (2019). Unsupervised clustering of Southern Ocean Argo float temperature profiles. Journal of Geophysical Research: Oceans, 124, 390–402. https://doi.org/10.1029/2018JC014629

Lévy, M., Shankar, D., André, J.-M., Shenoi, S. S. C., Durand, F., & de Boyer Montegut, C. (2007). Basin-wide seasonal evolution of the Indian Ocean's phytoplankton blooms. Journal of Geophysical Research, *112*, C12014, doi:10.1029/2007JC004090.

Lim, H.-G., Park, J.Y., & Kug, J.-S. (2018). Impact of chlorophyll bias on the tropical Pacific mean climate in an earth system model. Climate Dynamics, 51:2681-2694.

Manizza, M., Le Quéré,C., Watson, A.J., & Buitenhuis, E.T. (2005). Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. Geophysical Research Letters, 32, L05603, doi:10.1029/2004GL020778.

Maze, G., Mercier, H., Fablet, R., Tandeo, P., Radcenco, M.L., Lenca, P., Feucher, C., & Le Goff, C. (2017). Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the Northern Atlantic Ocean. Progress in Oceanography, 151, 275-292.

Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., & Ras, J. (2011). From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-a concentration. Biogeosciences, 8, 2391-2406.

Morel, A. (1988) Optical Modeling of the Upper Ocean in Relation to Its Biogenous Matter Content (Case I Waters). Journal of Geophysical Research, 93, C9, 10,749-10,768.

Morel, A. & Antoine, D. (1993). Heating Rate within the Upper Ocean in Relation to its Bio-Optical State. Journal of Physical Oceanography, 24, 1652-1665.

Morel, A. & Berthon, J.-F. (1989). Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. Limnology and Oceanography, 34(8), 1545-1562.

Ohlmann, J.C., Siegel, D.A., & Gautier, C. (1996). Ocean Mixed Layer Radiant Heating and Solar Penetration: A Global Analysis. Journal of Climate, 9, 2265-2280.

Organelli, E., Claustre, H., Bricaud, A., Barbieux, M., Uitz, J., D'Ortenzio, F., & Dall'Olmo, G. (2017). Bio-optical anomalies in the world's oceans: An investigation on the diffuse attenuation coefficients for downward irradiance derived from Biogeochemical Argo float measurements, Journal of Geophysical Research: Oceans, 122, 3543–3564, doi:10.1002/ 2016JC012629.

Park, J.-Y., Kug, J.-S., Bader, J., Rolph, R., & Kwon, M. (2015). Amplified Arctic warming by phytoplankton under greenhouse warming. PNAS, 112(19), 5921-5926.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Journal of Machine Learning Research, 12:2825-2830

Prasanth, R., Vijith, V., Thushara, V., George, J.V., & Vinayachandran, P.N. (2021). Processes governing the seasonality of vertical chlorophyll-a distribution in the central Arabian Sea: Bio-Argo observations and ecosystem model simulation. Deep-Sea Research Part II, 183, doi:10.1016/j.dsr2.2021.104926

Ravichandran, M., Girishkumar, M.S., & Riser, S. (2012). Observed variability of chlorophyll-*a* using Argo profiling floats in the southeastern Arabian Sea. Deep-Sea Research 1, 65, 15-25.

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013) Dynamic biogeochemical provinces in the global ocean. Global Biogeochemical Cycles, 27: 1046-1058, doi:10.1002/gbc.20089.

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmectig, C., Poteau, A., D'Ortenzio, F., Ras, J., Drapeau, S., Haëntjens, N., Barbieux, M. (2017). Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs Eco sensors. Limnology and Oceanography: Methods, 15, 572-585.

Rosso, I., Mazloff, M. R., Talley, L. D., Purkey, S. G., Freeman, N. M., & Maze, G. (2020). Water mass and biogeochemical variability in the Kerguelen sector of the Southern Ocean: A machine learning approach for a mixing hot spot. Journal of Geophysical Research: Oceans, 125, e2019JC015877, doi:10.1029/2019JC015877

Sathyendranath, S. & Platt, T. (1988). The Spectral Irradiance Field at the Surface and in the Interior of the Ocean: A Model for Applications in Oceanography and Remote Sensing. Journal of Geophysical Research, 93, C8, 9270-9280.

Schmechtig, C., Claustre, H., Poteau, A., D'Ortenzio F. (2018). Bio-Argo quality control manual for the Chlorophyll-A concentration. https://doi.org/10.13155/35385

Schott, F.A., & McCreary, J.P. (2001) The monsoon circulation of the Indian Ocean. Progress in Oceanography 51, 1–123.

Smith, R.C. (1981) Remote Sensing and Depth Distribution of Ocean Chlorophyll. Marine Ecology Progress Series, 5:3, 359-361.

Thomalla , S.J., Moutier, W., Ryan-Keogh, T.J., Gregor, L., & Schütt, J. (2018). An optimized method for correcting fluorescence quenching using optical backscatter on autonomous platforms. Limnology and Oceanography: Methods, 16, 132-144.

Thushara, V., Vinayachandra, P.N.M., Matthews, A.J., Webber, B.G.M., & Queste, B.Y. (2019). Vertical distribution of chlorophyll in dynamically distinct regions of the southern Bay of Bengal. Biogeosciences, 16, 1447-1468.

Tiano, L., Garcia-Robledo, E., Dalsgaard, T., Devol, A.H., Ward, B.B., Ulloa, O., Canfield, D.E., & Revsbech, N.P. (2014) Oxygen distribution and aerobic respiration in the north and south eastern tropical Pacific oxygen minimum zones. Deep-Sea Research 1, 94, 173-183.

Uitz , J., Claustre, H., Morel, A., & Hooker, S.B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. Journal of Geophysical Research, 111, C08005, doi:10.1029/2005JC003207.

Wang, B., Fennel, K., Yu, L., & Gordon, C. (2020). Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. Biogeosciences, 17, 4059-4074. Doi: 10.5194/bg-17-4059-2020

Xing, X., Claustre, H., Blain, S., D'Ortenzio, F., Antoine, D., Ras, J., & Guinet, C. (2012). Quenching correction for in vivo chlorophyll fluorescence acquired by autonomous platforms: A case study with instrumented elephant seals in the Kerguelen region (Southern Ocean). Limnology and Oceanography: Methods, 10, 483-495.

Xing, X, Claustre, H., Boss, E., Roesler, C., Organelli, E., Poteau, A., Barbieux, M., & D'Ortenzio, F. (2016). Correction of profiles of in-situ chlorophyll fluorometry for the contribution of fluorescence originating from non-algal matter. Limnology and Oceanography: Methods, 15, 80-93.

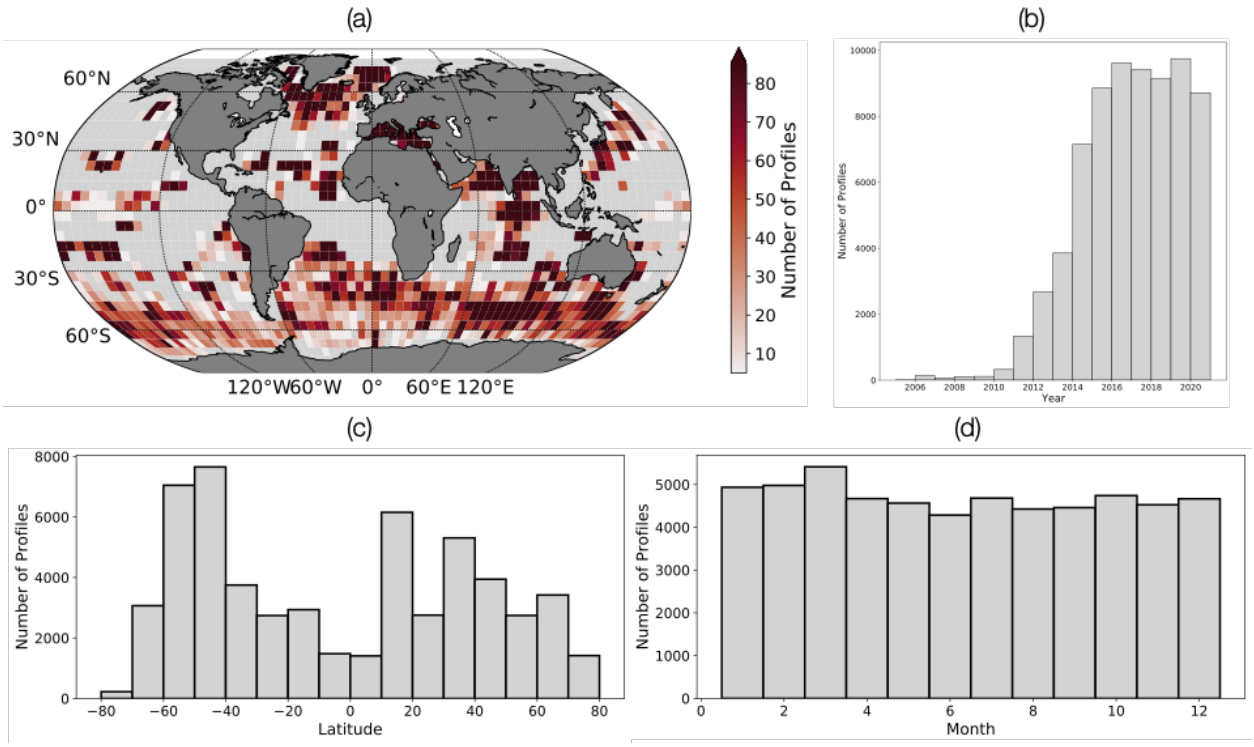

Figure 4.1. (a) Concentration of chlorophyll fluorescence profiles collected by biogeochemical Argo floats. Each box is 5x5 degrees. Gray shading indicates no (useable) profiles were gathered in that region. (b) Distribution of profiles by year. (c) Distribution of profiles by latitude. (d) Number of profiles gathered in each month.

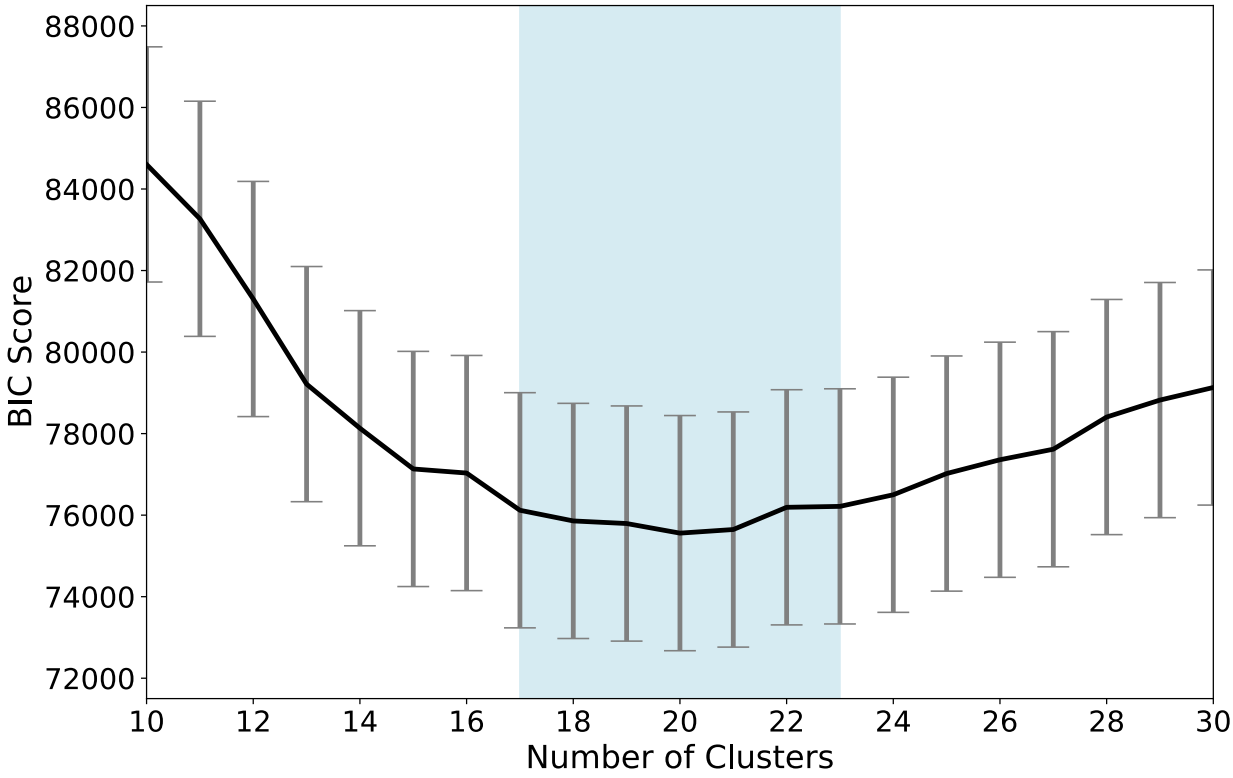

Figure 4.2. Bayesian Information Criterion (BIC) score calculated using 50 randomized subgroups. Blue shaded region highlights broad minimum of values considered for number of clusters. Gray bars are standard deviation.

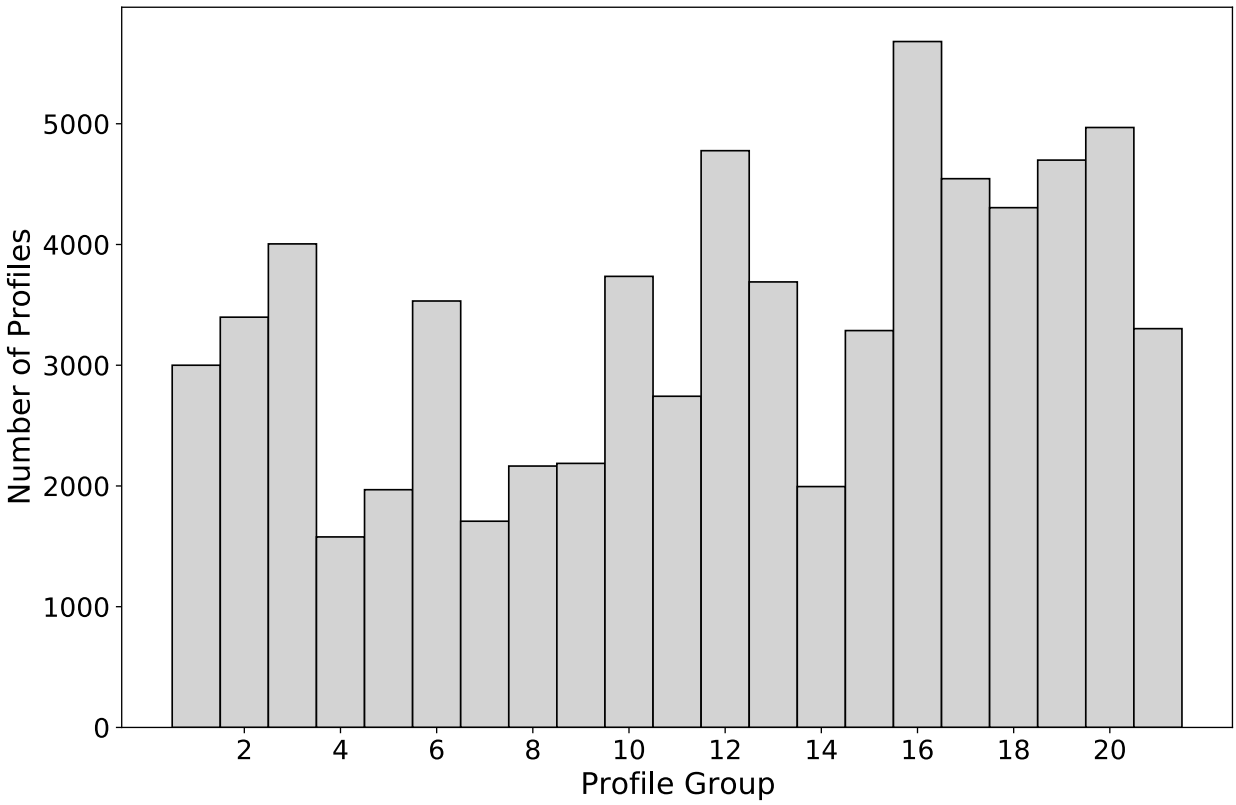

Figure 4.3. Number of profiles in each of the 21 clusters used in this study. Details of the exact number of profiles are also shown in Table 1.

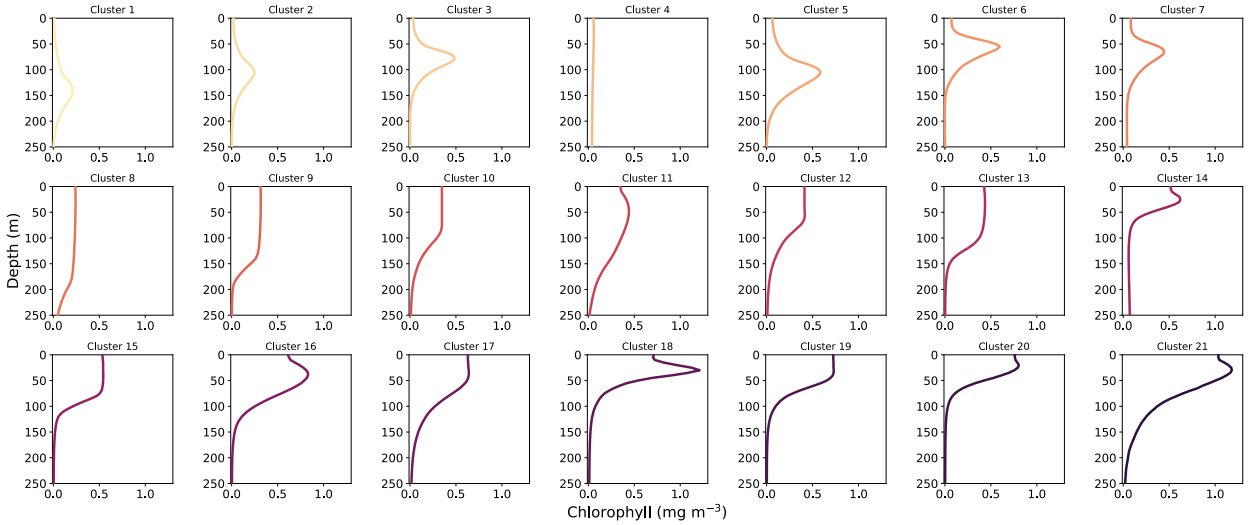

Figure 4.4. Weighted average profiles for each of the 21 clusters. Profiles are sorted from top left to bottom right by the surface chlorophyll concentration of the weighted average profile. Colors indicate sorting of profiles from lowest surface chlorophyll (light yellow) to highest surface chlorophyll (dark purple). Same color scheme is used in Figures 5-8 for ease of comparison.

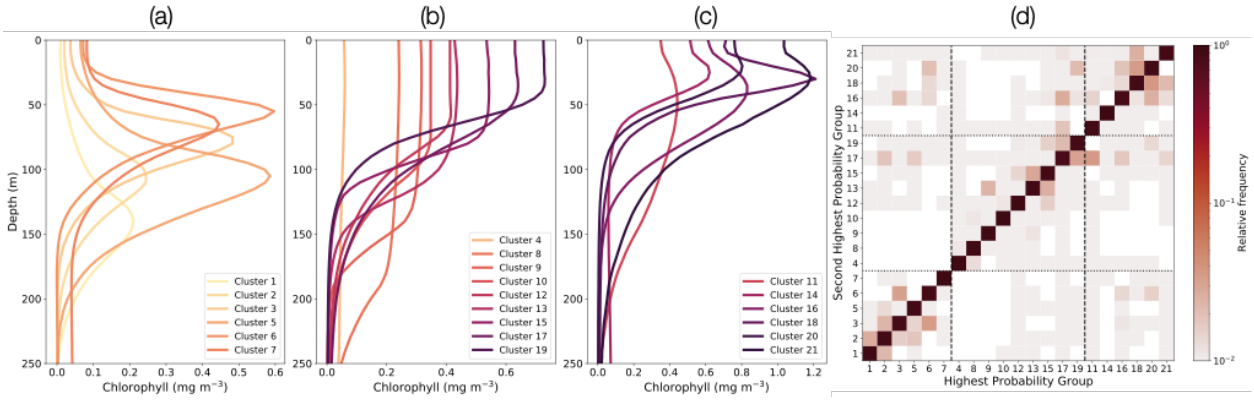

Figure 4.5. (a)-(c) weighted average profiles assembled in subgroups with generally similar profile characteristics: (a) deep subsurface chlorophyll maximum; (b) well-mixed; and (c) shallow sub surface chlorophyll maximum. (d) Highest posterior probability plotted vs. second highest posterior probability. Dotted/dashed lines separate major subgroups shown in (a)-(c). Boxes around center line indicate profiles for which the highest and second highest posterior probability fall within the same major subgroup. Diagonal values (all relative frequency = 1) are not meaningful in this context.

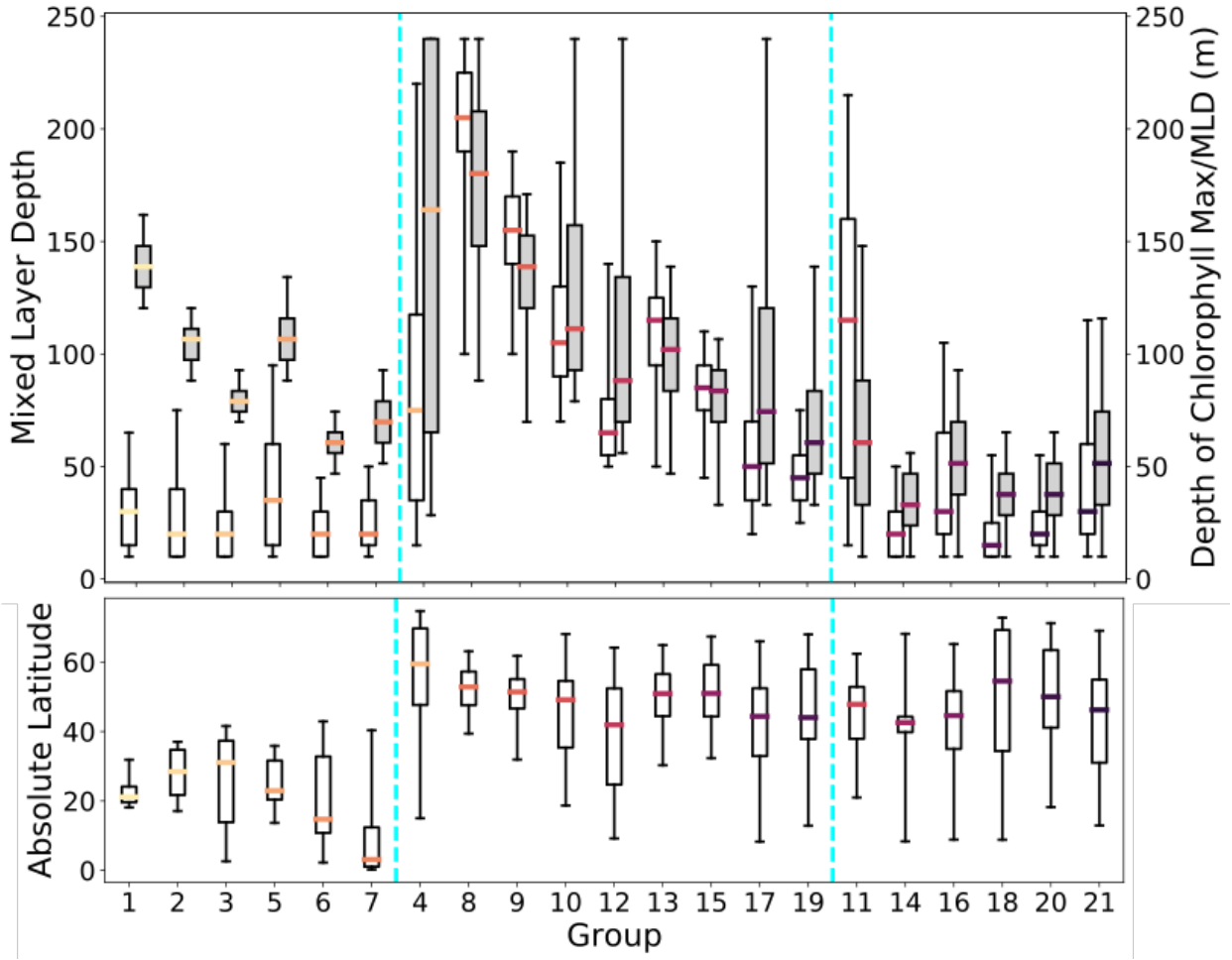

Figure 4.6. (a) Distribution of mixed layer depths (white fill) and chlorophyll maximum (gray fill) for each cluster. Vertical dashed lines mark the separation between each subgroup (from L to R: deep SCM; well-mixed; shallow SCM). For deep and shallow SCM clusters (L and R), gray fill boxes indicate depth of maximum chlorophyll; for well-mixed clusters (center), gray fill boxes mark approximate base of the well-mixed chlorophyll layer. (b) Absolute latitude distribution for each cluster.

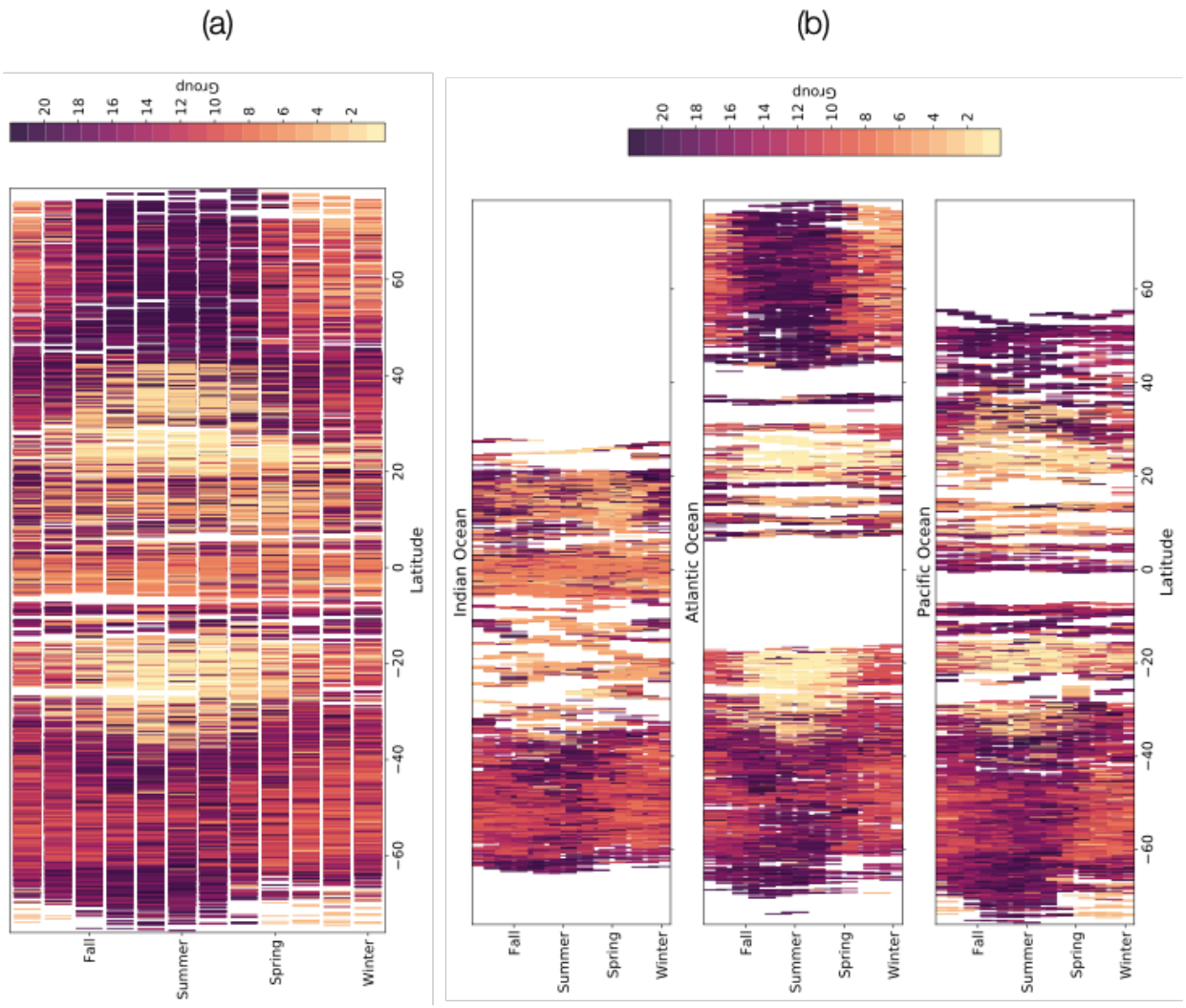

Figure 4.7. Latitude vs. season plot showing distribution of clusters derived from Gaussian Mixture Model for (a) the entire ocean and (b) individual basins.

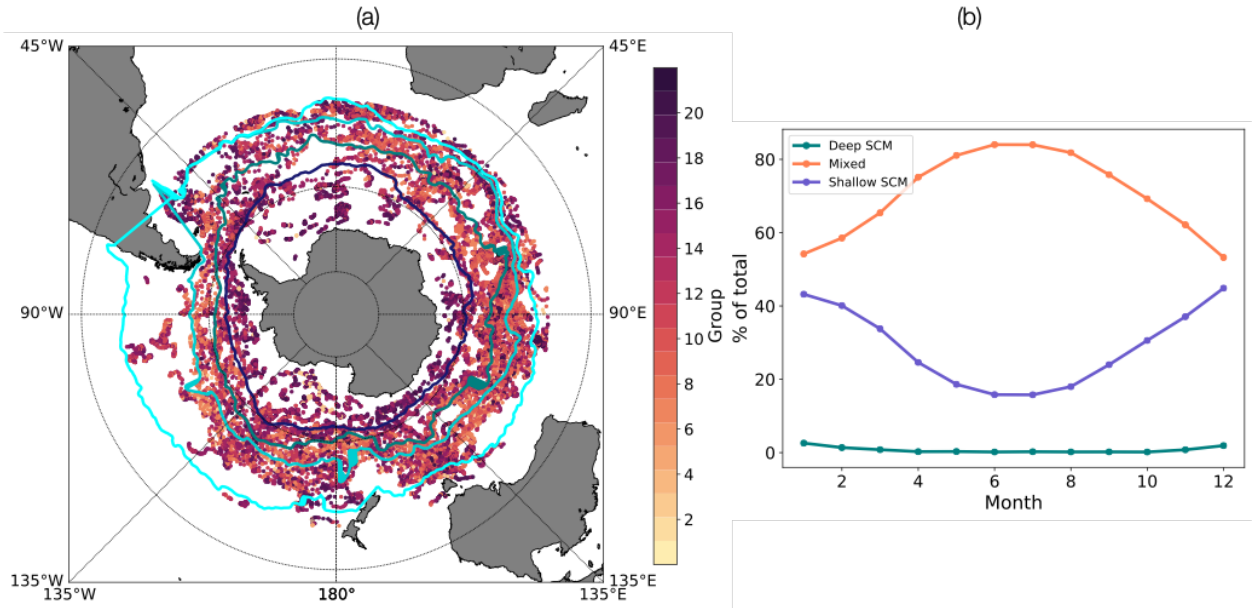

Figure 4.8. (a) Plot of Southern Ocean (latitude <40S) clusters, with locations of commonly observed fronts. (b) Percentage of profiles associated with each broad subgroup by month.

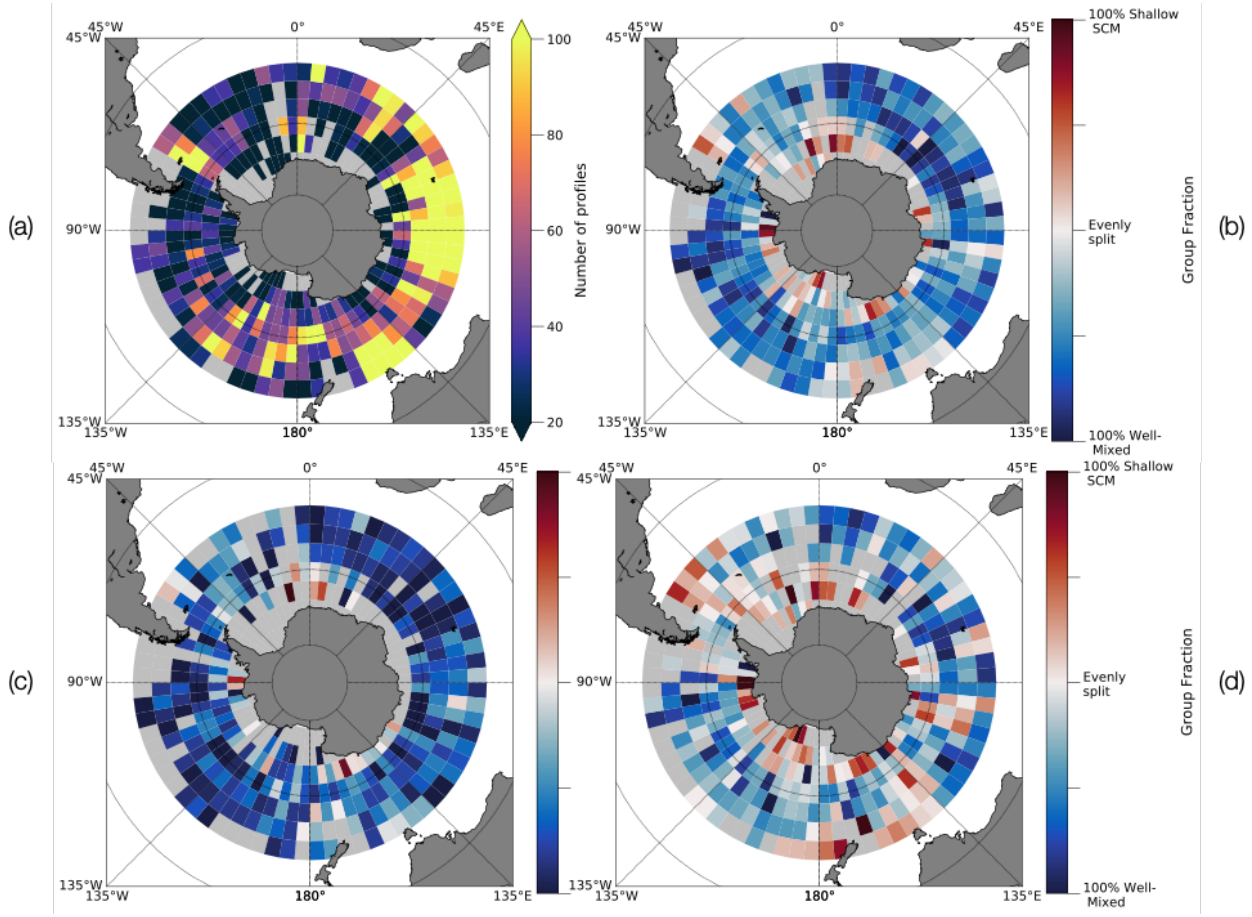

Figure 4.9. (a) Number of observations in each 5x5 box. (b) Percentage of profiles in each of the two major Southern Ocean subgroups (shallow SCM and well-mixed), annual average. (c) Same as (b), but "winter" months only. (d) Same as (b) but "summer" months only.

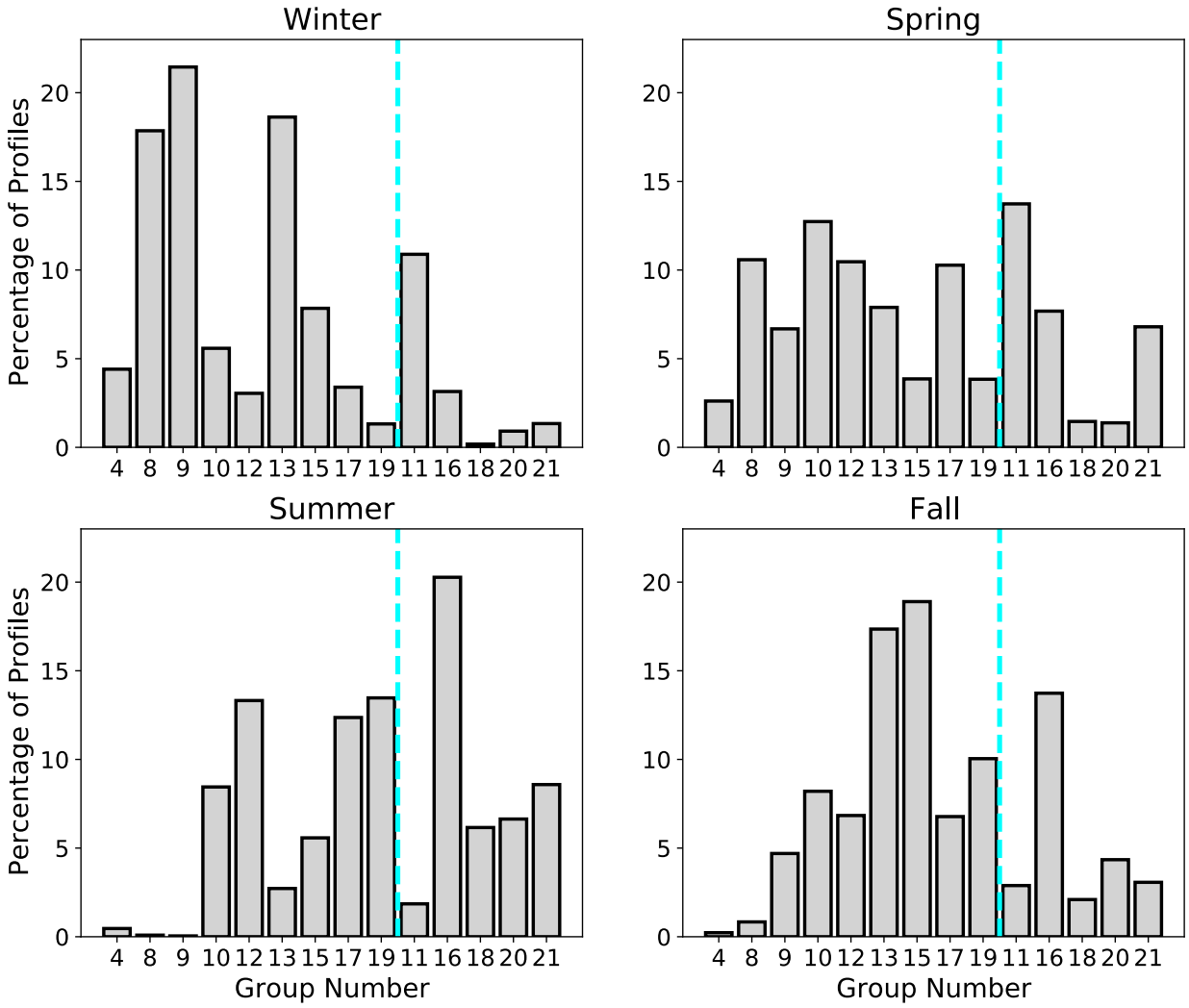

Figure 4.10. Histogram (by season) of percentage of profiles in clusters falling within each of the two major subgroups (shallow SCM and well-mixed). Clusters to the left of the dashed line are part of the well-mixed subgroup; clusters to the right are part of the shallow SCM subgroup.

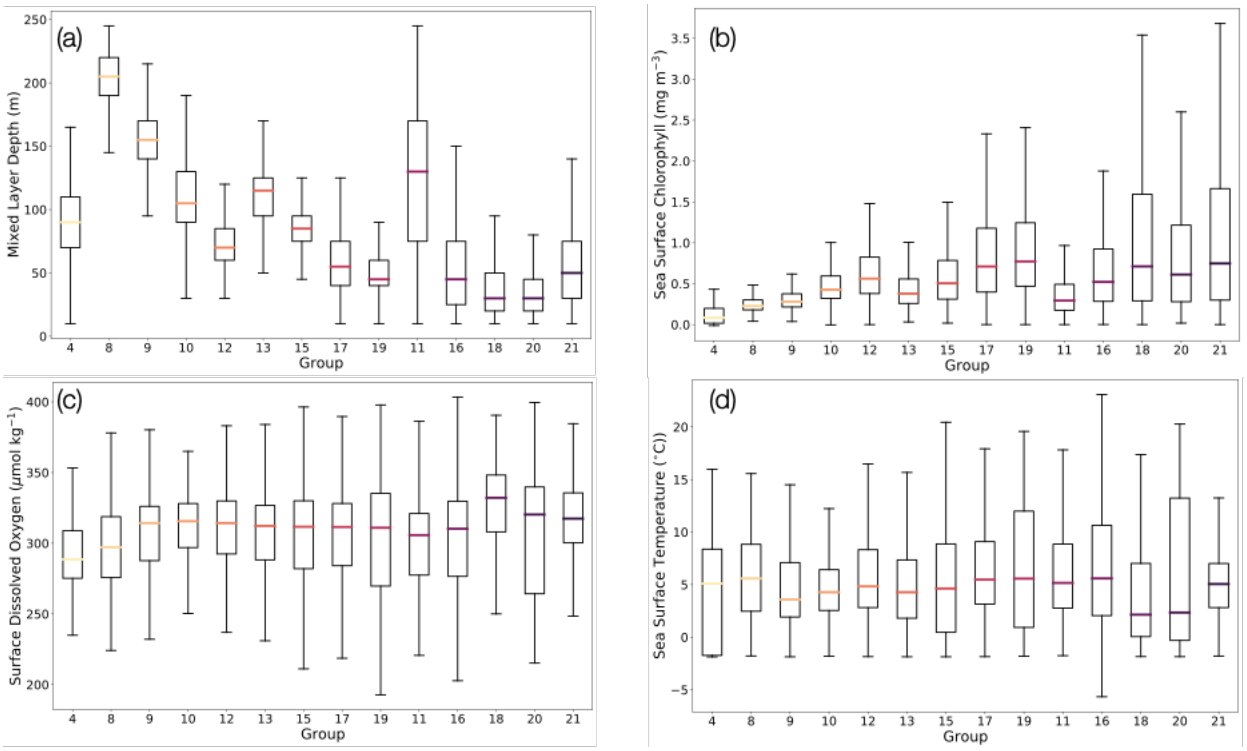

Figure 4.11. Distribution of physical and biogeochemical characteristics by cluster for Southern Ocean profiles: (a) Mixed layer depth (m); (b) sea surface chlorophyll (mg m^{-3}); (c) dissolved oxygen at the surface (µ mol kg^{-1}); (d) sea surface temperature (°C).

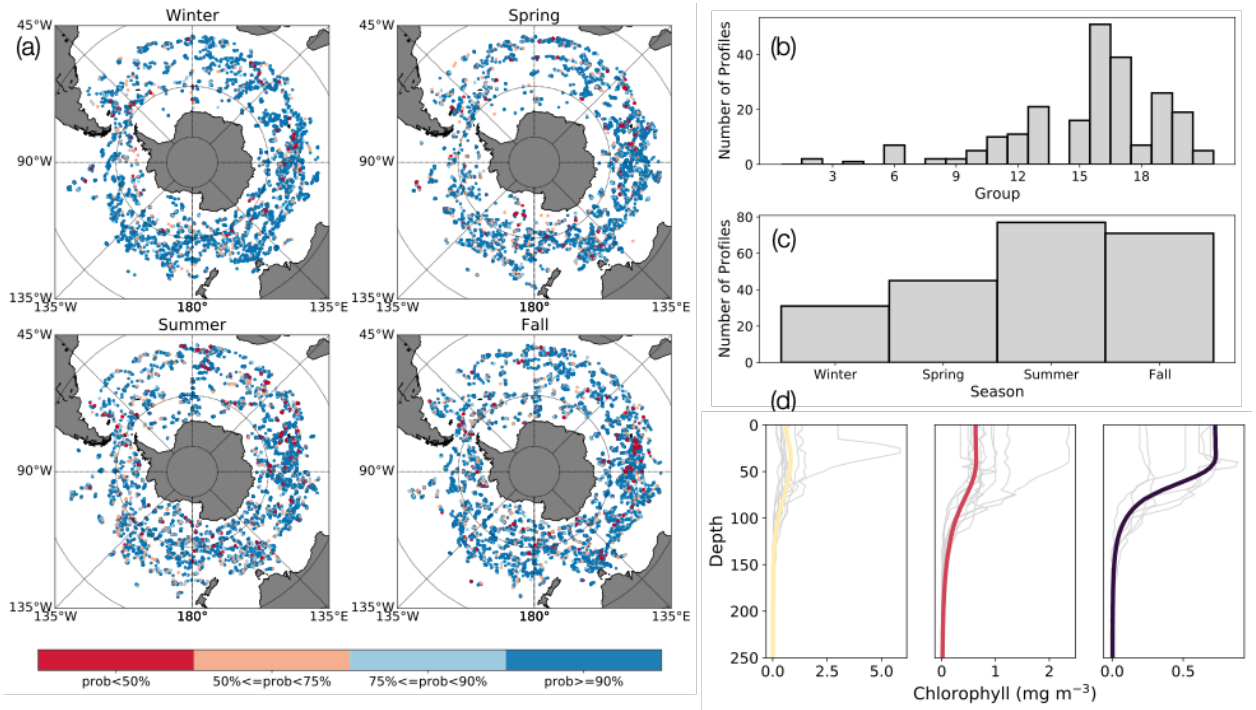

Figure 4.12. (a) Seasonal plot of posterior probabilities for each profile. (b) Number of profiles with probability <50% by cluster number. (c) Seasonal distribution of profiles with probability <50%. (d) Example profiles (light gray lines) from Cluster 16, 17, and 19 with weighted average profile in heavy line.

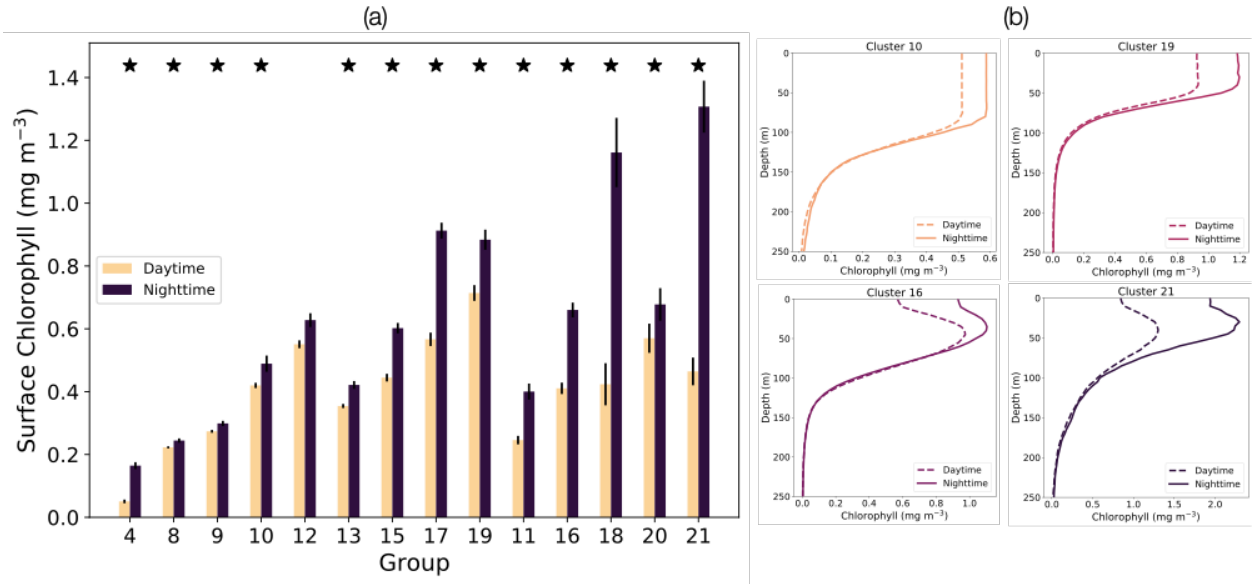

Figure 4.13. Comparison of daytime and nighttime behavior by cluster for the Southern Ocean. (a) Daytime vs. nighttime median sea surface concentrations. Error bars are standard error. Stars indicated clusters for which the difference between daytime and nighttime surface chlorophyll values is statistically significant. (b) Daytime vs nighttime weighted average profiles for four Southern Ocean cluster (10 and 19 from the well-mixed subgroup and 16 and 21 from the shallow SCM subgroup).

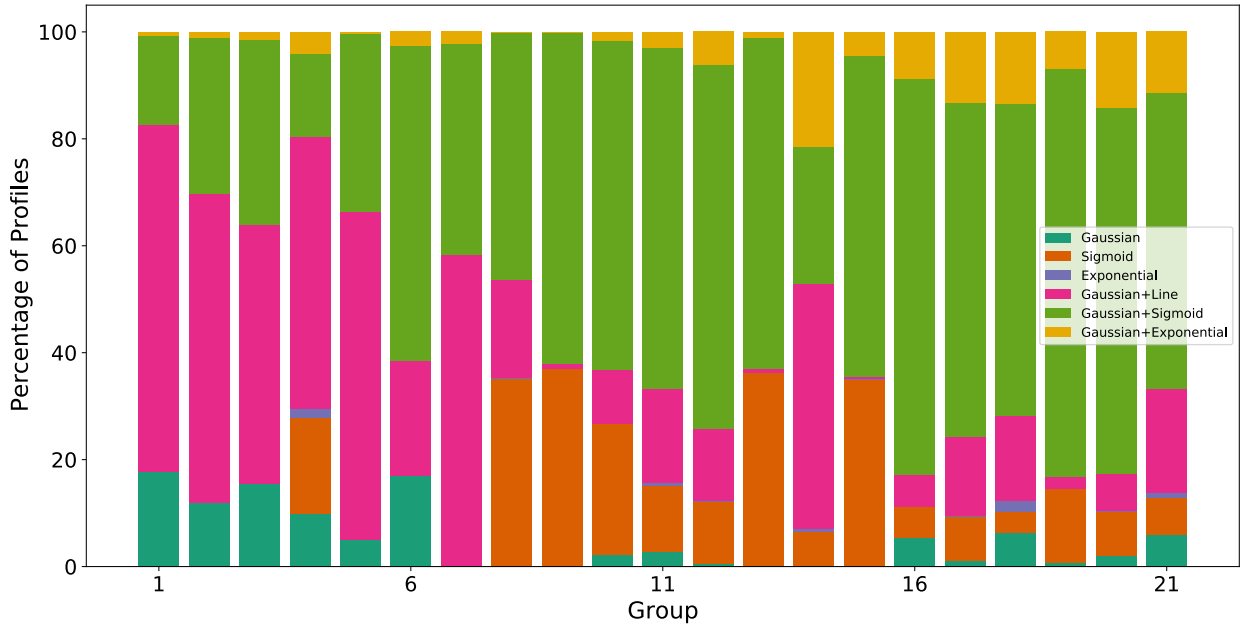

Figure 4.14. Best curve fit breakdown for profiles in each cluster, plotted as percentage of each group.

		Percentage of Profiles with Posterior Probabilities				Surface Chlorophyll Values				Percentage of Profiles by Season			
Cluster	Number of Profiles	0-0.5	0.5-0.75	0.75-0.9	0.9-1	Median	STD	Min	Max	Winter	Spring	Summer	Fall
1	3000	0.53%	8.30%	7.63%	83.53%	0.008	0.015	-0.019	0.171	6.60%	31.23%	40.20%	21.97%
2	3398	1.71%	14.74%	15.66%	67.89%	0.018	0.016	-0.015	0.125	11.83%	19.54%	38.14%	30.49%
3	4005	3.15%	17.58%	15.58%	63.70%	0.029	0.036	-0.018	0.385	10.21%	30.79%	35.11%	23.90%
4	1578	0.38%	5.39%	5.45%	88.78%	0.022	0.084	-0.020	0.533	58.17%	25.98%	6.34%	9.51%
5	1969	1.73%	12.95%	11.83%	73.49%	0.051	0.061	-0.004	0.558	18.64%	43.22%	22.45%	15.69%
6	3532	2.32%	17.72%	15.35%	64.61%	0.055	0.067	-0.016	0.646	18.32%	25.17%	27.21%	29.30%
7	1708	0.12%	2.93%	4.22%	92.74%	0.073	0.038	0.005	0.336	23.36%	27.11%	27.93%	21.60%
8	2165	0.23%	6.42%	7.39%	85.96%	0.215	0.120	0.011	1.525	54.36%	41.80%	0.37%	3.46%
9	2187	0.59%	7.96%	6.86%	84.59%	0.274	0.169	0.033	2.045	60.40%	23.96%	0.18%	15.45%
10	3736	0.24%	4.50%	4.60%	90.66%	0.296	0.341	-0.018	3.840	33.78%	31.34%	16.22%	18.66%
11	2743	1.90%	16.62%	13.16%	68.32%	0.246	0.460	-0.018	7.528	31.10%	49.73%	8.53%	10.65%
12	4777	0.88%	8.81%	10.01%	80.30%	0.308	0.440	-0.008	7.686	25.66%	29.37%	22.94%	22.02%
13	3690	1.68%	12.76%	11.90%	73.66%	0.353	0.325	0.000	4.712	41.00%	18.70%	5.34%	34.96%
14	1995	0.20%	3.46%	4.06%	92.28%	0.381	0.612	0.003	10.490	25.21%	22.06%	26.22%	26.52%
15	3287	1.73%	14.24%	12.60%	71.43%	0.436	0.441	0.011	6.372	28.23%	11.56%	11.99%	48.22%
16	5681	4.28%	19.80%	16.79%	59.13%	0.376	0.758	-0.017	10.252	15.58%	20.00%	36.28%	28.15%
17	4545	4.99%	23.28%	17.78%	53.95%	0.412	0.706	-0.009	9.266	17.14%	32.70%	26.93%	23.23%
18	4305	1.88%	16.12%	14.61%	67.39%	0.212	1.425	-0.020	19.136	9.22%	17.72%	45.74%	27.32%
19	4699	2.87%	16.49%	16.32%	64.31%	0.510	0.748	0.000	6.183	17.75%	15.54%	28.77%	37.94%
20	4969	2.17%	16.70%	16.12%	65.00%	0.419	0.912	-0.020	11.311	10.14%	14.95%	43.75%	31.15%
21	3303	1.57%	13.44%	11.72%	73.27%	0.442	1.605	-0.020	13.859	13.53%	38.39%	33.97%	14.11%

Table 4-1. Significant properties by cluster number, including percentage of profiles with certain posterior probabilities, surface chlorophyll concentration characteristics, and seasonal distribution.

Fit type	Percentage of profiles
Gaussian	5
Gaussian+Linear	21
Gaussian+Exponential	6
Gaussian+Sigmoid	55
Sigmoid	12
Exponential	0.3

Table 4-2. Distribution of profile curve fits across 6 possible fits.

Chapter 5. IMPACT OF PHYTOPLANKTON SELF-SHADING ON MIXED LAYER TEMPERATURE

5.1 ABSTRACT

Penetration of shortwave radiation exerts significant control on upper ocean heat balance, including the possibility of shifting the balance between positive and negative. Previous studies focusing on the impact of the depth of shortwave radiation penetration have suggested global and regional patterns resulting from different choices of parameterizations for shortwave radiation, but have not quantified the factors that contribute to these patterns. The goals of this study are to quantify how variations in the physical and biological structure of the upper ocean, as well as atmospheric forcing, contribute to differences in mixed layer heating when different parameterizations are used. In particular, we use observational profiles of chlorophyll data derived from biogeochemical Argo floats as the basis for a parameterization using a surface average value of chlorophyll in a simple one-dimensional model, as well as one where absorption is calculated according to chlorophyll concentration at each depth. We compare the results from these to the default constant depth parameterization, and find that some regions of the ocean see a difference in heating >1°C between parameterizations. Mixed layer depth is the most important factor in determining the simulated temperature difference, with additional significant impacts from surface chlorophyll value and wind stress. We find some regions of the ocean that are highly sensitive to the choice of parameterization year round, some that are only seasonally affected, and some where the choice of parameterization makes little difference. We identify the characteristics of vertical chlorophyll profiles that contribute to significant differences in mixed layer heating when a parameterization based on the complete vertical profile of chlorophyll is used rather than one based on a surface average profile, and show that

these are likely to exist in places sensitive to mixed layer heating effects, such as the Northern Indian Ocean and subpolar North Atlantic. Some of these regions overlap significantly with some regions where satellites do not provide very good estimates of subsurface chlorophyll distribution (and thus other quantities like net primary production).

5.2 Introduction

Penetration of shortwave radiation into the ocean plays an essential role in determining the net surface heat flux. Unlike latent, sensible, and longwave heat fluxes, which act as a boundary condition at the air-ocean interface, shortwave radiation can penetrate to depths of up to 100 meters into the ocean, thus causing a vertical transport of heat within the upper ocean. Several observational studies have found that in certain tropical regions, the quantity of shortwave radiation penetrating below the mixed layer can be of the same order of magnitude as the net air-sea heat flux (Lewis et al, 1990; Siegel at al. 1995, Echols and Riser, 2020). The mixed layer heat balance controls the evolution of sea surface temperature (SST) and although the shortwave radiation is just one element of the mixed layer heat balance, it often represents the single largest individual value. Thus, whether the net heat flux into the mixed layer is positive or negative can hinge on the depth to which shortwave radiation penetrates. Changing the upper ocean heat balance also affects upper ocean stratification and mixed layer depth (Yeh et al, Capotondi et al), which can feed back into how much shortwave radiation is retained within the mixed layer, as well as ocean oxygen content and biological productivity (Moore et al, 2018; Schmidtko et al 2017). Variations in mixed layer heating with biological activity and associated ocean optical clarity have been shown to influence upper ocean circulation (Gnanadesikan and Anderson, 2009; Oschlies, 2004) and can also influence feedback to the atmosphere in the form

of sensible heat flux (Jolliff et al, 2012). Thus a complex feedback loop exists between the physical and biological conditions of the upper ocean that requires careful examination.

Prior studies conducted using ocean general circulation models (OGCM) have explored the importance of penetrative radiation on ocean heating and circulation, either by varying the diffuse attenuation coefficient, K_d, in a simple exponential model for shortwave attenuation (Murtugudde et al, 2002; Sweeney et al, 2005), using parameterizations based on satellite observed chlorophyll (Nakamoto et al, 2000), or incorporating phytoplankton self-shading, typically using coupled physical-biogeochemical models (Manizza et al, 2005; Manizza et al, 2008; Oschlies, 2004; Park et al, 2015). Self-shading is the process by which the presence of phytoplankton at each depth alters the available light at subsequent depths. For quantifying penetrative shortwave radiation, self-shading means determining the integrated effect of vertical variations in the chlorophyll concentration. K_d is affected both by the optical properties of water as well as the chlorophyll concentration. The portion of the diffuse attenuation coefficient that depends on phytoplankton concentrations also varies with the actual concentration, potentially yielding substantial differences with an approach that uses an average concentration. The prior studies consistently show that variations in the value of K_d or the actual parameterization used for the absorption of shortwave radiation lead to variations in sea surface temperature (SST), mixed layer depth (MLD), and temperature and stratification below the mixed layer, which can in turn lead to changes in nutrient supply to the mixed layer, chlorophyll concentrations, overturning circulation, meridional transport, and ice melt (Manizza et al, 2005; Manizza et al, 2008; Oschlies, 2004; Park et al, 2015). Studies using simpler models have proposed mechanisms by which varying concentrations of phytoplankton, and thus variations in the way shortwave radiation attenuates, can induce circulation at fronts as a consequence of differential

heating (Edwards et al, 2001; Edwards et al., 2004). In general, incorporating chlorophyll into the parameterization of shortwave radiation leads to warmer sea surface temperatures (and cooler temperatures at depth) due to heightened absorption associated with the presence of organic matter. This also typically increases near-surface stratification and in some cases, leads to net ocean cooling (Oschlies, 2004). However, the approaches in these modeling studies taken vary widely, typically using vastly different "control" parameterizations. The results are rarely related back to the observed vertical structure or surface concentration of chlorophyll measured in situ.

Figure 5.1 shows how the vertical decay of radiation differs depending on whether only the surface chlorophyll concentration is used in determining the distribution of irradiance in the upper ocean or the full vertical variability (self-shading). Self-shading has implications both for the biological community, as it moderates the amount of light received by phytoplankton at different depths, as well as influencing how much solar radiation is absorbed at depth, particularly when there are large variations in the chlorophyll content. Both chlorophyll profiles in **Figure 5.1** have the same surface chlorophyll, and the same depth and peak value for the subsurface chlorophyll maximum, differing only in the width of the gaussian-shaped peak. The distribution of heating differs between the two profiles if the absorption of shortwave radiation is parameterized using only surface average chlorophyll, due to different upper ocean averages (blue lines, right panel). However, the distribution of heating differs more if the vertical distribution of chlorophyll is factored in, particularly in where the heating is concentrated (green lines, right panel). The distribution of phytoplankton within the top 200m of the ocean varies widely (Uitz et al, 2006; Mignot et al, 2011; Carranza et al, 2018), and profiles of physical properties such as temperature and salinity do not always reflect the distribution of phytoplankton (Carranza et al, 2018). Subsurface and deep chlorophyll maxima (SCM and

DCM) are common features in the world ocean (Lewis et al 1983; Cullen, 2015), forming in response to nutrient availability, near surface salinity stratification, and variations in zooplankton grazing. SCM and DCM also evolve with time in response to subsurface physical processes such as motion of the thermocline (Cullen, 2015; Ravichandran et al, 2012).

It is thus well-established that the vertical attenuation of shortwave radiation varies with seawater optical properties, which is related to the biological content, and can have significant effects on ocean circulation. Despite this evidence from OGCMs that the vertical profile of chlorophyll can lead to significant effects on a variety of ocean properties (Manizza et al, 2005), few studies have explored this latter point and those that do have only used coupled physical-biogeochemical models, which often show bias in their chlorophyll concentrations relative to observations (Lim et al, 2018). The reliance on biogeochemical models is due in part to the scarcity of vertical profiles of chlorophyll relative to the fairly comprehensive coverage of the surface ocean achieved by satellites. Aside from repeat hydrographic sections (which are typically separated by roughly a decade for a given cruise track), cruise data from subseasonal regional cruises, and station data from repeat occupations of a single site such as the Hawaii Ocean Time-series (HOT), ongoing vertical profiles of chlorophyll have not been readily available prior to the advent of the use of biogeochemical (BGC) sensors on increasing numbers of Argo floats. However, evidence from the available vertical measurements show that the surface expression of chlorophyll (satellite average values) does not reflect distribution over the remainder of the water column (Arrigo et al, 2011) and may lead to inaccurate predictions of the distribution of heating as well as other quantities of interest such as Net Primary Production (NPP). Therefore, exclusive reliance either on satellites or biogeochemical models to estimate

the distribution of subsurface chlorophyll, and consequent effects on heating, may not provide an accurate understanding.

The value of chlorophyll assessed by satellites provides an average upper water column value (Smith, 1981), and does not fully capture the complexities of the vertical chlorophyll profile and thus its potential impact on shortwave radiation. Several studies have suggested that the vertical profile and integrated chlorophyll content could be established based on the surface concentration alone (Morel and Berthon, 1989; Uitz et al, 2006; Morel et al, 2007). However, these studies were based on a dataset that did not include most of the Indian or Southern Oceans or the Arctic, and thus the patterns observed in those studies may not be generalizeable. Other studies have indicated that mixed layer chlorophyll concentration can change on time scales different from physical mixing processes (e.g. Carranza et al, 2018), or that variations in the subsurface chlorophyll maximum depth and magnitude are tied to subsurface processes (e.g., Ravichandran et al, 2012). Thus near-surface layers that are well-mixed in temperature and salinity are not always associated with uniform upper ocean chlorophyll concentrations. Therefore, attempting to diagnose vertical profile shape universally based on surface chlorophyll expression may falter depending on the relative time scales of physical and biological processes. As such, analysis of a more comprehensive dataset of vertical profiles of chlorophyll has the potential to provide important insight into the variability of vertical chlorophyll expression with physical processes, and how this might impact the absorption of shortwave radiation by the ocean.

Ultimately, a limitation to the existing work is lack of validation with observations of vertical chlorophyll distribution, and an assessment of how time-dependent variations in these profiles might alter the estimates. Even in simulations involving coupled biogeochemical-

physical models (e.g. Manizza et al, 2005; Manizza et al, 2008), there has been little discussion of how the estimates derived relate to realistic observations of the ocean. While some studies attempt to provide regional assessments of appropriate attenuation coefficients based on the how well the model SST matches observations (Murtugudde et al, 2002), and many use either surface chlorophyll values derived from satellites or coupled biogeochemical models, none have attempted to look for spatial and temporal patterns using observational data. It is here that there is a significant opportunity for using the data from an observational program like Argo to assess the significance of self-shading on shortwave radiation absorption in the upper ocean, as well as how this might vary across different regions. Specifically, this database allows for a single data source to be used as basis both for a surface-chlorophyll-only parameterization and one that incorporates information about the vertical distribution of chlorophyll. Considering previous explorations on this topic, we anticipate that both the magnitude and subsurface distribution of chlorophyll (which by necessity impact local water column heating) are likely to impact the evolution of the mixed layer temperature in conjunction with other variables.

The goals of this study are to use a simple model paired with representative vertical chlorophyll profiles derived from the BGC Argo dataset (see Chapter 4) to explore systematic relationships between the evolution of mixed layer temperature and atmospheric forcing, the physical structure of the ocean, and the chlorophyll distribution under several different shortwave radiation parameterizations. In doing so, we hope to provide predictive tools for classifying regions of the ocean where consideration of surface chlorophyll in the parameterization of shortwave radiation is important, and to further clarify which regions require the most careful consideration of vertical chlorophyll distribution. This study differs from those discussed previously in that we restrict our investigation to a one dimensional model and use an

approximation of surface chlorophyll tied directly to the observed vertical distribution of chlorophyll, rather than relying on separate data sources for the surface and vertical distribution of chlorophyll or satellite only estimates of shortwave penetration depths.

5.3 Data and Methods

We used a modified Price-Weller-Pinkel (PWP) model (Price et al. 1986) to simulate the one-dimensional response of the upper ocean to different shortwave radiation absorption parameterizations. The goal of these simulations was to disentangle the relationships between variations in physical or biological characteristics in the ocean (e.g., mixed layer depth, vertical chlorophyll distribution) and atmospheric forcing (heat flux, wind stress) in contributing to the evolution of SST under different parameterizations. PWP is a one-dimensional diagnostic mixed layer model, in which atmospheric heat and freshwater fluxes, as well as initial temperature and salinity profiles, are prescribed, and then allowed to develop following simple heat and salt budgets. This model has been used effectively to explore a range of topics, including the diurnal cycle in the tropics (Price et al, 1986), ice ocean interactions (modified PWP, Wilson and Riser 2019), and the effect of diurnal cycling on intraseasonal variability (Shinoda 2005). We performed a series of idealized simulations using different combinations of realistic forcing and vertical ocean structure over a range of relevant values identified from sensitivity tests. With this set of simulations, we tested three different models for the absorption of shortwave radiation, deriving values for the chlorophyll-dependent absorption cases from a set of 21 realistic chlorophyll profiles that were obtained from our previous work (see Chapter 4).

Using the simulation results, we performed three major analyses to identify the difference in mixed layer temperature produced by the different choices of shortwave radiation absorption models. First, we develop a multiple linear regression using the simulation data to quantify

seasonal patterns in the expected temperature difference as well as the individual contributions of each model variable, and to determine what other effects might contribute to temperature differences that are not considered here. We then perform a clustering analysis on the simulation data, and used the results to classify oceanic regions where different effect sizes were expected based on the climatological ocean conditions, heat fluxes, and wind stress. This produces a broad categorization of ocean regions where we expect incorporating vertical chlorophyll distribution into the model to be more or less important. Finally, we examine the differences in mixed layer temperature produced by considering the surface average chlorophyll value (or "satellite" chlorophyll) rather than the full vertical chlorophyll profile. In this final analysis, we identify a subset of chlorophyll profile types where the temperature difference is most pronounced, use clustering analysis to determine which features are most responsible for the observed differences, and then compare these with the available in situ Argo chlorophyll profiles.

5.3.1 Model Setup

In order to address the questions for this study, the original PWP model was modified to provide three different options for how the absorption of shortwave radiation in the water column was parameterized. The first parameterization is double exponential-decay, as proscribed in the original model (Price et al, 1986):

$$I(z) = I_{IR} \cdot e^{-k_{IR}z} + I_{VIS} \cdot e^{-k_{VIS}z} \quad (1)$$

where $I(z)$ is the irradiance at depth z, I_{IR} and I_{VIS} are the proportions of light distributed in the infrared and visible range respectively (0.6 and 0.4 here), and k_{IR} and k_{VIS} are the vertical attenuation coefficients for the open ocean (1.66 m^{-1} and 0.05 m^{-1} respectively). These differ slightly from those used by Manizza et al (2005) for the "Dead Ocean" scenario, but produce

very similar exponential curves (with a maximum difference of 6% occurring at a depth of 1 m, which is always within the mixed layer). These simulations are referred to as the "Control", or "C".

The second parameterization is a chlorophyll-dependent triple exponential decay that incorporates a surface-average value of chlorophyll, as seen by satellites:

$$I(z) = I_{IR} \cdot e^{-k_{IR}z} + I_{RED} \cdot e^{-k_{RED}z} + I_{BLUE} \cdot e^{-k_{BLUE}z} \quad (2)$$

All terms are the same as in the previous equation, except that I_{IR} and I_{VIS} are now 0.58 and 0.42 and I_{VIS} is evenly partitioned into I_{RED} and I_{BLUE} (so they are each 0.21). This difference in partitioning would, if anything, slightly counteract the effect of increased absorption related to the presence of chlorophyll. We follow the approach Manizza et al (2005) in calculating the light attenuation coefficient as a function of chlorophyll concentration [Chl] in a red and blue band (Morel, 1988),

$$k(\lambda) = k_{sw} + \chi_{(\lambda)}[Chl]^{e_\lambda} \quad (3)$$

where λ is either red or blue/green, k_{sw} is the light attenuation coefficient for optically pure seawater (k_{red} is 0.225 m^{-1} and k_{blue} is 0.0232 m^{-1}), $\chi_{(\lambda)}$ is 0.037 and 0.074 m^{-2} mgChl m^{-3} for the red and blue/green band respectively, and e_λ is 0.629 and 0.674 (unitless) for red and blue/green band respectively (Foujols et al 2000, Morel, 1988). These simulations are referred to as "No self-shading (surface chlorophyll only)" or "NO_SS".

The final parameterization is also a chlorophyll-dependent triple exponential decay, but in this case relies on values of chlorophyll at each vertical level:

$$I(z) = I_{IR} \cdot e^{-k_{IR}z} + I_{RED(z-1)} \cdot e^{-k_{RED}\Delta z} + I_{BLUE(z-1)} \cdot e^{-k_{BLUE}\Delta z} \quad (4)$$

Here everything is the same as in the previous equation except that the light at each level z is computed as a function of the light level of the previous level ($z–1$), Δz is the thickness of the

layer (1m for these simulations). The chlorophyll concentration used for Equation (4) follows the vertical chlorophyll profile that changes with z rather than using a single surface chlorophyll value. These simulations are referred to as "Self-Shading" or "SS". Values for the vertical chlorophyll profiles were based on the 21 representative profiles described in Chapter 4 and were not allowed to change during the simulation. The minimum and maximum surface chlorophyll represented by these profiles are 0.01 mg m^{-3} and 1.035 mg m^{-3}. This represents more than 90% of the surface chlorophyll conditions annually from the Argo chlorophyll profiles and approximately 95% of climatological MODIS-Aqua surface chlorophyll values (NASA GSFC, 2014), and thus captures the overwhelming majority of ocean conditions. To compute values of surface average chlorophyll for Equation (2) that are equivalent to satellite chlorophyll values, we first determined the euphotic zone depth, z_{eu}, from the vertical chlorophyll profile used for Equation 4 (Morel and Maritorena, 2001) and then averaged chlorophyll concentration over the satellite penetration depth z_{pd}, where $z_{pd} = z_{eu}/4.6$ (Uitz et al, 2006; Gordon and McCluney, 1975). This is distinct from previous studies where the vertical chlorophyll profile and surface average (satellite) chlorophyll were derived from different sources (Manizza et al, 2005).

We performed a series of 5 day simulations to provide a realistic time window during which a chlorophyll profile might be relatively static, particularly for low wind stress values (Carranza et al, 2018). We used four different initial temperature (**Figure 5.2**a) and salinity (**Figure 5.2**b) profiles reflecting the range of mixed layer depths over which the greatest effect was observed in sensitivity tests: 15m, 20m, 40m, and 60m. Profiles were taken from MIMOC climatology at 4 locations in the Atlantic Ocean (Schmidtko et al, 2013), and mixed layer depth was determined using a 0.03 kg m^{-3} threshold from 10m density (de Boyer Montegut et al, 2004). Hourly shortwave radiation, Q_{sw}, forcing profiles were constructed following a sawtooth pattern

modeled after the climatologies from ERA5 Reanalysis (Figure 3a), with daily average shortwave radiation varying across simulations between 100 W/m^2 and 300 W/m^2 in increments of 50 W/m^2 (**Table 5-1**). This average shortwave radiation was spread out over a 12-hour daytime periods during the 24 hour day. Surface net heat flux, Q_{net}, varied independently between –50 W/m^2 and 150 W/m^2 in increments of 50 W/m^2 using constant values for latent, sensible, and longwave heat fluxes. Precipitation was set to 0 mm/hr for all simulations so that isolated rain events would not be confused with other factors. Net wind stress, τ, was varied across simulations between 0.025 N/m^2 and 0.3 N/m^2 in increments of 0.025 N/m^2 between 0.025 N/m^2 and 0.125 N/m^2, and increments of 0.05 N/m^2 over 0.125 N/m^2. Each of these settings was paired with each of the 21 vertical average chlorophyll profiles and the three parameterizations of shortwave radiation. **Figure 5.2**c shows the 21 chlorophyll profiles used. In total, this produced 18,900 comparisons among the three parameterizations. **Figure 5.3**a shows a sample heat flux timeseries and **Figure 5.3**b shows a sample wind stress and evaporation-precipitation timeseries (in this case only dependent on evaporation); all of the other variations are simply scaled variations on these shapes. **Figure 5.3**c-d shows sample timeseries of the resulting output for the control simulation, "C" (**Figure 5.3**c) and self-shading simulation, "SS" (**Figure 5.3**d). **Figure 5.3**e shows the difference between **Figure 5.3**c and 3d ("SS"-"C") where positive values indicate a warmer result from using Equation (4) (self-shading) to calculate the absorption of shortwave radiation rather than Equation (1).

For each simulation, the following values were calculated: average shortwave radiation, average net heat flux, average wind stress, average evaporation-precipitation, average mixed layer depth, surface chlorophyll, and integrated chlorophyll. To quantify the difference in heating between simulations, we calculate the average mixed layer temperature over the last 24 hours of

each simulation (see **Figure 5.3** for example), and then take the difference to compare the simulation. We refer to this quantity as ΔT_{ML}(A-B), where A and B indicate the shortwave parameterizations being compared ("SS", "NO_SS", or "C"). Although these are 5 day simulations, ΔT_{ML} values are shown on monthly timescales. **Figure 5.4**a shows a summary scatter plot of each of the input variables plotted against the ΔT_{ML}(SS-C), as well as a correlation heatmap between the model variables and final temperature outcome. The heatmap in **Figure 5.4**b is for used for preliminary variable selection for the subsequent analysis and hypothesis testing (discussed below). The same information for the NO_SS-C comparison is shown in Figures S1 and S2. Evaporation-Precipitation (E-P) was omitted from **Figure 5.4**a because it shows a very low correlation with ΔT_{ML}. The shading of the dots in **Figure 5.4**a aligns with the colors used for the chlorophyll profiles in **Figure 5.2**c. Qualitatively, **Figure 5.4** highlights the general patterns we see in the simulations: higher values of ΔT_{ML}(SS-C), where the use of the self-shading parameterization leads to higher final mixed layer temperature than the control, correlates with higher Q_{sw}, $Q_{net,}$ and surface chlorophyll concentration and lower MLD and wind stress. For mixed layer depths deeper than 60m, very little difference is observed, which supports the choice to restrict the simulations to profiles with initial MLD<=60m. Wind stress, which may contribute to changes in mixed layer depth, seems to have very little impact above 0.2N/m^2, a fairly large effect below 0.2N/m^2, and a somewhat ambiguous effect for the lowest values. Integrated chlorophyll concentration does not show a clear pattern, suggesting that the distribution of the chlorophyll (surface vs. depth) may be more important than the total quantity in the water column except where the latter is a direct function of the former.

5.3.2 Multiple Linear Regression

The goal of performing a multiple linear regression (MLR) is to take the results shown in Figure 4a and develop more quantitative relationships that can be used make predictions in realistic ocean conditions. MLR also allows us to identify both individual and collective contributions to ΔT_{ML} by various factors from the simulation, explore seasonal patterns, and to identify the limits of such a model. MLR takes the form of:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n \quad (6)$$

where Y represents the output variable (in this case, ΔT_{ML}), X represents input variables (i.e. Q_{sw}, MLD, etc), and α represent the coefficients. α_0, the intercept, is the expected value of the output variable with no additional information about the input variables. MLR is most effective if the individual input variables demonstrate a linear relationship with $\Delta T_{ML.}$ However, the scatter plots in Figure 4 suggest otherwise in some cases: surface chlorophyll behaves more as a logarithmic relationship, mixed layer depth reflects an inverse relationship, and wind stress is scaled to $\sqrt{\tau}$, so these variables were scaled accordingly. Each of the raw correlations shown in **Figure 5.4**b improves following these scalings (not shown). Prior to developing the regression each input variable was normalized so that the relative size and distribution of each variable was roughly comparable using the formula:

$$var_{norm} = \frac{var - \overline{var}}{std_{var}} \quad (5)$$

where *var* is one of the input variables (e.g. Q_{sw}), $\overline{var}$ is the numerical average of all the values, and std_{var} is the standard deviation of all the values. This is similar to the methodology taken by

Rosso et al (2019) in their clustering study that used both temperature and salinity. Normalizing the input variables is important when incorporating multiple input variables that have widely different magnitudes and distributions (i.e. Q_{sw}, which varied between 100 and 300 W/m^2 in these simulations and surface chlorophyll concentration, which varied between 0 and 1.035 mg/m^{-3}). Due to the final design of the simulations, each of the input variables was uncorrelated with the others (with the exception of wind speed and mixed layer depth, which is unavoidable in a model that allows for physical development of the mixed layer). We followed a hypothesis testing approach to determine which of the available input variables should be retained in the final model, by successively evaluating the R^2 values and residual sum of squares (*RSS*) for individual and then combinations of variables. **Table 5-2** shows these values, while **Table 5-3** shows the coefficients and intercept for an equation of the form from Equation (6) using scaled/normalized input variables. Ultimately, we retained all of the simulation variables except E-P.

We then applied the resulting MLR to global climatologies of heat flux, wind stress, mixed layer depth, and surface chlorophyll. For heat fluxes and wind stress we used monthly mean data on single levels from 1990-2020 to create 30 year climatologies for eastward and northward turbulent surface stress (to produce net wind stress), surface latent heat flux, surface net long wave radiation flux, surface net shortwave radiation flux, and surface sensible heat flux from ERA5 Reanalysis (Hersbach et al, 2019). For MLD, we used the Argo Mixed Layer depth climatology (Holte et al, 2017; Holte and Talley 2009). For surface chlorophyll, we used MODIS-Aqua 9km binned monthly climatology of chlorophyll-a from 2003-2020 (NASA GSFC, 2014). To make these data compatible with the simulation data, we scaled the mixed layer depth, surface chlorophyll, and wind stress data as described above and then normalized

this test data to the simulation data used to create the MLR. This ensures that the values used for prediction are on the same relative scale as the values used in designing the MLR. In addition to constructing monthly zonal predictions, we also quantify interannual variability using monthly values of each of these variables over the MODIS-Aqua lifetime.

5.3.3 Clustering analysis

Clustering analysis is useful for delineating general regions where different effect sizes are likely to be observed, and can thus be a useful tool when making choices about whether or not to include a more complicated parameterization in a model of observational study. It can also be used to improve predictions in regions or for scenarios where other methods fall short. We perform a supervised K-nearest neighbors (KNN) classification to identify broad patterns in the simulation data. Supervised K-nearest neighbors classification determines the appropriate classification of an unknown point by determining its distance from previously classified points and selecting the closest fit. In this case, the known points are the ΔT_{ML} values calculated from the PWP simulations, and the unknown points are the expected ΔT_{ML} for the real ocean. ΔT_{ML} must first be sorted into bins in order to create the training data. The goal is to create bins that will ultimately produce useful estimates of the magnitude of the impact of incorporating vertical chlorophyll structure into shortwave radiation parameterizations on ΔT_{ML}. For this, we consider both whether the impact of incorporating chlorophyll is positive or negative. We create one bin for all negative effects (i.e. a bin for simulations where the final temperature when incorporating chlorophyll was lower than when using the control parameterization), and divide all positive results into five additional bins. Although the simulations themselves were only 5 days long, we extrapolate to monthly values for the clustering purposes as the data used in making the final

predictions for the real ocean are climatological values, and therefore will roughly account for any variability that might occur between individual 5-day segments. The distribution of ΔT_{ML} values among the bins is shown in Figure S2. The largest number of ΔT_{ML} values from the simulations fall into bin 2: $0<\Delta T<0.15$ºC month^{-1} and the smallest number of ΔT_{ML} values fall into bin 6: $\Delta T>1.2$ºC month^{-1}. However, this alone does not indicate the overall importance of this effect in the global ocean.

We validated the classification scheme by dividing our simulation data into training and testing data and looked at the accuracy of the clustering algorithm in replicating the labels of the pre-labeled testing data. We divided the complete simulation dataset into 80% training and 20% testing subsets, using 100 different subsets to develop a statistical spread. We also used this to test out how many neighbors to use in making the classification (5, 10, 15, or 20), and whether to weight the classification by distance from the nearest neighbors or assign all the nearest neighbors equal weights. Model evaluation results are shown in Figure S3 for ΔT_{ML}(SS-C). Regardless of the number of neighbors used, median classification accuracy was greater than 91% when weighting the impact of the neighbors by distance from the test point. The remaining ~8% were almost entirely assigned to a neighboring bin. For simplicity, we used 5 neighbors for the final model. To create the final classification scheme, we use all of the ΔT_{ML} data points.

Finally, we apply the classification scheme developed from the simulation data to the global ocean using the same data as described for the MLR. This allows us to directly compare the magnitude and geographic distribution of the predictions from each case. In addition, there are some cases (discussed below) where MLR performs poorly or is not a good fit for the data, and in these cases the clustering analysis can compensate.

5.3.4 Curve fitting

The previous approaches do not consider chlorophyll profile shape in making predictions except to the extent that the simulation results are partially due to profile shape. Given that we expect that the differences between the SS and NO_SS simulations are likely to be due in some part to profile shape, predictive models based solely on surface variables may not provide adequate insight, particularly if the ΔT_{ML}(SS-NO_SS) trends are not consistent. To investigate differences in heating that were not well captured purely by surface variables and MLD, or for which a single uniform pattern was not obvious (SS-NO_SS case), we performed a set of simplified calculations similar to those shown in Lewis et al (1983) for a subset of profile types that exhibited the largest ΔT_{ML}(SS-NO_SS). For a set of profiles (discussed more below), we determined the distribution of irradiance predicted by Equation (4) and Equation (2) (as in **Figure 5.1**b,e), determined the expected temperature change at each level (as in **Figure 5.1**c,f) and averaged those temperature changes over the selected MLD to determine the expected mixed layer temperature change for each parameterization. This analysis assumes thorough mixing of temperature down to the base of the mixed layer, and follows the Lewis et al (1983) approach to identifying chlorophyll distributions that might lead to unstable heating configurations. Temperature change is calculated using an assumption of an average shortwave radiation of 300 W/m^2, although this is purely a scaling factor in this case.

In order to determine which profiles to apply this approach to, we first identified the profiles with the largest ΔT_{ML}(SS-NO_SS). We then applied mathematical fits to the relevant chlorophyll profiles from **Figure 5.2**. The profiles in question were well represented mathematically by either a Gaussian combined with linear decrease in chlorophyll from the surface,

$$[Chl](z) = [Chl]_{surf} + \frac{d([Chl])}{dz} z + [Chl]_{max} e^{-\frac{(z-z_{\max})^2}{dz^2}} \quad (7)$$

where [*Chl*] represents the concentration of chlorophyll, *surf* indicates the surface value, *max* indicates the maximum value, $\frac{d([Chl])}{dz}$ represents the linear rate of decrease of chlorophyll from the surface, z_{max} is the depth at which the maximum chlorophyll concentration is observed, and z is the vertical coordinate; or a Gaussian combined with a sigmoid:

$$[Chl](z) = \frac{[Chl]_{surf}}{1+e^{\left(z_{1/2}-z\right)*s}} z + [Chl]_{max} e^{-\frac{(z-z_{\max})^2}{dz^2}} \quad (8)$$

where $z_{1/2}$ represents the depth of the inflection point and *s* represents the sigmoid fit slope. These represent a subset of the 5 (Carranza et al, 2018) or 6 (Chapter 4 of this work) profile descriptions used in previous studies. In order to test how variations in profile shape might affect ΔT_{ML}(SS-NO_SS), artificial chlorophyll profiles were then constructed across a large range of curve-fit parameters by independently varying each parameter. The range of values and increment of change used to construct these profiles is shown in **Table 5-4**, and are based on a realistic range of values identified for chlorophyll profiles derived from Argo floats. We used a value for slope, $\frac{d([Chl])}{dz}$, that ensures 0 chlorophyll at a depth of 250m in Equation (7) and constant value for *s* in Equation (8) to avoid non-physical irregularities in the artificial chlorophyll profiles. For each constructed profile, we calculated the decay of shortwave radiation and the expected temperature change in each 1m bin for a single day using both Equation (2) and

Equation (4). We selected a range of mixed layer depths across the range where the largest ΔT_{ML} was observed (10-80m), and average the temperature change across this layer for each case, retaining the difference as ΔT_{ML}(SS-NO_SS).

We bin the resulting temperature differences into 6 bins, multiplying the daily values by 30 to create an estimate for the monthly ΔT_{ML}(SS-NO_SS) expected. Because these temperature differences only consider differences in shortwave heating over a particular time scale, they are of more value relative to one another than in comparison to the values above or to other work. We use KNN to perform supervised clustering on the normalized curve-fit parameters, using the each ΔT bin as a cluster. We apply the method separately to each of the two curve-fit types discussed above, following the same training-testing approach described in Section 5.3.3. This allows us to determine which combination of profile parameters for a chlorophyll profile and mixed layer depth leads to the largest expected ΔT_{ML}(SS-NO_SS). Based on the training-testing, for the Gaussian/Line profiles, using 10 or more neighbors produces superior results to using 5, but the differences between using uniform and distance-based weights is relatively small (Figure S8), whereas for the Gaussian/Sigmoid profiles all of the options are very similar. We then deploy the KNN model to the Argo chlorophyll profiles after curve-fitting each of the 71272 profiles. We identify profiles that are best represented by either a Gaussian and line (15268 or 21.4% of the total) or a gaussian and sigmoid (39351, or 55.2% of the total), and only apply the relevant KNN model to each subset. In total, this represents 76.6% of the Argo chlorophyll profiles as of the end of 2020.

5.4 Results

5.4.1 MLR (SS-C):

We use a hypothesis testing approach for the multiple linear regression, following the patterns suggested by the correlation heatmap (**Figure 5.4**b). The heatmap shows that average MLD has the highest correlation with ΔT_{ML}(SS-C). If this variable alone is used to form a linear regression, we find that R^2 (SS-C) is 0.34. **Table 5-2** shows the individual R^2 values for each of the independent variables alone. The remaining variables all have a similar magnitude of R^2, and the cumulative R^2 increases with the addition of each variable, so we opted to retain all 5 variables in the MLR. **Table 5-3** shows the intercept and slopes for each of the scaled and normalized input variables in the MLR. Each of these variables has unique geographical distributions that vary seasonally in different ways. **Figure 5.5** and **Figure 5.6** show the geographical distribution and the corresponding expected monthly ΔT_{ML}(SS-C) for each individual term of the MLR for January (**Figure 5.5**) and July (**Figure 5.6**). Positive values indicate that warmer mixed layer temperatures are expected when using the self-shading parameterization (Equation (4)) rather than the control (Equation (1)). These figures highlight the seasonal shifts in each individual variable that eventually contribute to the total. MLD is shallow year round in the eastern tropical Pacific, but shows significant variation in other equatorial regions and at higher latitudes. Thus the temperature difference attributable to MLD follows similar patterns (**Figure 5.5**a, **Figure 5.6**a). Surface chlorophyll shows elevated concentrations at high latitudes in their respective summers, as well as high coastal concentrations during upwelling seasons, but very little variability in the subtropics, making the temperature patterns associated with this variable distinct from that from MLD (**Figure 5.5**b, **Figure 5.6**b). Each of the other three variables also exhibits distinct seasonal cycles. Wind stress (**Figure 5.5**c, **Figure 5.6**c) shows some regions with large effect year round (e.g. in the Antarctic Circumpolar Current (ACC) region, where we expect large wind stress and thus a negative impact on ΔT_{ML}(SS-C),

and other regions with significant seasonal variability, such as the Gulf Stream and Kuroshio regions. Both net shortwave radiation and net heat flux are almost exclusively positive in the hemisphere experiencing summer and negative in the hemisphere experiencing winter, and thus only contributes a positive effect in one hemisphere (**Figure 5.5**c-d, **Figure 5.6**c-d). However, during fall and spring (Figure S4 and S5), we see that net heat flux has less uniform patterns than shortwave radiation, indicating regions where individual outgoing heat flux terms (sensible, latent, or longwave) become dominant in the heat flux balance. Thus, while each variable does not necessarily have the same absolute global importance, they may have elevated importance regionally relative to other variables depending on their seasonal patterns.

Figure 5.7 shows the monthly spatial predictions of ΔT_{ML} (SS-C) from the full MLR, representing the sum of the individual plots in **Figure 5.5** and **Figure 5.6** (Figures S4 and S5) as well as the intercept. The impact of using Equation (4) rather than Equation (1) to parameterize shortwave radiation is large and widespread. In some months, ΔT_{ML} (SS-C) exceeds 0.6°C over most of a hemisphere. In comparing **Figure 5.7** to the individual variable plots, we can pick out which terms are dominant in particular regions. For example, in both January and July, MLD in the Arabian Sea and western equatorial Indian Ocean are fairly deep and thus have negligible effect on ΔT_{ML}(SS-C), but surface chlorophyll is elevated along the coast of Somalia and Arabian Peninsula, contributing to an overall positive ΔT_{ML}. Meanwhile, with the exception of the northernmost section, surface chlorophyll is very low year round in the Bay of Bengal, but has very shallow mixed layers, leading to a strong positive ΔT_{ML}(SS-C) in most seasons. Meanwhile, at higher latitudes, in the ACC region in January, both MLD and wind stress show strong zonal patterns (**Figure 5.5**a, c) but the pattern of high surface chlorophyll concentration downstream of the Drake Passage (**Figure 5.5**b) leads to a more widespread positive ΔT_{ML}(SS-

C) in that region. While much of the northern hemisphere in July experiences relatively shallow MLD, within the range where we expect a significant ΔT_{ML}(SS-C), each of the other variables is distributed more irregularly, leading to an uneven distribution of ΔT_{ML}(SS-C).

Figure 5.8a shows the monthly zonal averages of predicted ΔT_{ML}(SS-C) between 60ºS and 50ºN, with interannual variability indicated by the error bars (standard deviation for the years 2004-2020). The relatively small error bars highlights that the expected ΔT_{ML}(SS-C) is fairly consistent across the years even with variations in individual factors. **Figure 5.8**b shows the zonal SST average calculated from Reynolds OISST (Reynolds et al, 2002), by month. This puts ΔT_{ML}(SS-C) in context with the actual SST in each region. In equatorial regions, the expected temperature difference hovers around 0.25 ºC month^{-1}, with SST on the order of 26-28 ºC. While small compared to the mean, such small changes can contribute to the onset of deep convection if they push the temperature above 28 ºC, and are thus important to get right. Furthermore, tropical variability, such as the intraseasonal Madden-Julian Oscillation (MJO), has a comparable magnitude (Shinoda et al, 1998), suggesting that the ΔT_{ML} due to the choice of shortwave radiation parameterization is equally important to frame accurately. This is doubly the case in scenarios where other patterns of variability, like MJO or interannual variability in the Arabian Sea, exhibit a strong correlation between changes to MLD and surface chlorophyll concentration (Keerthi et al, 2017). Higher latitudes show greater seasonal shifts, which is to be expected from the wider seasonal variation in the variables observed at higher latitudes (**Figure 5.5** and **Figure 5.6**). At 60 ºN for example, the peak summer temperature is roughly 10 ºC, while the ΔT_{ML}(SS-C) is greater than 0.6 ºC, indicating a large effect size relative to the observed surface temperature.

Figure 5.8a also shows some predicted monthly values for ΔT_{ML}(SS-C) that are negative for high latitude bands, implying that using the self-shading parameterization leads to cooler mixed layers. This runs counter to the expectation that incorporating chlorophyll into our parameterization should lead to warmer temperatures. There are a number of possible explanations. First, as discussed in the Section 2.1, the slightly different choice of partitioning between infrared and visible light may lead to different heating distributions where none should exist, but if anything that should skew the results positively towards self-shading since infrared is absorbed closer to the surface. Second, a close look at **Figure 5.4** shows that all of the negative values are concentrated at very low wind, but are distributed across a range of values in the other variables. It is possible that in constructing the MLR, these negative values alter the relationships exhibited by the remainder of the simulation data. The MLR performs best in the middle of the range of ΔT_{ML}(SS-C) (Figure S7, showing the results of applying the MLR to the simulation data), and tends to overpredict negative effects (which are not expected to be as significant as shown in these plots given the increased absorption associated with chlorophyll) and underpredict the most positive effects. Figure S7b shows in particular that for the largest ΔT_{ML}(SS-C) from the simulation data, the MLR produces the best results where the mixed layers in each of the simulation runs (SS, C) remained the most similar. Situations where the MLD diverged significantly led to the least agreement between the simulated values and those predicted by the MLR for the inputs used in the simulation. If anything, this suggests that the predictions at the upper end of the spectrum are likely conservative on monthly timescales (or that any overprediction due to the extrapolation to monthly timescales is compensated for by underprediction of effect size). This is also an example of why a complementary approach, such as clustering, is valuable: while it does not produce the same types of individual relationships

outlined in **Figure 5.5** and **Figure 5.6**, in some cases it better handles patterns such as those highlighted in **Figure 5.4** for very low wind speeds.

5.4.2 Clustering (SS-C and NO_SS-C):

Figure 5.9 shows the breakdown of how each input variable is distributed for each of the temperature bins described earlier. In contrast with MLR, the clustering approach highlights scenarios where individual variables are more and less widely distributed for a particular class of ΔT_{ML}(SS-C) from the simulations. For example, **Figure 5.9**c shows that Bin 1: $\Delta T_{ML}<0$, almost exclusively occurs with very low surface chlorophyll scenarios. As a core feature of that cluster, this means that only data that shares that particular feature will be assigned to that cluster. Thus, even thought $\Delta T_{ML}< 0$ can occur at almost any value of the other variables according to the simulations (**Figure 5.9**a,b,d,e), surface chlorophyll is likely to greatly restrict the distribution of that cluster when applying the clustering scheme to real data. The fact that a subset of the $\Delta T_{ML}<$ 0 values are very negative (**Figure 5.4**) also does not affect the distribution of input variables related to the other bins, and therefore may not suffer from the same skewing observed with the MLR. At the other extreme, we observe that the largest ΔT_{ML} values ($\Delta T_{ML}>1.2$°C month^{-1}) can occur for almost any surface chlorophyll concentration, but are more restricted in all of the other variables: low wind, shallow MLD, large positive Q_{sw}. In setting the distribution of that cluster, higher median surface chlorophyll is likely to contribute (**Figure 5.9**c), but other variables demonstrate narrow permissible ranges. Preliminary work using both a smaller and larger number of bins and different segmentation between the bins showed similar patterns to those shown in **Figure 5.9**; this delineation was ultimately selected because of its efficacy in highlighting the regions of maximum impact.

Figure 5.10shows monthly plots of the predicted ΔT_{ML} (SS-C) from the clustering algorithm applied to monthly climatological data for January, April, July, and October. This figure is equivalent to **Figure 5.7**, showing the predictions for clustering rather than MLR. As with **Figure 5.7**, positive values of ΔT_{ML} indicate warmer temperatures expected when using Equation (4) rather than Equation (1). While many of the patterns between **Figure 5.7** and **Figure 5.10** are similar, including the general magnitude and seasonal patterns of the ΔT_{ML} predicted, there are some pronounced differences. As with the MLR, the clustering reveals that regions experiencing winter tend to have low ΔT_{ML} (SS-C), and are thus unlikely to see major differences between using Equation (1) and Equation (4), likely due to the combination of low chlorophyll, deep winter mixed layers, and low surface irradiation. Subtropical gyre regions, notable for very low surface chlorophyll year round, tend to exhibit $\Delta T_{ML}<0$ in **Figure 5.10**, with an expected temperature difference close to 0 or slightly negative. Unlike **Figure 5.7**, the $\Delta T_{ML}<0$ cluster is restricted to center of these gyre regions, as opposed to being widespread across a whole hemisphere, in keeping with the regions that exhibit the very lowest chlorophyll. In addition, while the location of the largest ΔT_{ML} (SS-C) is similar between **Figure 5.7** and **Figure 5.10**, the regions with the highest ΔT_{ML} (SS-C) are more widespread in **Figure 5.10** than in **Figure 5.7**. This is likely related to the under-prediction of the MLR discussed previously. For example, the eastern equatorial Pacific is a region of large ΔT_{ML} (SS-C) in both plots, largely due to the shallow mixed layers. However, when using the clustering method, the meridional extent of this region is somewhat larger and the overall magnitude of ΔT_{ML} (SS-C) is larger. In contrast, if we compare the October plots from both figures, we see a narrow band in the southern Atlantic and Indian Oceans (30°S) of $0.3°C<\Delta T_{ML}<0.6°C$ suggested by the clustering (**Figure 5.10**), with the surrounding areas are uniformly close to 0. However, in **Figure 5.7** we see a much broader

band, also with a predicted size around 0.3°C. The MLR in this case therefore predicts more widespread, if perhaps somewhat smaller, effects.

We use all of the ΔT_{ML}(SS-C) bin assignments calculated for each of the months, a subset of which are shown in **Figure 5.10**, to determine the concentration of each cluster in the global ocean (**Figure 5.11**). The bin representing the most significant warming, ΔT_{ML}(SS-C)>1.2°C, constitutes a relatively small fraction (~5%) of the behaviors (**Figure 5.11**a), with a slight seasonal cycle peaking in June and December/January. This effect is concentrated in equatorial regions year round, and high latitudes during local summer. Combined with the next largest bin, ΔT_{ML}(SS-C)>0.6°C, comprises more than 20% of the data for much of the year. This is a significant percentage of the data and represents a large difference in expected mixed layer temperature as a result of using a different parameterization. Furthermore, many of these large effect sizes are concentrated in regions that play an important role in oceanic circulation such as the tropical Pacific. The tropical Pacific is also a region that exhibits SST bias in some models, making it a region of particular interest for representing correctly.

Figure 5.11b shows the monthly concentration of each bin except for ΔT_{ML}(NO_SS-C). The patterns for all of the bins where ΔT_{ML}>0.15°C appear almost identical, with the most significant differences appearing in the two lower bins. The relative frequency of ΔT_{ML}<0 increases while 0<ΔT_{ML}<0.15°C decreases. Although this suggests that the cumulative effects of using SS rather than NO_SS might be large, since they are distributed over such a large area, replicating the previous analyses for ΔT_{ML}(SS-NO_SS) yields maps showing a uniformly small effect everywhere, and thus neither the clustering approach nor the MLR provides very much insight into the SS-NO_SS scenario. ΔT_{ML}(SS-NO_SS) does not show distinct patterns for individual forcing variables as used previously (Figure S7). For example, increasing surface

chlorophyll does not correlate with a clear increase or decrease in ΔT_{ML}(SS-NO_SS). Many of the final temperature differences are nominally zero, and even the largest ΔT_{ML} appear to be roughly an order of magnitude smaller than those observed for either ΔT_{ML}(SS-C) or ΔT_{ML}(NO_SS-C) (~0.1°C compared to ~1°C). However, as with ΔT_{ML}(SS-C), even relatively small effects can be important depending on location. Therefore we turn to categorizing ΔT_{ML}(SS-NO_SS) by vertical chlorophyll profile rather than surface features in order to facilitate identifying where significant ΔT_{ML}(SS-NO_SS) might appear in the real ocean.

5.4.3 SS-NO_SS

Figure 5.12 shows the distribution of ΔT_{ML}(SS-NO_SS) from the PWP simulations broken down according to the chlorophyll profile used in the simulation (see **Figure 5.2**). Profiles are positioned adjacent to those with which they have the greatest shape similarity (as described in Chapter 4 of this work). Profiles exhibiting a well-mixed surface behavior (4, 8, 9, etc; **Figure 5.2**c, center) show the smallest ΔT_{ML}(SS-NO_SS), which we reasonably expect given that vigorous mixing of phytoplankton near the surface performs the same purpose physically that vertical averaging of chlorophyll does mathematically. For these types of chlorophyll profiles, it is likely that there is little difference between using Equation (2) and Equation (4), since the average chlorophyll concentration employed in Equation (2) would be identical to the actual concentration over the mixed layer. For other profile shapes, we see that chlorophyll profiles 3, 6, 7, and 18 have the highest median difference between the two simulations (SS, NO_SS) but also exhibit the largest spread in possible outcomes, indicating that profiles of these types could produce either a positive or negative relative effect depending on other physical factors. These profiles all share some generally common features: large (relative

to surface chlorophyll) subsurface peak in chlorophyll occurring relatively close to the surface (**Figure 5.2**c).

Within the group of profiles exhibiting the largest temperature differences, we find that the profiles are best characterized mathematically either as combination Gaussian + Linear (Cluster 3 and 7) or Gaussian + Sigmoid (cluster 6 and 18) (Figure S9). We bin the ΔT_{ML}(SS-NO_SS) values calculated for each set of artificial profiles as described in section 2.4 (using Q_{sw} only), and plot the distribution of the individual curve fit parameters for each temperature bin (**Figure 5.13**, Gaussian+Line only). **Figure 5.13** shows ΔT_{ML}(SS-NO_SS) as monthly values, and we see that the largest ΔT_{ML} value is an order of magnitude smaller than the ΔT_{ML}(SS-C) discussed earlier. However, the values calculated for ΔT_{ML}(SS-NO_SS) using this approach are comparable to those derived from the PWP simulations, and thus give a realistic view of the impact of varying types of profiles. Both as individual effects in critical regions or as a cumulative effect, the relatively smaller values of ΔT_{ML}(SS-NO_SS) compared with ΔT_{ML}(SS-C) are still important, particularly if the former is compounding the latter. **Figure 5.13** shows that scenarios with shallower mixed layer depths, relatively small subsurface chlorophyll peaks, low surface chlorophyll, and shallow chlorophyll peaks are more likely to be associated with a negative ΔT_{ML}(SS-NO_SS). This is qualitatively reasonable because a surface average chlorophyll parameterization (Equation (2)) skews the surface chlorophyll values higher than what is observed at the shallowest depths, concentrating more heating near the surface than would be expected if heating is calculated level-by-level. This seems especially true for the lowest surface chlorophyll values. These plots also show that many of the features of the chlorophyll profile can contribute to a large positive or negative effect depending on the values of other features, as indicated by the somewhat symmetric distribution of surface chlorophyll,

depth of maximum chlorophyll, and width of Gaussian peak. The largest positive ΔT_{ML}(SS-NO_SS) are associated with the largest subsurface chlorophyll peaks, relatively shallow maximum chlorophyll depth, and very low surface chlorophyll and gaussian width. The key distinction between the positive and negative ΔT_{ML}(SS-NO_SS) seems to be MLD and amplitude of the subsurface chlorophyll maximum: negative effects (SS-NO_SS<0) are associated with extremely shallow mixed layer depths and lower amplitude SCM whereas the positive effects (SS-NO_SS>0) are more typically found at intermediate depths and with higher peak chlorophyll values. The same patterns are observed for the Gaussian+Sigmoid calculations (not shown), where the additional parameter z_{half} shows now significant trend. Performing a train-test-fit for the labeled data for each set of profiles, using the temperature bins as labels shows predictive accuracy of approximately 96%.

After applying each KNN model to the appropriate subset of Argo chlorophyll profiles, we can examine the location of profiles showing the largest effects. **Figure 5.14**a shows the location of measured Argo profiles whose shape produced a relatively large difference in heating between the self-shading and average surface chlorophyll parameterizations. Profiles in the Southern Ocean and Kuroshio region always experience a positive ΔT_{ML}(SS-NO_SS), indicating that mixed layer temperatures derived from Equation (4) are warmer than those derived from Equation (2). The Mediterranean Sea, however, is almost exclusively characterized by profiles with ΔT_{ML}(SS-NO_SS) < 0, such that using surface average chlorophyll values leads to warmer temperatures. Most other regions are characterized by a combination of positive and negative effects. In total, there are 214 profiles, which is a very small fraction (0.3%) of the total available profiles in this dataset. However, the profiles are concentrated in certain regions, and seasons, as indicated by the months in **Figure 5.14**b and **Figure 5.14**c. The profiles are most frequently

found in Baffin Bay and the far North Atlantic (reaching 10x the global percentage), northern Indian Ocean, eastern boundary regions of the south Atlantic and Pacific, and the Indian Ocean segment of the Southern Ocean (**Table 5-5**). The largest effects are observed in local summer months, which is consistent with the increased likelihood of significant chlorophyll in the water column in high latitudes in those months. Spring and fall prevalence is more commonly observed in the Northern Indian Ocean (the intermonsoon seasons) and near-coastal regions. Although still a relatively small percentage of the total profiles in each region (**Table 5-5**), several of these regions (notably Baffin Bay/North Atlantic and the Northern Indian Ocean) where the evolution of SST and heat exchange with the atmosphere can have significant consequences. **Figure 5.14**d shows the cluster numbers (associated with profile shape, as in **Figure 5.2**c) where profiles with the larger effects are most commonly observed, and it is clear that cluster 18, a Gaussian+Sigmoid cluster with a large subsurface (but near surface) chlorophyll maximum is the cluster that dominates these results. **Figure 5.14**e shows an example profile from cluster 18 with a relatively large ΔT_{ML}(SS-NO_SS) that is best represented by a Gaussian+Sigmoid, from the Arabian Sea. As expected from **Figure 5.13**, this profile has very low surface chlorophyll but a large amplitude SCM located close to the surface (z<50m) and an intermediate MLD (z~35m). Meanwhile, **Figure 5.14**f shows a profile from the same cluster with a relatively large negative effect from the Arabian Sea region in October. In comparison with **Figure 5.14**e, this profile has a shallower mixed layer and slightly lower SCM amplitude. The depth of the SCM peak in **Figure 5.14**f is also significantly below the base of the mixed layer, whereas the peak in **Figure 5.14**e coincides precisely with the MLD, meaning that any heating happening as a result of shortwave radiation absorbed at this depth could directly affect mixed layer heating and mixing. Both profiles are from the same region and season, indicating how slight changes in both the

physical and biological profiles in a single region can lead to significant changes in ΔT_{ML}(SS-NO_SS). A shift from a positive ΔT_{ML} to a negative ΔT_{ML} as profiles evolve shows the importance of treating shortwave radiation parameterization with nuance in these regions, because unlike ΔT_{ML}(SS-C) where we almost always see a positive value, ΔT_{ML}(SS-NO_SS) can swing both ways.

Furthermore, given that in the world ocean there are more than 4000 profiles assigned to cluster 18 (see Chapter 4) and they are found in several different regions, such as the subpolar North Atlantic, Mediterranean Sea, Northern Indian Ocean, and even the Southern Ocean, it is possible that the current results under-represent profiles that are likely to produce a significant ΔT_{ML}(SS-NO_SS). Furthermore, in the initial data processing of these profiles, a 5 point median filter was used on the data, which can dampen the magnitude of the SCM, particularly for profiles with a relatively large narrow peak. Given that large narrow SCMs are associated with the largest positive ΔT_{ML}(SS-NO_SS), it is also possible that this analysis undercounts these profiles in the global ocean.

5.4.4 Effects on mixed layer depth and stratification

Many of the simulations show a change in mixed layer depth over the course of the five-day period, and as shown in **Figure 5.3**, the change in mixed layer depth may differ across the three different parameterizations. While average MLD of the control simulation is a key feature of both the MLR and the KNN clustering, it does not account for diverging MLD across groups of simulations (SS, NO_SS, and C). **Figure 5.15** shows how differences in average mixed layer depth between the self-shading and control simulations relate to ΔT_{ML}(SS-C), shaded by average mixed layer depth. For a given average mixed layer depth, for all values of ΔT_{ML}(SS-C) > 0, a greater difference in average mixed layer depth between the SS and C simulations (ΔMLD, y-

axis) is highly correlated with larger ΔT_{ML}(SS-C). Negative values of ΔMLD indicate that the self-shading simulation had shallower average mixed layers than the control simulation. Depending on model forcing, especially wind stress, this could be because mixed layers in both simulations deepened, but the control simulation deepened more (as in **Figure 5.3**, where ΔMLD is relatively small). It can also occur if one MLD shoals more than the other over the course of the simulation. For relatively shallow MLD, on the order of 15-20m, relatively large ΔT_{ML}(SS-C) can result even if mixed layers between the simulations remain fairly similar (as in **Figure 5.3**). However, due to the primary role of MLD in determining the impact of selecting different parameterizations, much large differences in average mixed layer depth between simulations are observed for the larger ΔT_{ML}(SS-C) values. Thus, in addition to the general effect of MLD, secondary effects related to changes in MLD during the simulation, and differences in this between the three parameterizations, may play a significant role in ΔT_{ML}(SS-C).

5.5 Discussion

It is well-established that a single value for the exponential decay of shortwave radiation is inadequate to describe upper ocean heating effects in the global ocean. Studies have explored this from many different perspectives: providing regionally specific decay constants (e.g. Murtugudde et al, 2002), using surface chlorophyll to parameterize the decay (Sweeney et al, 2005; Gera et al, 2020), and coupling biogeochemical models to physical models to assess differences in several possible parameterization options (Park et al 2015; Manizza et al, 2005; Manizza et al, 2008; Gnanadesikan and Anderson, 2009). The current study complements this work by assessing the impact of observed vertical chlorophyll profiles, and their satellite observation equivalents (based on the same chlorophyll profile, rather than a separate dataset

entirely), on the one-dimensional evolution of SST. We find the biggest departure from parameterizing the absorption of shortwave radiation with a constant K_d in regions with shallow mixed layers, low wind speeds, and relatively high surface chlorophyll, typically when Q_{sw} and Q_{net} are positive and relatively large. This is consistent with the expectation that in a one-dimensional model, the irradiance curve *I(z)*, and thus differences resulting from the choice of how *I(z)* is parameterized, becomes especially important when wind speeds are low (L. Johnson, personal communication, April 24, 2022).

Direct comparison with studies based on global circulation models is complex, given that a one-dimensional model does not allow for the horizontal redistribution of heat, but we attempt to put these results in context with previous studies here, as well as assess the limitations and useful next steps suggested by this work. One notable result from Manizza et al (2005), which formed the initial motivation for this work, is their observation that the results were not robust to different bio-optical models, particularly in the tropics where the presence of phytoplankton spurred changes to the upwelling system. These same regions are highlighted here as regions that are particularly likely to exhibit large ΔT_{ML} (regardless of whether surface chlorophyll only is used or a self-shading parameterization). These are also regions where biogeochemical models have shown chlorophyll bias in the past (Lim et al, 2018), but did not appear as the regions where self-shading was likely to produce a significantly different outcome from consideration of surface chlorophyll only. This makes a strong supporting argument that chlorophyll should be carefully considered in the highlighted regions when calculating the absorption of shortwave radiation, particularly since these are also regions that show strong SST errors in models and the extent of the regions with the largest ΔT_{ML}(SS-C) changes monthly (e.g. Gnanadesikan and Anderson, 2009, Davey et al, 2002). The zonal mean patterns of monthly differences in SST

show similar magnitudes and patterns in each hemisphere, although we also show that these patterns vary significantly with latitude. In addition, we expect that most of the temperature differences shown in this study should be positive (since there are no changes to circulation, such as upwelling and it is likely that most of the negative values reported here are due to overprediction by the MLR), whereas studies using a GCM are likely to show both positive and negative differences. These results also expand on those of studies that have used a spatially variable K_d but no vertical consideration of chlorophyll (e.g. Murtugudde et al, 2002, which used a single exponential decay) and those that significantly changed the K_d value from a seasonal climatology (e.g. Gnanadesikan and Anderson, 2009), all of which suggest equatorial regions as prone to large changes in SST and subsequent circulation. Other regions highlighted here as of more seasonal importance are also reflected in previous GCM studies, notably the North Atlantic (Oschlies, 2004) and Northern Indian Ocean (Sathyendranath et al, 1991).

Changes in circulation are likely to be associated with changes to the physical structure of the ocean, e.g. mixed layer depth (Sathyendranath et al, 1991). **Figure 5.15** showed that changes in the expected mixed layer temperature are concurrent with changes in mixed layer depth and are likely inter-related to some extent. Shallower initial mixed layers show smaller changes to the mixed layer depth over the course of a simulation, and regardless of overall magnitude of the difference in mixed layer depth. The relative shoaling of the mixed layer observed here when parameterizing shortwave radiation to include chlorophyll rather than a constant value is of the same order of magnitude as that found in previous studies, which have shown changes to meridional transport as a result of shoaling mixed layers (Sweeney 2005). The near-equatorial regions showing the largest SST impacts and thus also (we expect) changes to mixed layer depth, extend well off the equator in some cases particularly in certain zonal bands, and support

previous claims that off-equatorial chlorophyll effects (and/or mis-estimates of heating off-equator) contribute to circulation changes.

One thing we also cannot account for here is feedbacks to the atmosphere due to changes in surface temperatures, but there is evidence that the effects discussed here may have additional effects beyond merely altering upper ocean temperature. Numerous studies have shown that atmospheric feedback is a consequence of changes to heating rate, influencing wind stress (Gnanadesikan and Anderson, 2009) and turbulent heat transfer to the atmosphere (Joliff et al, 2012; Seo et al 2014; Sathyendranath et al, 1991; Shell et al, 2003), often in regions that were highlighted here as having particularly large ΔT_{ML}(SS-C), such as the tropical Pacific or Arabian Sea. In regions such as the eastern tropical Pacific, where SST and air temperature near the surface are very similar, a 1°C ΔT_{ML}(SS-C) means potentially doubling the expected sensible heat flux, and possibly more. Regions with a large atmosphere-ocean temperature difference but a small ΔT_{ML}(SS-C), or a small atmosphere-ocean temperature difference but large ΔT_{ML}(SS-C) may also see cumulative impacts that are significant. In modeling scenarios that permit feedback to the atmosphere, increased sensible heat flux due to elevated SST, coupled with cooler temperatures below the mixed layer as shown in **Figure 5.3**, could ultimately lead to cooler SSTs when later mixed events occurred (Oschlies, 2004).

Studies have also shown that diurnal cycles are of particular importance for intraseasonal variability of SST as well as atmospheric feedback (Shinoda, 2005; Seo et al, 2014; Bernie et al, 2007), especially at low wind speeds (Dickey, 1983). As seen in **Figure 5.3**, one consequence of different choices of parameterization for shortwave radiation is different diurnal cycles of SST and MLD. Anderson et al (2007) showed that equatorial zonal winds weaken in runs without chlorophyll due to weakened SST gradient, and Gera et al (2020) showed that changes to heating

due to the consideration of chlorophyll affected evaporation, with consequences for salinity and circulation. We can see in our results that the Eastern Tropical Pacific shows larger effects from the incorporation of chlorophyll than the west, which would also lead to stronger zonal gradients and concurrent changes in wind stress, complementing that finding. The influence of turbulent heat transfer to the atmosphere can contribute to changes in atmospheric circulation as well through the onset of convection, although it is beyond the scope of this study to address that.

This work also highlights the critical importance of accurate estimates of surface chlorophyll concentration and vertical distribution both in biogeochemical models, from satellites, and from in situ observational platforms. Groeskamp and Iudicone (2018) showed that the effect of including chlorophyll concentration in the parameterization of shortwave radiation had a more significant impact on water mass transformation than differences in freshwater flux products, but applying this result to models or observational studies requires accurate representations of chlorophyll. Lim et al (2019) and others (e.g. Zhang et al, 2011) highlighted the ways in which biases in chlorophyll content in biogeochemical models can affect seasonal cycles in SST when coupled to the physical models. Differentiating real feedbacks and changes to thermal structure and circulation from changes produced by incorrect distributions of chlorophyll can therefore be a significant challenge, particularly when paired with the note by Manizza et al (2005) that different bio-optical models can produce widely differing results. To this end, significant work has been done to demonstrate the importance of data assimilation in biogeochemical models (Gregg, 2008, Ford et al, 2012, Wang et al, 2020), although much of this has focused on using satellite data, which does not always lead to an accurate understanding of subsurface chlorophyll distribution, especially in some of the more sensitive regions identified here.

However, autonomous floats, such as those used in the biogeochemical Argo program, have also shown regional biases in calibration that may affect the accuracy of chlorophyll values used for model assimilation and studies such as this using profile details (Roesler et al 2017, Mignot et al 2019, Terzic 2019). Furthermore, several recent studies have highlighted that changes to surface chlorophyll often result from redistribution of chlorophyll rather than changes in quantity (Wang et al 2020) and that there are significant overlaps in the ranges of surface chlorophyll concentrations observed in very different types of profiles (Chapter 4 of this work). The evidence provided here that the shape of the chlorophyll profile is a key determinant in whether or not a surface chlorophyll based parameterization of shortwave radiation is adequate requires that observations of the vertical distribution of chlorophyll be reflected in data assimilation efforts and also that any concrete values of chlorophyll used from these sources is accurately calibrated (as is true of any other data source or model).

Arrigo et al (2011) showed that certain types of profiles were likely to show the largest discrepancy between satellite and in-situ estimates of NPP, and there is significant overlap with the profile characteristics shown here to produce major discrepancies in heating: low surface chlorophyll and a large subsurface peak in chlorophyll (but not necessarily extremely shallow). The biggest differences between NPP estimated from satellites and from in situ observations occurred where the SCM was slightly deeper and the surface chlorophyll concentration was low. This study showed that errors in satellite NPP estimates could be offset by the errors in satellite estimates, but it is not clear that this same mechanism works in relation to heating, particularly given the range of available parameterizations. The Arctic is also a region that is not well observed by in situ platforms such as the biogeochemical Argo program at the moment (outside

of specific regions like Baffin Bay, e.g. Randelhoff et al, (2020)) as well as one that is undergoing significant changes to the background physical structure as sea ice extent changes.

Global estimates of long term changes in chlorophyll content show that even in cases where there is no global average change, some regions may be particularly affected. Gregg and Rousseaux (2014) indicate that while a 15 year record from the SeaWiFS satellite mission did not show a significant global trend, some regions that we find to be of particular interest here (such as the North Atlantic and equatorial Indian Ocean) did show significant decreases. Global models have shown sensitivity to climate change, predicting a decrease in primary production and expansion of oligotrophic regions, particularly in equatorial and subtropical regions, and is fairly consistent across different models (Cabré et al, 2014). As in the Arctic, changes in SST, meltwater quantities and distribution, upper ocean stratification, and the resulting distribution and availability of nutrients therefore may alter the patterns in phytoplankton growth and chlorophyll distribution we see as well, which is one of the key factors in determining the magnitude of the impacts of different shortwave radiation parameterizations. In this study we did not allow for influx of freshwater or the presence of barrier layers and fresh lenses, phenomena known to be common in the Northern Indian Ocean, and increasingly common in high-latitude regions adjacent to meltwater, such as Greenland (Luo et al, 2016). This may be particularly relevant for the results related to ΔT_{ML}(SS-NO_SS), considering the concentration of the highest impact profiles in these regions. Given the expected changes to the hydrological cycle that produces these phenomena, this forms the basis of ongoing future work: to explore how salinity stratification obstructing access to a deep thermocline and nutrients works in tandem with differing distributions of shortwave radiation to alter heating profiles.

Beyond restricting the focus of this study to a one-dimensional model, there are several other limitations to this study that may influence the generality and applicability to some regions. First, we did not allow the chlorophyll profile to evolve over the course of the 5-day simulation and extrapolated the results to a monthlong climatological estimate. Comparison with other versions of the simulations showed that the estimates here are smaller than those achieved by running monthlong simulations. Given that the global predictions are based on monthly climatologies, and that in many regions profiles of a similar nature are observed in certain months (even if there is intermittent mixing at times that alters the profile shape somewhat), we think that these limitations are unlikely to affect the magnitude of the estimates here. Furthermore, the regions and types of profiles showing the greatest impact will remain the same, even if slight alterations in the magnitude of difference is observed.

Finally, there are many opportunities to expand on this work to explore how model vertical resolution affects the magnitude and overall sign of the impacts as well as how these changes are related to differences in the diurnal cycle. Previous model studies relying on OGCM are better able to represent the 3D circulation but have much lower vertical resolution, and therefore may not see the same effects described here. Completing the same simulations outlined here but on a coarser grid will provide more insight into how the results observed here would manifest in a different modeling scenario. Furthermore, differences in diurnal cycle have been shown to have significant effect under different forcing scenarios and optical characteristics (Dickey, 1983; Shinoda 2005) and this warrants further investigation to determine the importance of transient effects on top of the average effects described here.

5.6 ACKNOWLEDGMENTS

Many thanks to my lab group who looked at many rough drafts of analysis, figures, and presentations in the process of putting this chapter together. The Argo data used to create this climatology were collected and made freely available by the International Argo Program and the national programs that contribute to it. (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. doi.org/10.17882/42182#56126. NOAA_OI_SST_V2 data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their Web site at https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html

References

Anderson, W.G., Gnanadesikan, A., Hallberg, R., Dunne, J., & Samuels, B.L. (2007). Impact of ocean color on the maintenance of the Pacific Cold Tongue. Geophysical Research Letters, 34, L11609, doi:10.1029/2007GL030100

Arrigo, K.R., Matrai, P.A., & van Dijken, G.L. (2011) Primary productivity in the Arctic Ocean: Impacts of complex optical properties and subsurface chlorophyll maxima on large-scale estimates. Journal of Geophysical Research, 116, C11022, doi:10.1029/2011JC007273

Bernie, D.J., Guilyardi, E., Madec, G., Slingo, J.M., Woolnough, S.J. (2007). Impact of resolving the diurnal cycle in an ocean-atmosphere GMC. Part 1: a diurnally forced OGCM. Climate Dynamics, 29: 575-590.

Cabré, A., Marinov, I., & Leung, S. (2014) Consistent global responses of marine ecosystems to future climate change across the IPCC AR5 earth system models. Climate Dynamics, 45, 1253-1280, doi:10.1007/s00382-014-2374-3.

Capotondi, A., Alexander, M.A., Bond, N.A., Curchitser, E.N., & Scott, J.D. (2012). Enhanced upper ocean stratification with climate change in the CMIP3 models. Journal of Geophysical Research, 117, C04031, doi:10.1029/2011JC007409

Carranza, M. M., Gille, S. T., Franks, P. J., Johnson, K. S., Pinkel, R., & Girton, J. B. (2018). When mixed layers are not mixed. Storm-driven mixing and bio-optical vertical gradients in mixed layers of the Southern Ocean. Journal of Geophysical Research: Oceans, 123, 7264–7289. https://doi.org/10.1029/2018JC014416.

Cullen, J. J. (2015). Subsurface chlorophyll maximum layers: enduring enigma or mystery solved? Annual Review of Marine Science, 7, 207-239, doi: 10.1146/annurev-marine-010213-135111

Davey, M.K., Huddleston, M., Sperber, K.R., Braconnot, P., Bryan, F., Chen, D., Colman, R.A., Cooper, C., Cubasch, U., Delecluse, P., DeWitt, D., Fairhead, L., Flato, G., Gordon, C., Hogan, T., Ji, M., Kimoto, M., Kitoh, A., Knutson, T.R., Latif, M., Le Treut, H., Li, T., Manabe, S., Mechoso, C.R., Meehl, G.A., Power, S.B., Roeckner, E., Terray, L., Vintzileos, A., Voss, R., Wang, B., Washington, W.M., Yoshikawa, I., Yu, J.-Y., Yukimoto, S., & Zebiak, S.E. (2002). STOIC: a study of coupled model climatology and variability in tropical ocean regions. Climate Dynamics,18: 403-420, doi: 10.1007/s00382-001-0188-6

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of Geophysical Research, 109, C12003, doi:10.1029/2004JC002378.

Dickey, T.D. (1983). The influence of optical water type on the diurnal response of the upper ocean. Tellus, 35B, 142-154.

Echols, R., and Riser, S.C., 2020: The Impact of Barrier Layers on Arabian Sea Surface Temperature Variability. *Geophysical Research Letters*, **47**, e2019GL085290, https://doi.org/10.1029/2019GL085290

Edwards, A. M., Platt, T., & Wright, D. G. (2001). Biologically induced circulation at fronts. *Journal of Geophysical Research: Oceans*, *106*(C4), 7081-7095.

Edwards, A. M., Wright, D. G., & Platt, T. (2004). Biological heating effect of a band of phytoplankton. *Journal of Marine Systems*, *49*(1-4), 89-103.

Ford, D.A., Edwards, K.P., Lea, D., Barciela, R.M., Martin, M.J., & Demaria, J. (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. Ocean Science, 8, 751-771.

Foujols, M. A., Lévy, M., Aumont, O., & Madec, G. (2000). OPA 8.1 Tracer model reference manual. *Institut Pierre Simon Laplace*, 39.

Gera, A., Mitra, A.K., McCreary, J.P., Hood, R., & Momin, I.M. (2020). Impact of chlorophyll concentration on thermodynamics and dynamics of tropical Indian ocean. Deep-Sea Research Part II, 179.

Gnanadesikan, A. & Anderson, W.G. (2009). Ocean Water Clarity and the Ocean General Circulation in a Coupled Climate Model. Journal of Physical Oceanography, 39, 314-332.

Gordon, H.R. & McCluney, W. R. (1975). Estimation of the Depth of Sunlight Penetration in the Sea for Remote Sensing. Applied Optics, 14, 413-416.

Gregg, W.W. (2008). Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. Journal of Marine Systems, 69: 205-225.

Gregg, W.W., & Rousseaux, C.S. (2014). Decadal trends in global pelagic ocean chlorophyll: A new assessment integrating multiple satellites, in situ data, and models. Journal of Geophysical Research: Oceans, 119, 5921-5933, doi:10.1002/2014JC010158.

Groeskamp, S., & Iudicone, D. (2018). The effect of air-sea flux products, shortwave radiation depth penetration, and albedo on the upper ocean overturning circulation. Geophysical Research Letters, 45, 9087–9097. https://doi.org/10.1029/2018GL078442

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2019): ERA5 monthly averaged data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on < 14-06-2021 >), 10.24381/cds.f17050d7

Holte, J., & Talley, L.D. (2009) A New Algorithm for Finding Mixed Layer Depths with Applications to Argo Data and Subantarctic Mode Water Formation. Journal of Atmospheric and Oceanic Technology, 26, 1920-1939, doi:10.1175/2009JTECHO543.1

Holte, J., Talley, ., L. D., Gilson, J., & Roemmich, D. (2017), An Argo mixed layer climatology and database, Geophys. Res. Lett., 44, 5618–5626, doi:10.1002/2017GL073426.

Jolliff, J.K., Smith, T.A., Barron. C.N., deRada, S., Anderson, S.C., Gould, R.W., & Arnone, R.A. (2012). The impact of coastal phytoplankton blooms on ocean-atmosphere thermal energy exchange: Evidence from a two-way coupled numerical modeling system. Geophysical Research Letters, 39, L24607, doi:10.1029/2012GL053634.

Keerthi, M.G., Lengaigne, M., Levy, M., Vialard, J., Parvathi, V., de Boyer Montégut, C., Ethé, C., Aumont, O., Suresh, I., Akhil, V.P., Muraleedharan, P.M. (2017). Physical control of interannual variations of the winter chlorophyll bloom in the northern Arabian Sea. Biogeosciences, 14, 3615-3632, https://doi.org/10.5194/bg-14-3615-2017.

Lewis, M.R., Carr, M.E., Feldman, G.C., Esaias, W., & McClain, C. (1990). Influence of penetrating solar radiation on the heat budget of the equatorial Pacific Ocean. Nature, 347.

Lewis, M.R., Cullen, J.J., & Platt, T. (1983). Phytoplankton and Thermal Structure in the Upper Ocean: Consequences of Non-uniformity in Chlorophyll Profile. Journal of Geophysical Research, 88, C4, 2565-2570.

Lim, H.-G., Park, J.Y., & Kug, J.-S. (2018). Impact of chlorophyll bias on the tropical Pacific mean climate in an earth system model. Climate Dynamics, 51:2681-2694.

Luo, H., Castelao, R.M., Rennermalm, A.K., Tedesco, M., Bracco, A., Yager, P., & Mote, T. (2016). Oceanic transport of surface meltwater from the southern Greenland ice sheet. Nature Geoscience, 9, 528-532.

Manizza, M., Le Quéré,C., Watson, A.J., & Buitenhuis, E.T. (2005). Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. Geophysical Research Letters, 32, L05603, doi:10.1029/2004GL020778.

Manizza, M., Le Quéré,C., Watson, A.J., & Buitenhuis, E.T. (2008). Ocean biogeochemical response to a phytoplankton-light feedback in a global model. Journal of Geophysical Research, 113, C10010, doi:10.1029/2007JC004478.

Moore, J.K., Fu, W., Primeau, F., Britten, G.L., Lindsay, K., Long, M., Doney, S.C., Mahowald, N., Hoffman, F., & Randerson, J.T. (2018). Sustained climate warming drives declining marine biological productivity. Science, 359, 1139-1143

Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., & Ras, J. (2011). From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-a concentration. Biogeosciences, 8, 2391-2406.

Mignot, A., D'Ortenzio, F., Taillandier,V., Cossarini, G., & Salon, S. (2019).Quantifying observational errors in Biogeochemical-Argo oxygen, nitrate ,and chlorophyll *a* concentrations. Geophysical Research Letters, 46, 4330–4337, doi: 10.1029/2018GL080541

Morel, A. (1988) Optical Modeling of the Upper Ocean in Relation to Its Biogenous Matter Content (Case I Waters). Journal of Geophysical Research, 93, C9, 10,749-10,768.

Morel, A. & Berthon, J.-F. (1989). Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. Limnology and Oceanography, 34(8), 1545-1562.

Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sensing of the Environment, 111, 69-88.

Morel, A. & Maritorena, S. (2001) Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research, 106, C4, 7163-7180

Murtugudde, R., Beauchamp, J., McClain, C.R., Lewis, M., & Busalacchi, A.J. (2002). Effects of Penetrative Radiation on the Upper Tropical Ocean Circulation. Journal of Climate, 15, 470-486.

Nakamoto, S., Prasanna Kumar, S., Oberhuber, J.M., Muneyama, K., & Frouin, R. (2000). Chlorophyll Modulation of Sea Surface Temperature in the Arabian Sea in a Mixed-Layer Isopycnal General Circulation Model. Geophysical Research Letters, 27(6), 747-750.

NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group; (2014): MODIS-Aqua Ocean Color Data; NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing
Group. http://dx.doi.org/10.5067/AQUA/MODIS_OC.2014.0 (Accessed on 6/1/2021)

Oschlies, A. (2004). Feedbacks of biotically induced radiative heating on upper-ocean heat budget, circulation, and biological production in a coupled ecosystem-circulation model. Journal of Geophysical Research, 109, C12031, doi:10.1029/2004JC002430.Joliff et al (2012)

Park, J.-Y., Kug, J.-S., Bader, J., Rolph, R., & Kwon, M. (2015). Amplified Arctic warming by phytoplankton under greenhouse warming. PNAS, 112(19), 5921-5926.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Journal of Machine Learning Research, 12:2825-2830.

Price, J.F., Weller, R.A., & Pinkel, R. (1986). Diurnal Cycling: Models and Observations of the Upper Ocean Response to Diurnal Heating, Cooling, and Wind Mixing. Journal of Geophysical Research, 91, C7, 8411-8427.

Randelhoff, A., Lacour, L., Marec, C., Leymarie, E., Lagunas, J., Xing, X., Darnis, G., Penkerc'h, C., Sampei, M., Fortier, L., D'Ortenzio, F., Claustre, H., & Babin, M. (2020). Arctic mid-winter phytoplankton growth revealed by autonomous profilers. Science Advances, 6: eabc2678.

Ravichandran, M., Girishkumar, M.S., & Riser, S. (2012). Observed variability of chlorophyll-*a* using Argo profiling floats in the southeastern Arabian Sea. Deep-Sea Research 1, 65, 15-25.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. J. Climate, 15, 1609-1625.

Roesler, C. Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., D'Ortenzio,F., Ras, J., Drapeau, S., Haëntjens, N, Barbieux, M. (2017) Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs Eco sensors. Limnology and Oceanography: Methods, 15, 572-585.

Rosso, I., Mazloff, M. R., Talley, L. D., Purkey, S. G., Freeman, N. M., & Maze, G. (2020). Water mass and biogeochemical variability in the Kerguelen sector of the Southern Ocean: A machine learning approach for a mixing hot spot. Journal of Geophysical Research: Oceans, 125, e2019JC015877, doi:10.1029/2019JC015877

Sathyendranath, S., Gouveia, A.D., Shetye, S.R., Ravindran, P, & Platt, T (1991). Biological control of surface temperature in the Arabian Sea. Nature, 349, 54-56.

Schmidtko, S., Johnson, G. C. & Lyman, J. M. (2013). MIMOC: A Global Monthly Isopycnal Upper-Ocean Climatology with Mixed Layers, *J. Geophys. Res.*, **118**, 1658-1672, https://doi.org/10.1002/jgrc.20122.

Schmidtko, S., Stramma, L., & Visbeck, M. (2017). Decline in global oceanic oxygen content during the past five decades. Nature, 542, 335-339, doi:10.1038/nature21399

Seo, H., Subramanian, A.C., Miller, A.J., & Cavanaugh, N.R. (2014). Coupled Impacts of the Diurnal Cycle of Sea Surface Temperature on the Madden–Julian Oscillation. Journal of Climate, 27, 8422-8443.

Shell, K.M., Frouin, R., Nakamoto, S., & Somerville, R.C.J. (2003). Atmospheric response to solar radiation absorbed by phytoplankton. Journal of Geophysical Research, 108, D15, 4445, doi:10.1029/2003JD003440.

Shinoda, T. (2005). Impact of the Diurnal Cycle of Solar Radiation on Intraseasonal SST Variability in the Western Equatorial Pacific. Journal of Climate, 18, 2628-2636.

Shinoda, T., Hendon, H.H., & Glick, J. (1998). Intraseasonal Variability of Surface Fluxes and Sea Surface Temperature in the Tropical Western Pacific and Indian Oceans. Journal of Climate, 11, 1685-1702.

Siegel, D.A., Ohlmann, J.C., Washburn, L., Bidigare, R.R., Nosse, C.T., Fields, E., & Zhou, Y. (1995). Solar radiation, phytoplankton pigments and the radiant heating of the equatorial Pacific warm pool. Journal of Geophysical Research, 100, C3, 4885-4891.

Smith, R.C. (1981). Remote Sensing and Depth Distribution of Ocean Chlorophyll. Marine Ecology Progress Series, 5(3), 359-361.

Sweeney, C., Gnanadesikan, A., Griffies, S.M., Harrison, M.J., Rosati, A.J., & Samuels, B.L. (2005). Impacts of Shortwave Penetration Depth on Large-Scale Ocean Circulation and Heat Transport. Journal of Physical Oceanography, 35, 1103-1119.

Terzic, E., Lazzari, P., Organelli, E., Solidoro, C., Salon, S., D'Ortenzio, F., & Conan, P. (2019). Merging bio-optical data from Biogeochemical-Argo floats and models in marine biogeochemistry. Biogeosciences, 16, 2527-2542, doi: 10.5194/bg-16-2527-2019.

Uitz , J., Claustre, H., Morel, A., & Hooker, S.B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. Journal of Geophysical Research, 111, C08005, doi:10.1029/2005JC003207.

Wang, B., Fennel, K., Yu, L., & Gordon, C. (2020). Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. Biogeosciences, 17, 4059-4074. Doi: 10.5194/bg-17-4059-2020

Wilson, E.A., Riser, S.C., Campbell, E.C., & Wong, A.P.S. (2019). Winter Upper-Ocean Stability and Ice-Ocean Feedbacks in the Sea Ice-Covered Southern Ocean. Journal of Physical Oceanography, 49(4), 1099-1117.

Yeh, S.-W., Yim, B.Y., Noh, Y. & Dewitte, B. (2009). Changes in mixed layer depth under climate change projections in two CGCMs. Climate Dynamics, 33, 199-213, doi: 10.1007/s00382-009-0530-y

Zhang, R.H., Chen, D., & Wang, G. (2011). Using Satellite Ocean Color Data to Derive an Empirical Model for the Penetration Depth of Solar Radiation (H_p) in the Tropical Pacific Ocean. Journal of Atmospheric and Oceanic Technology, 28, 944-965.

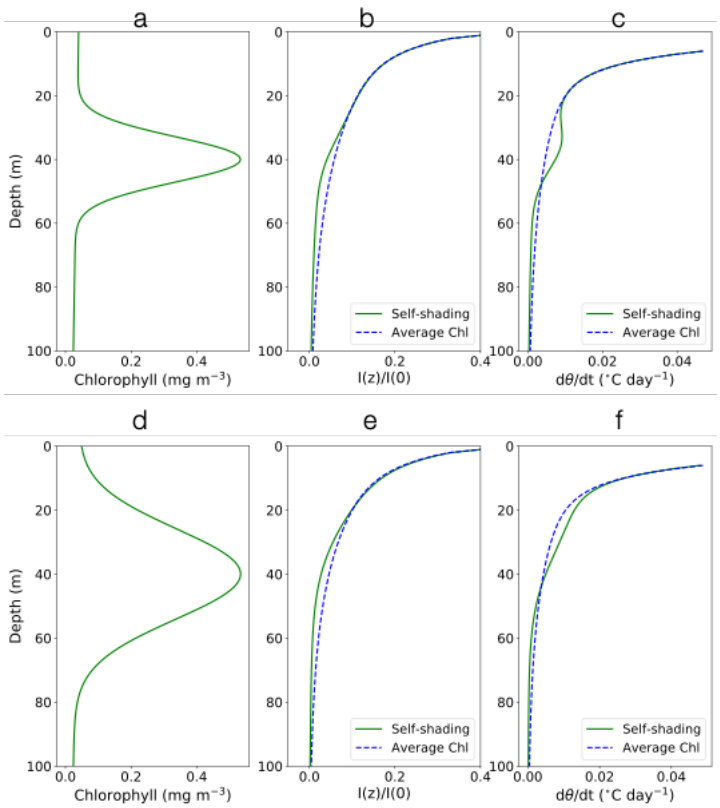

Figure 5.1. (a), (d) sample chlorophyll profiles showing two different subsurface chlorophyll maximum; (b), (e) corresponding irradiance vs depth curves illustrating the difference in absorption profiles between a self-shading parameterization (green) and average surface chlorophyll based parameterization (blue); and (c), (f) expected rate of temperature change by depth for each of the two parameterizations (in °C day^{-1}).

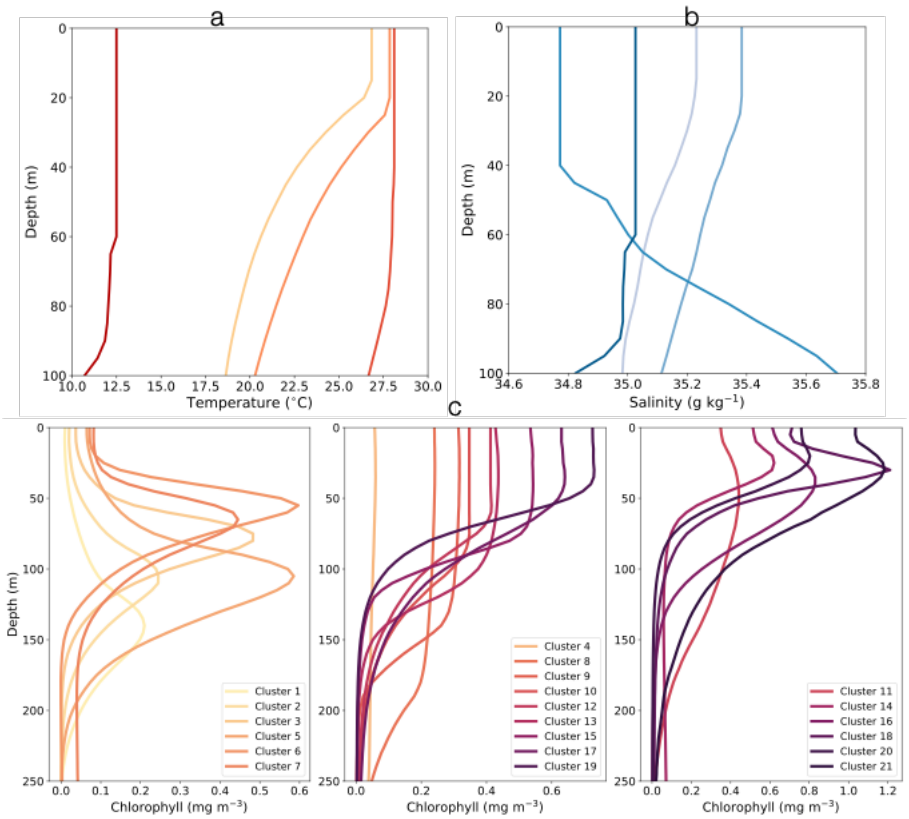

Figure 5.2. Initial conditions for the model: (a) Temperature profiles for MLD_i = 15m (yellow), MLD_i = 20m (light orange), MLD_i = 40m (dark orange), and MLD_i = 60m (red); (b) Salinity profiles (MLD_i = 15m in lightest blue; MLD_i = 60m in darkest blue), and (c) 21 chlorophyll profiles from Echols et al (202__). Profile with MLD_i = 40m is the only one where salinity exerts a strong control on stratification in conjunction with temperature. Colorcoding in (c) indicates progression of surface chlorophyll from lowest (yellow) to highest (purple).

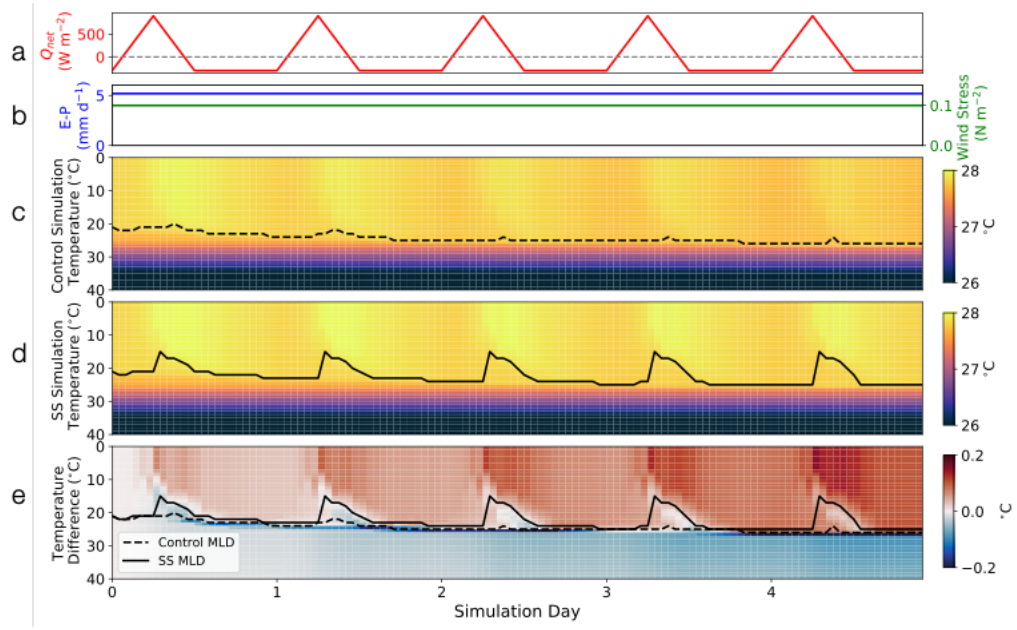

Figure 5.3. Example forcing and model output. (a) Net heat flux (12 hour daytime). (b) Evaporation-precipitation (blue line) and wind stress (green line). (c) Results from control simulation (PWP default parameterization of Q_{sw}). (d) Results from self-shading parameterization. (e) Temperature difference between panels (d) and (c) (specifically, (d)-(c)). In panels (c) and (d), most of the change in mixed layer depth occurs during the first day of the simulation, whereas temperature in the mixed layer changes throughout. Black boxes in panel (c) and (d) represent the values averaged together to determine the final average temperature for each simulation.

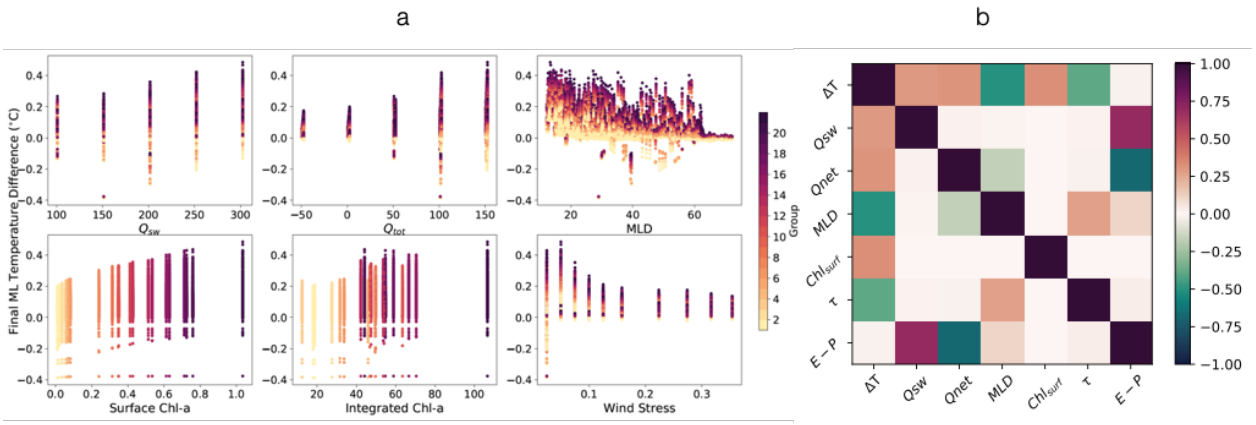

Figure 5.4. (a) Scatter plot of model variables vs. final temperature mixed layer temperature differences for all simulation results (SS-C comparison). Coloring of dots corresponds to chlorophyll profiles (Group) shown in Figure 2 (coloring is sequenced from low surface chlorophyll in light yellow to high surface chlorophyll in dark purple). (b) Heatmap showing correlation between model variables and the final temperature difference between simulations for SS-C simulations. All correlations between ΔT and the independent variables (first column and/or first row) are significant at the $p<0.01$ level using a Pearson correlation coefficient and p-value test.

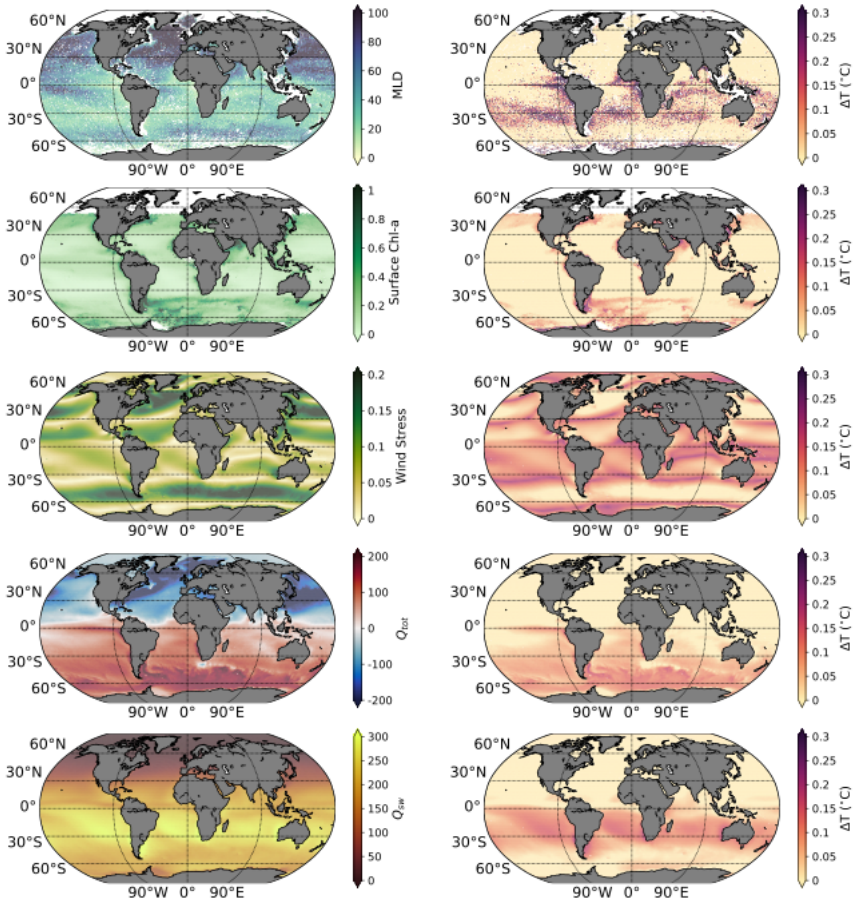

Figure 5.5. Temperature difference patterns (SS-C) associated with each independent variable for January. (a) Mixed layer depth. (b) Surface chlorophyll. (c) Wind stress. (d) Total heat flux. (e) Shortwave radiation.

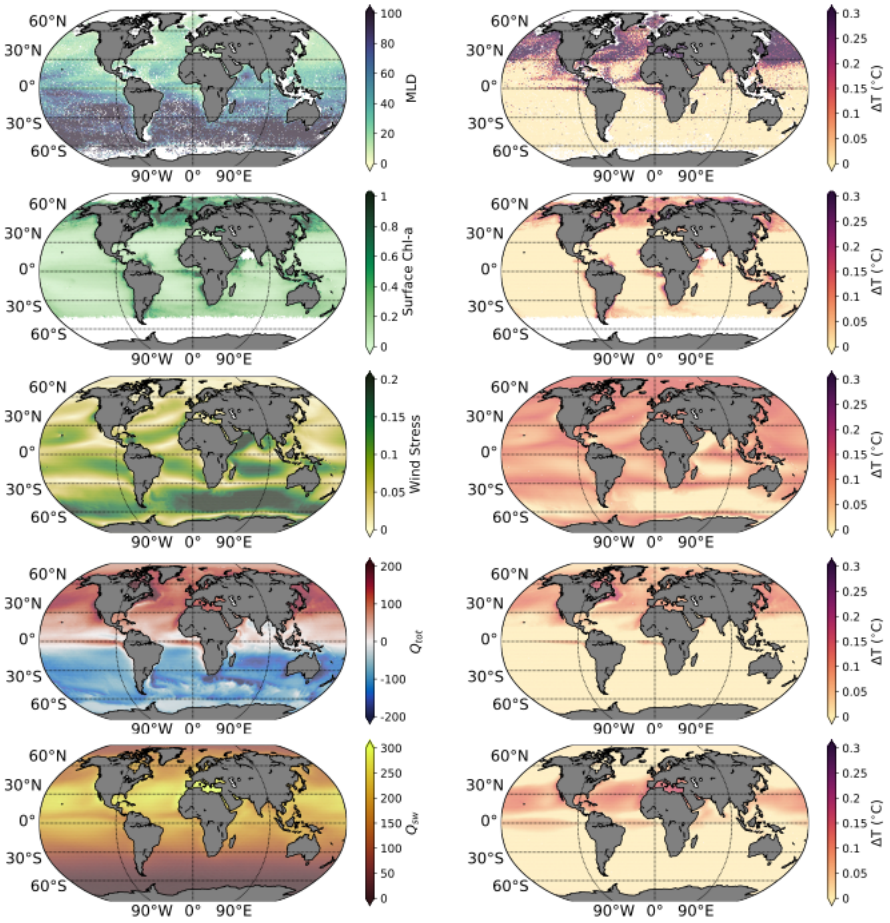

Figure 5.6. Same as Figure 5.5 but for July.

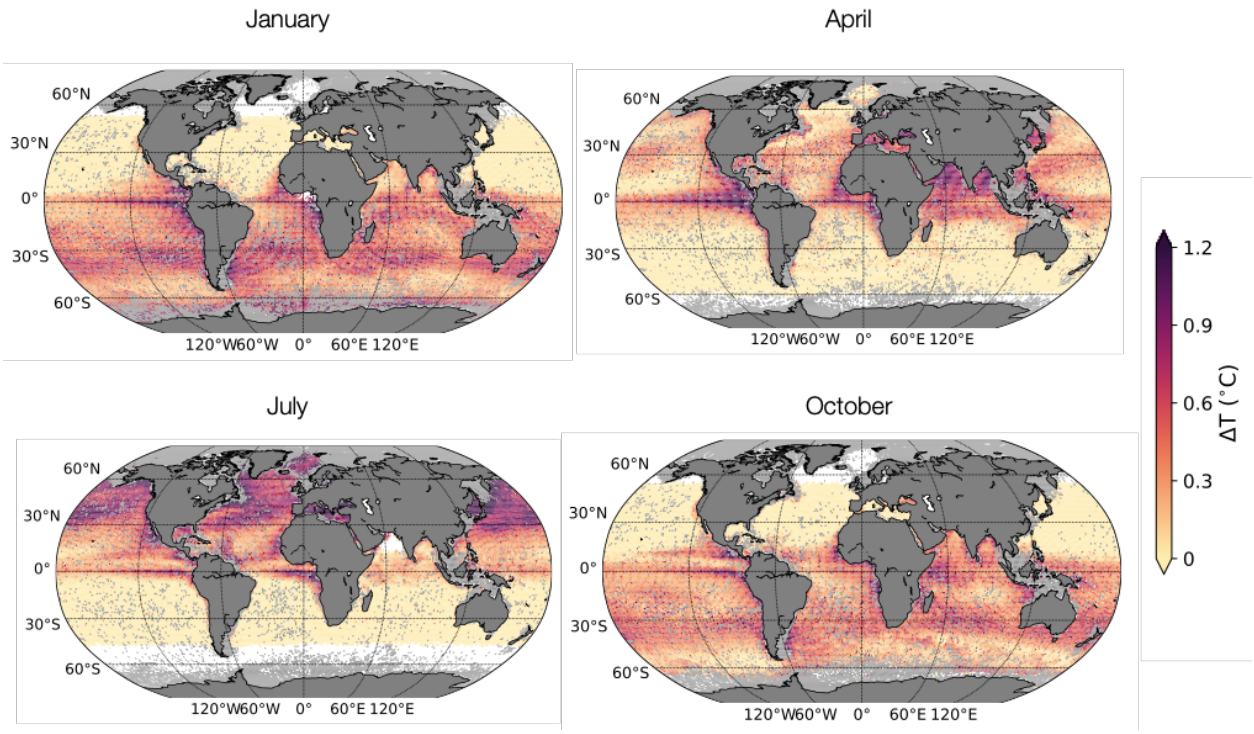

Figure 5.7. Monthly maps of expected temperature differences predicted from multiple linear regression for SS-C comparison. Grayed out regions are those with MLD>70m.

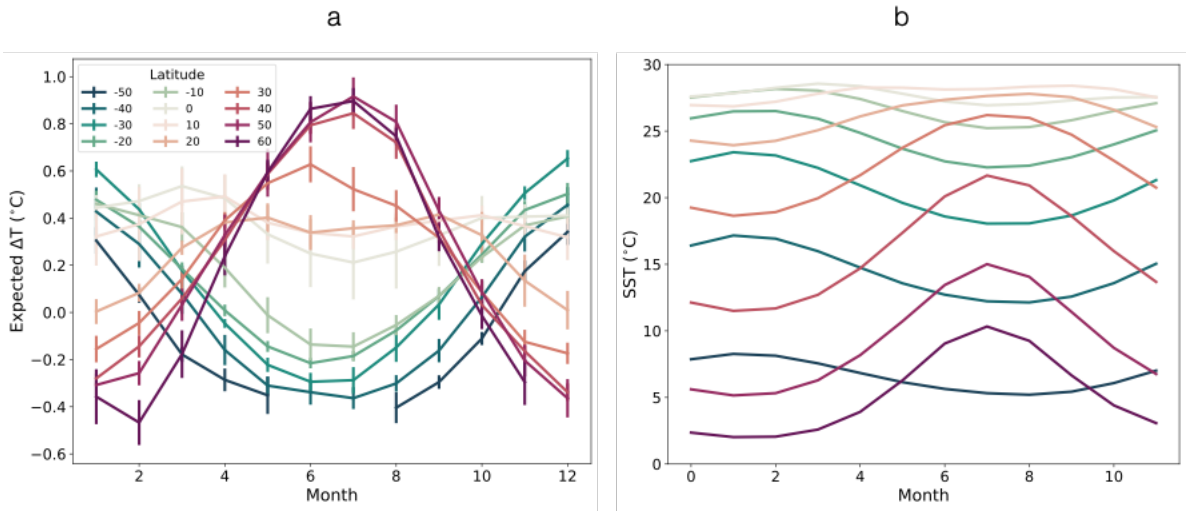

Figure 5.8. (a) Zonal means of expected temperature difference predicted by multiple linear regression calculated using climatological values of independent variables. Error bars indicate standard deviation calculated using monthly values for the years 2004-2020. (b) Seasonal SST. Both shown for 10 degree zonal boxes centered around the latitudes shown in the legend. Higher latitudes show more variation in predicted temperature differences between parameterizations as well as greater seasonal variation in SST.

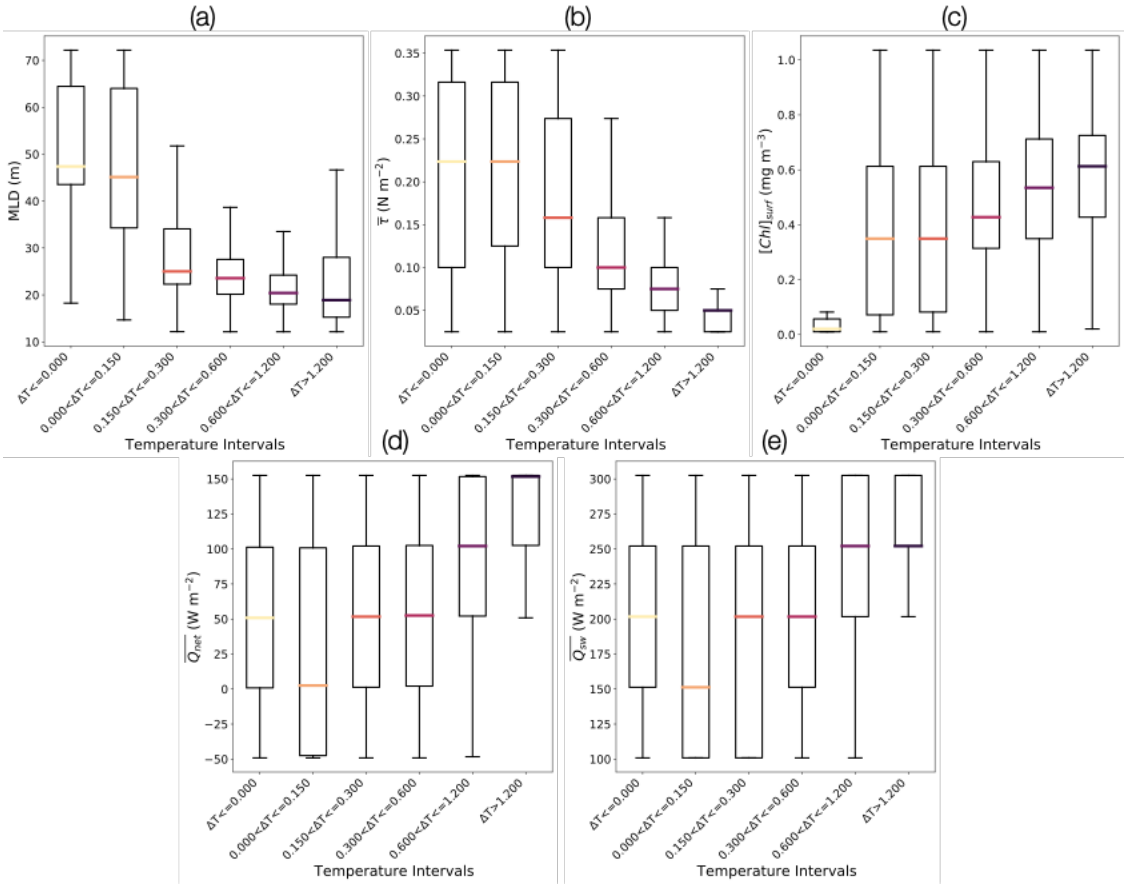

Figure 5.9. Boxplots showing relationship between final mixed layer temperature difference between simulations (x-axis) and model input variable (y-axis) for SS-C comparison (plotted as monthly temperature difference values). (a) Mixed layer depth. (b) Wind stress. (c) Surface chlorophyll concentration. (d) Net average heat flux. (e) Net average shortwave radiation.

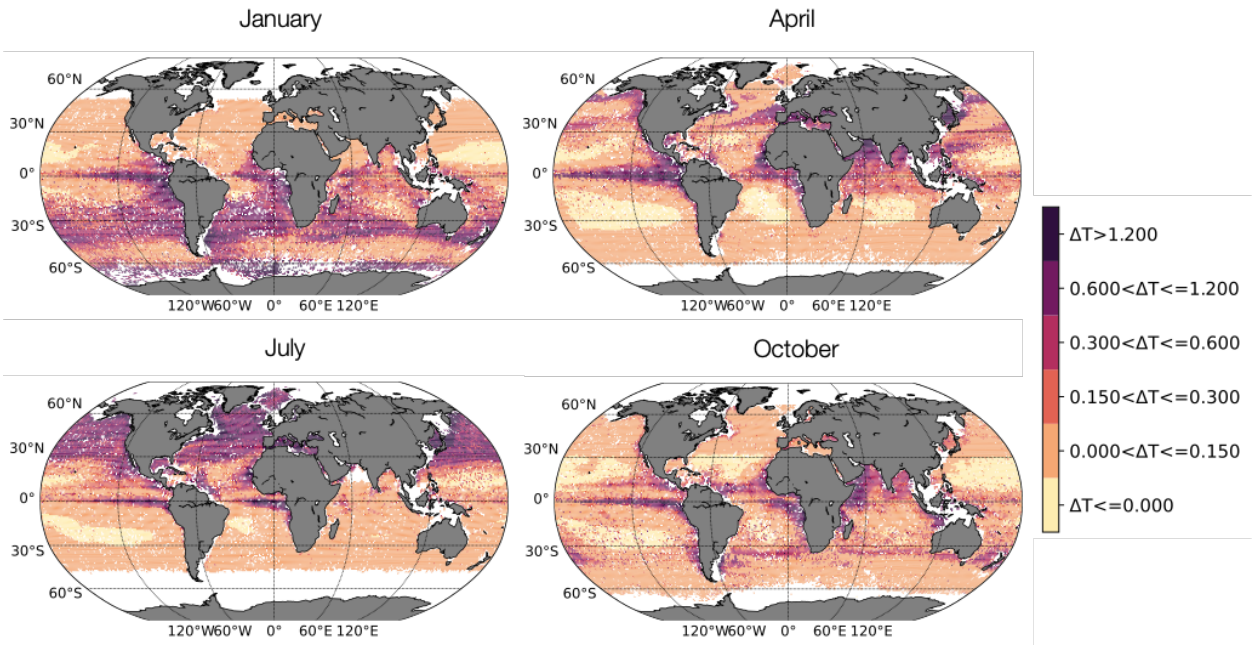

Figure 5.10. Monthly maps of expected temperature differences predicted from KNN clustering algorithm for SS-C comparison. Each color indicates the expected monthly difference in mixed layer temperature between the self-shading and control simulations, corresponding to the bins shown in Figure 9.

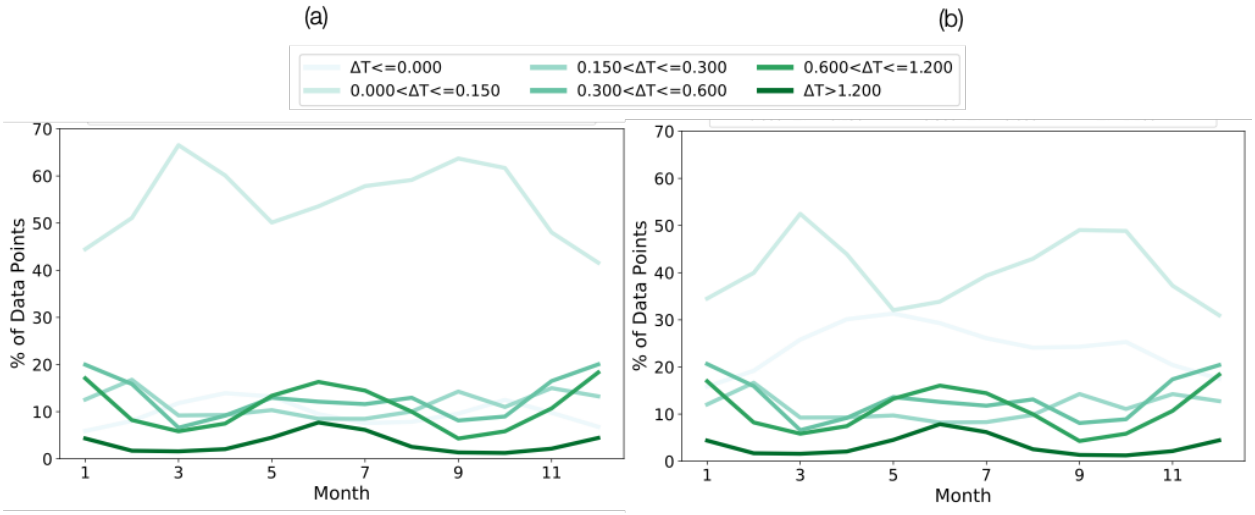

Figure 5.11. Prevalence of each temperature difference cluster by month for (a) SS-C comparison and (b) NO_SS-C comparison. Clusters representing $\Delta T > 0.15°C\ month^{-1}$ show very similar patterns in both plots.

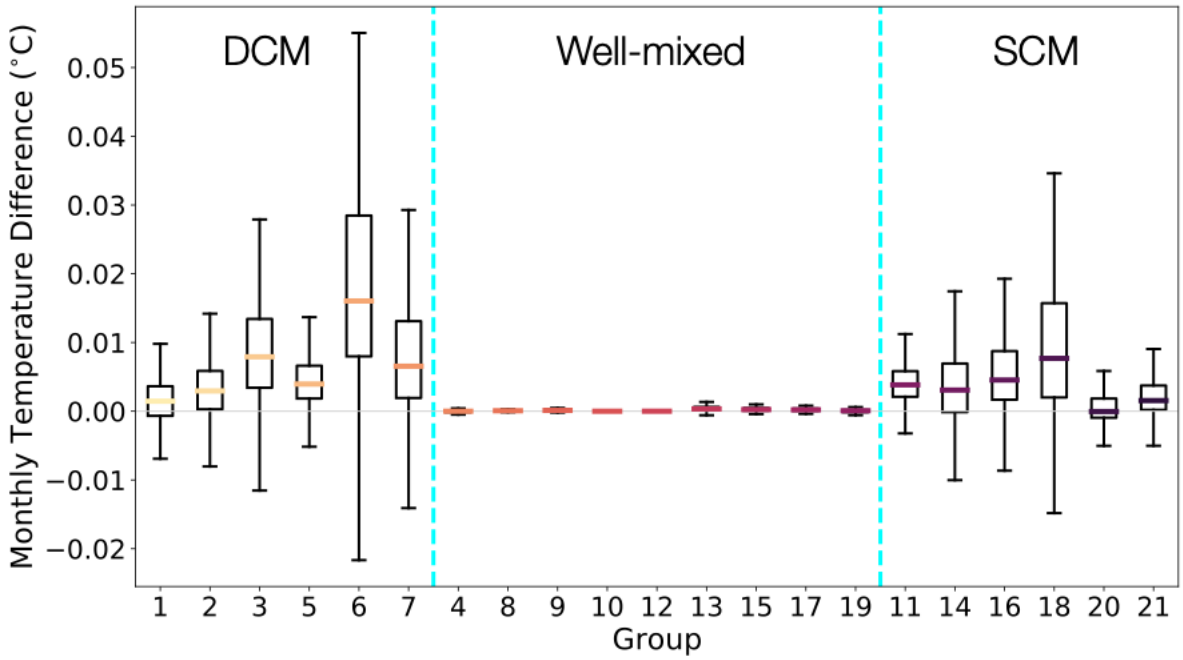

Figure 5.12. Difference in final mixed layer temperature for SS-NO_SS organized by group number and group type (based on work in Chapter 4; see Figure 2). DCM indicates profiles with very low surface chlorophyll and a deep chlorophyll maximum. Well-mixed indicates profiles that exhibit well-mixed near-surface chlorophyll features. SCM indicates profiles with a higher surface chlorophyll concentration as well as a subsurface chlorophyll maximum (typically <50m depth). These correspond to the profiles shown in Figure 2c.

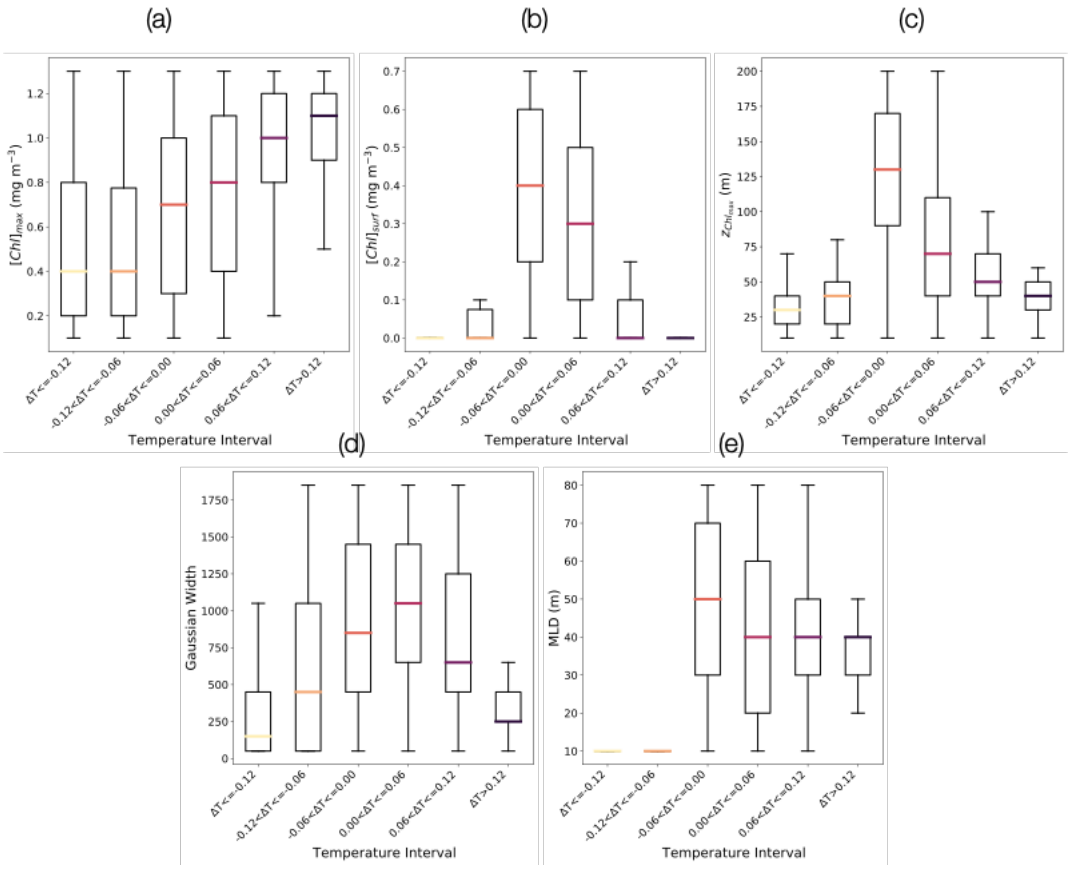

Figure 5.13. Box-plots illustrating relationships between different curve-fit parameters for the Gaussian+linear curve fit combination and expected temperature difference between the two methods of calculating heating due to shortwave radiation. Positive values indicate greater heating when calculated using self-shading (a) Amplitude of subsurface chlorophyll maximum. (b) Surface chlorophyll concentration. (c) Depth of subsurface chlorophyll maximum. (d) Width of gaussian peak. (d) Mixed layer depth over which temperature change is calculated.

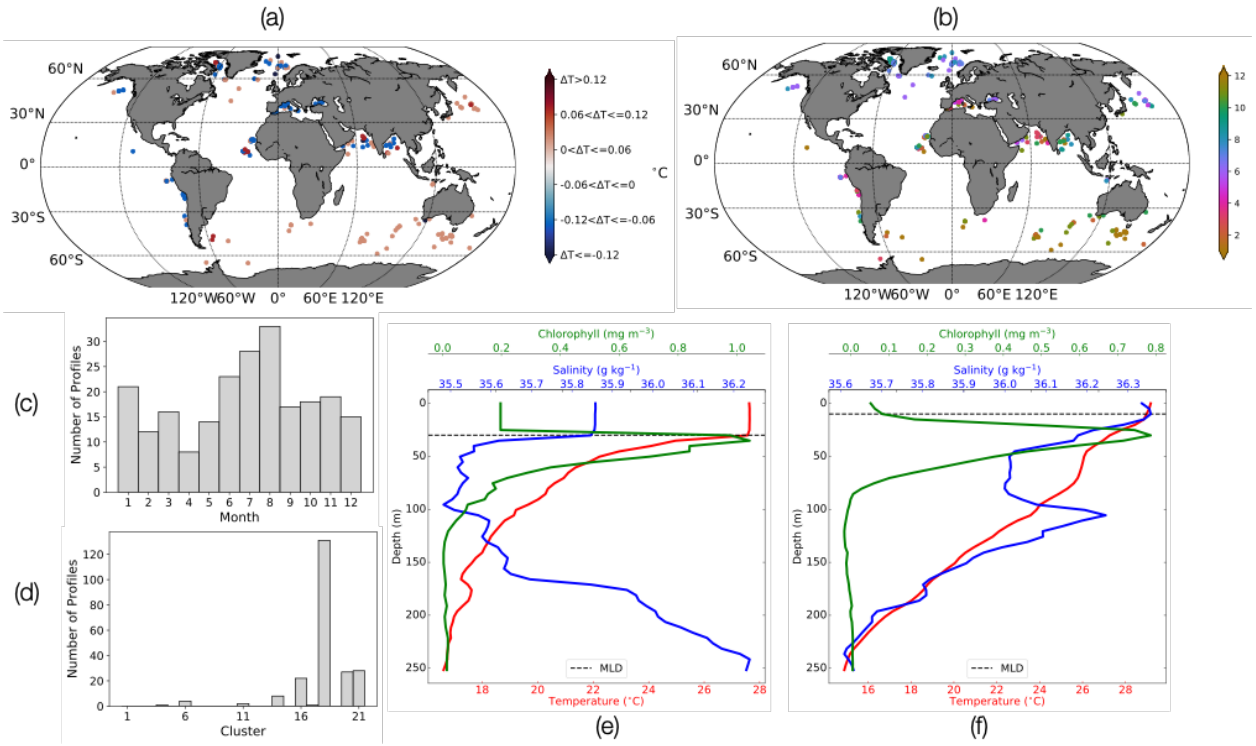

Figure 5.14. Predictions of Argo float profiles likely to exhibit maximum difference between SS and NO_SS parameterizations based on KNN clustering of constructed chlorophyll profiles. (a) Magnitude of mixed layer temperature difference predicted using Q_{sw} only (currently in °C day^{-1}). (b) Same profiles as in (a) but showing month in which those profiles were measured. (c) Distribution of profiles by month. (d) Distribution of profiles by cluster according to clusters shown in Figure 2. (e) Example profile associated with large positive effect from cluster 18 (17N, 57E, November). (f) Example profile with large negative effect from cluster 18 (13N, 65E, October).

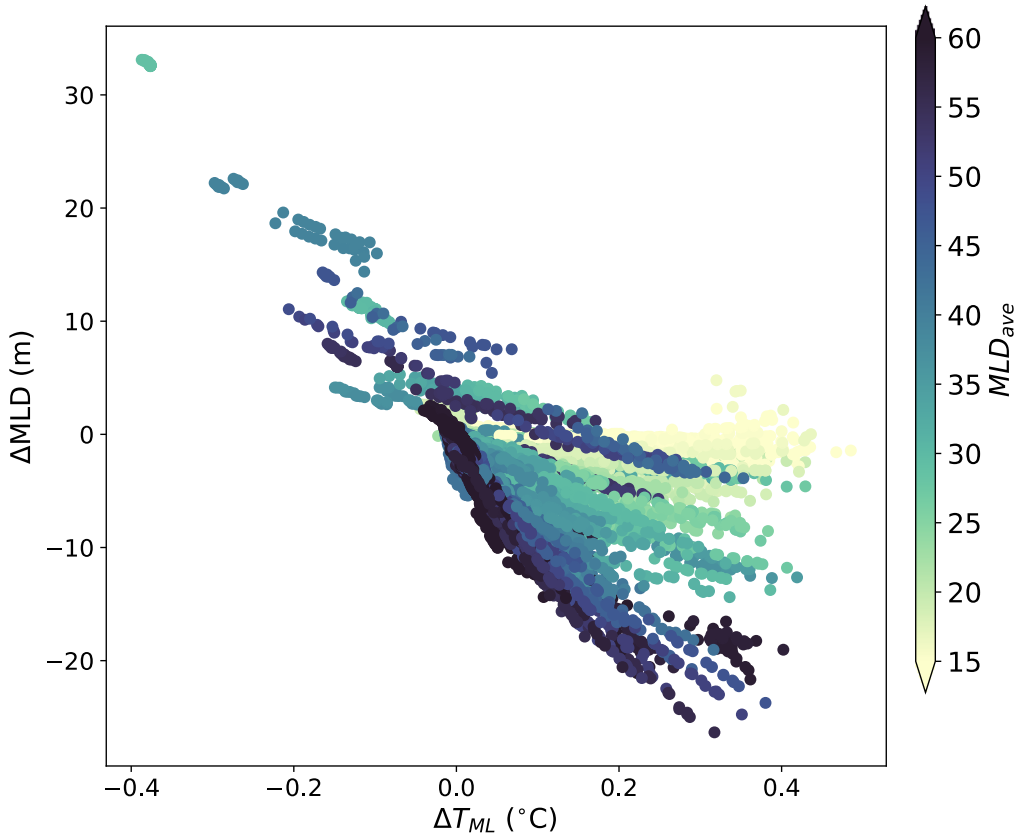

Figure 5.15. ΔT_{ML}(SS-C) plotted versus ΔMLD (SS-C). Shading of dots indicates average mixed layer depth of control simulation. ΔMLD is calculated as the difference in mixed layer depths throughout the entire simulation.

Variable	Minimum	Maximum	Increment
Q_{sw} (W m^{-2}) Shortwave Radiation	100	300	50
Q_{net} (W m^{-2}) Net Heat Flux	-50	150	50
Wind Stress (N m^{-2})	0.025	0.35	0.025 for τ <=0.125 0.15 <= τ <=0.3
Mixed Layer Depth (m)	15	60	15, 20, 40, 60

Table 5-1. Model forcing scenarios, showing range of heat flux, wind stress, and mixed layer depth values. Q_{sw} was distributed over a 12-hour daytime period.

Variable	R^2	*RSS*
Q_{sw}	0.084	208
Q_{net}	0.094	205
MLD_{ave}	0.34	149
chl_{surf}	0.11	201
τ_{ave}	0.16	190

Table 5-2. R^2 and residual sum of squares (RSS) values for the individual components of the multiple linear regression with ΔT_{ML} (SS-C).

Intercept	**0.0477**
$\boldsymbol{Q_{sw}}$	0.0174
$\boldsymbol{Q_{net}}$	0.0130
$\boldsymbol{MLD_{ave}}$	0.0282
$\boldsymbol{chl_{surf}}$	0.0207
$\boldsymbol{\tau_{ave}}$	-0.0143
$\boldsymbol{R^2}$	0.621

Table 5-3. Intercept and coefficients for multiple linear regression with cumulative R^2 value.

	Gaussian + Line			Gaussian + Sigmoid		
	Min	**Max**	**Increment**	**Min**	**Max**	**Increment**
$[Chl]_{max}$ (mg m^{-3})	0.1	1.3	0.1	0.1	3	0.25
z_{max} (m)	10	200	10	10	90	20
dz (m)	50	1950	200	50	1950	200
$[Chl]_{surf}$ (mg m^{-3})	0	0.65	0.1	0.1	2	0.2
$z_{1/2}$	-	-	-	10	150	20
MLD	10	80	10	10	80	10

Table 5-4. Curve fit parameters for Equations (7) (Gaussian + Line) and (8) (Gaussian + Sigmoid) used to construct artificial profiles. Each parameter was varied independently.

Region	Number of Profiles	Total Profiles	Months
Baffin Bay	34	1024	7, 8, 9
North Atlantic	17	812	6, 7, 8
Arabian Sea	11	912	
Bay of Bengal	11	912	
Southern Ocean (Indian Sector)	42	6302	10, 11, 12, 1, 2

Table 5-5. Regions with significant numbers of high impact profiles from Argo for ΔT_{ML}(SS-NO_SS) and their observed seasons.

Chapter 6. CONCLUSION

One of the core themes running through this dissertation is the question of how we compare the effects of different phenomena in ways that are consistent within a single study and also consistent across multiple studies. Differences in datasets, models, or methods can yield different results for scenarios we might otherwise expect to be the same. Although there is benefit to using different approaches in contexts where they are warranted, we miss many opportunities to meaningfully discuss patterns and trends when every study on a topic uses a different approach to address the same phenomenon. Barrier layer studies suffer from a huge range in definitions for what constitutes the relevant depths. Are the differences across studies due to these differences in methods? The different datasets available when each study was published? Genuine physical effects? And this challenge is not limited to observational studies. Modeling studies exploring the differences in shortwave radiation absorption also show a range of outcomes, but given the range in data sources used for chlorophyll data or K_d, parameterizations of shortwave radiation used, the model resolution, and biases that already exist in the models, it is hard to come to a unifying conclusion in many cases about the magnitude of effects like self-shading and how important they are for upper ocean heat balance. Some studies, such as Manizza et al (2005), even note that their results for the change in sea surface temperature in some regions change sign if a different bio-optical model is used. This makes attribution to the phenomenon under investigation quite difficult. This issue has been discussed in other fields as well, such as by Palevsky and Doney (2018), who note that differences in results about carbon export to the deep ocean may be as much due to methodological differences as anything else. As a result, in this work we have tried to highlight approaches where we can resolve some of the inconsistencies in methods.

Consistency in how we identify barrier layers is essential for understanding the question of how barrier layers impact local mixed layer heating, biological productivity, and atmospheric feedback. The approach outlined in Chapter 2 here produces results that show some similarity with other studies in the Arabian Sea, but does not require a choice of temperature threshold. This is particularly useful in regions like the Arabian Sea where the stratification below the mixed layer changes quite significantly for different seasons, and using a single threshold value may end up selecting very different sections of the water column. The approach defined here performs well in selecting for the two essential depths of a barrier layer across seasons, and therefore may also be useful in expanding the study of barrier layers to other questions, such as their impact on biological productivity. In addition to the observation that they separate the mixed layer from the cooler waters at the top of the thermocline, and therefore play an important role in the evolution of mixed layer temperature, as discussed in Chapter 3, they may also separate the mixed layer from mixing with deeper water and therefore exert some control on the distribution of nutrients and phytoplankton in the water column (**Figure 6.1**) as well as the concentration of dissolved oxygen.

If we apply our spice-based barrier layer identification method to the global ocean, we are successfully able to identify the dominant barrier layer regions, such as the Bay of Bengal, Western Pacific Warm Pool, and Amazon outflow region (**Figure 6.2**). We are also able to identify several less-well studied regions, such as the eastern tropical north Pacific (Katsura et al, 2020) and high latitude regions, particularly in the north Pacific (Kara et al, 2000). The applicability of this method to the global ocean opens the door for studies identifying the impact of barrier layers throughout the ocean without requiring regionally or seasonally specific threshold values. This is of great interest given the potential impacts discussed here in Chapter 3,

but also considering the possible feedback to tropical storms (Balaguru et al, 2012). Tropical storms appear frequently in several of the core regions shown in **Figure 6.2**, such as the Bay of Bengal (Chacko, 2017; Navaneeth et al, 2019) and Amazon outflow region in the Atlantic Ocean (Balaguru et al, 2012, Reul et al, 2021), but have also been observed more recently in the Arabian Sea (Murakami et al, 2017). Understanding both the changes in sea surface temperature and barrier layer intensity and distribution may lead to a better understanding of tropical storm intensity as the climate warms.

In Chapter 4 I evaluated global patterns in the vertical distribution of chlorophyll using data derived from biogeochemical profiling floats. These floats provide data in regions and seasons that have been traditionally not well covered, especially regions such as the Southern Ocean. They also provided data with a temporal resolution and longevity that contributes to a comprehensive understanding of seasonal patterns of chlorophyll distribution throughout the water column. Performing a clustering analysis allows us to understand seasonal patterns not purely in terms of surface chlorophyll concentrations but also in terms of how the chlorophyll is distributed and its relation to physical factors. Broad subgroups of clusters showed distinct geographic separation, furthering the idea that chlorophyll profiles evolve for different reasons in different regions. As this dataset expands, this method may be viable for future studies on biogeography and evaluating long-term trends in the behavior of various regions.

The representative profiles developed in Chapter 4 lay the groundwork to better understand how chlorophyll in the water column influences mixed layer heating, and how this varies under different atmospheric forcing and ocean structures. They support the modeling study of Chapter 5 by providing a reasonable, observation-based estimate for a range of possible chlorophyll distributions in the ocean. Unlike previous modeling studies, this means

parameterizing shortwave radiation using chlorophyll concentrations as they are observed in the ocean, both for the parameterization that relies on the vertical distribution of chlorophyll as well as the one based on surface average chlorophyll values, as observed from satellites. Previous studies have used different sources of data for these two chlorophyll parameterizations (e.g. Manizza et al, 2005), and therefore it is not always clear which results arise from the chlorophyll concentration itself as opposed to the choice of parameterization. In Chapter 5, I showed that choice of parameterization leads to significant effects, and that the magnitude of these effects depends not only on the concentration of chlorophyll, as expected, but also on the mixed layer depth, wind stress, and heat fluxes. Some regions experience significant effects year round, such as the eastern tropical Pacific, whereas higher latitude regions are largely only effected during times of high insolation and increased chlorophyll concentration. This is valuable insight for observational studies, where the choice of how to parameterize shortwave radiation can have significant impacts on the final results from a mixed layer heat budget. It may also prove useful for comprehensive modeling studies that require a trade-off between accuracy and computational efficiency. As the dataset of chlorophyll profiles expands, we also expect them to provide a valuable resource for model validation, such as that discussed by Wang et al (2020) for regional studies.

In the work outlined in these chapters, I explored the influence of salinity stratification and the influence of chlorophyll as two distinct phenomena that impact the absorption of shortwave radiation. However, some regions of the ocean exhibit both the strong salinity stratification discussed with barrier layers as well as chlorophyll profiles that are most likely to require a more detailed parameterization for shortwave radiation. These include the northern Indian Ocean, coastal regions near river outflow, and potentially some high-latitude regions

where significant meltwater exists. This prompts the question of what we expect the combined effect to be. Do shallower mixed layers observed with barrier layers allow more radiation to penetrate below the mixed layer? Do the subsurface chlorophyll features observed in the Bay of Bengal alter the predicted impact of barrier layers or vice versa? Lotliker et al (2016) explored some of these questions using cruise data in the Bay of Bengal, but to date these are otherwise highly understudied questions. As observational coverage of these regions expands, increasing the availability of biogeochemical data and ultimately providing multi-decadal timeseries of both physical and biogeochemical data, we will be better equipped to clarify the effects we can expect from these complex systems.

Furthermore, phenomena that are heavily influenced by freshwater fluxes and biological activity may see changes in future decades, as we expect both the hydrologic cycle and biological productivity in the ocean to change (Trenberth, 1999; Cabré et al, 2014; Gregg and Rousseaux, 2014; Moore et al, 2018). Regions that currently experience strong barrier layers may see even stronger salinity stratification if the hydrologic cycle intensifies and the conditions producing the barrier layers, such as strong riverine outflow, become more extreme. What may be an intermittent or occasional phenomenon in some regions may become a more significant feature, altering the interaction of the upper ocean with the atmosphere and potentially impacting biological activity as the physical structure of the ocean changes.

Finally, the high latitude regions highlighted in **Figure 6.2** and Figure 5.14 (Chapter 5) are likely to experience increasing effects of freshwater input, as glacial melt from Greenland becomes more pronounced (Luo et al, 2016). This is also a region where subsurface chlorophyll features are not only common, but are also not well represented by satellite measurements (Arrigo et al, 2011) and represents one of the regions where we see the greatest difference

between different parameterizations that incorporate chlorophyll measurements. Tracking the seasonal patterns and long term trends in the distribution of freshwater along with the collection of biogeochemical measurements in the region may provide insight into how changes in the physical structure of the ocean in high latitude regions is impacting biological productivity. We can ask questions about how freshwater content is changing, how it is distributed, and whether or not the peak freshwater content aligns with productive seasons. We can then turn to asking questions about where biomass is located and how this might affect local heating, looking for combined impacts from changes to the physics and biology of these regions. The observational methods employed here can complement modeling studies that attempt to understand future changes in the ocean.

References

Arrigo, K.R., Matrai, P.A., & van Dijken, G.L. (2011) Primary productivity in the Arctic Ocean: Impacts of complex optical properties and subsurface chlorophyll maxima on large-scale estimates. Journal of Geophysical Research, 116, C11022, doi:10.1029/2011JC007273

Cabré, A., Marinov, I., & Leung, S. (2014) Consistent global responses of marine ecosystems to future climate change across the IPCC AR5 earth system models. Climate Dynamics, 45, 1253-1280, doi:10.1007/s00382-014-2374-3.

Chacko, N. (2017). Chlorophyll bloom in response to tropical cyclone Hudhud in the Bay of Bengal: Bio-Argo subsurface observations. Deep-Sea Research Part 1, 124, 66-72.

Gregg, W.W., & Rousseaux, C.S. (2014). Decadal trends in global pelagic ocean chlorophyll: A new assessment integrating multiple satellites, in situ data, and models. Journal of Geophysical Research: Oceans, 119, 5921-5933, doi:10.1002/2014JC010158.

Kara, A.B., Rochford, P.A., & Hurlburt, H.E. (2000). An optimal definition for ocean mixed layer depth. Journal of Geophysical Research, 105, C7, 16803-16821.

Katsura, S. & Sprintall, J. (2020). Seasonality and Formation of Barrier Layers And Associated Temperature Inversions in the Eastern Tropical North Pacific. Journal of Physical Oceanography, 50, 791-808, DOI: 10.1175/JPO-D-19-0194.1.

Lotliker, A.A., Omand, M.M., Lucas, A.J., Laney, S.R., Mahadevan, A., & Ravichandran, M. (2016). Penetrative radiative flux in the Bay of Bengal. Oceanography 29(2):214–221, http://dx.doi.org/10.5670/oceanog.2016.53.

Luo, H., Castelao, R.M., Rennermalm, A.K., Tedesco, M., Bracco, A., Yager, P., & Mote, T. (2016). Oceanic transport of surface meltwater from the southern Greenland ice sheet. Nature Geoscience, 9, 528-532.

Manizza, M., Le Quéré,C., Watson, A.J., & Buitenhuis, E.T. (2005). Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model. Geophysical Research Letters, 32, L05603, doi:10.1029/2004GL020778.

Moore, J.K., Fu, W., Primeau, F., Britten, G.L., Lindsay, K., Long, M., Doney, S.C., Mahowald, N., Hoffman, F., & Randerson, J.T. (2018). Sustained climate warming drives declining marine biological productivity. Science, 359, 1139-1143

Murakami, H., Vecchi, G., & Underwood, S. (2017). Increasing frequency of extremely severe cyclonic storms over the Arabian Sea. Nature Climate Change, 7, 885-889, https://doi.org/10.1038/s41558-017-0008-6

Navaneeth, K.N., Martin, M.V., Jossia Joseph, K., & Venkatesan, R. (2019). Contrasting the upper ocean response to two intense cyclones in the Bay of Bengal. Deep-Sea Research Part 1, 147, 65-78.

Palevsky, H. I., & Doney, S. C. (2018). How choice of depth horizon influences the estimated spatial patterns and global magnitude of ocean carbon export flux. Geophysical Research Letters, 45, 4171–4179. https://doi.org/10.1029/2017GL076498

Reul, N., Chapron, B., Grodsky, S.A., Guimbard, S., Kudryavtsev, V., Foltz, G. R., & Balaguru, K. (2021). Satellite observations of the sea surface salinity response to tropical cyclones. Geophysical Research Letters, 48, e2020GL091478. https://doi. org/10.1029/2020GL091478

Trenberth, K.E. (1999). Conceptual Framework for Changes of Extremes of the Hydrological Cycle with Climate Change. Climatic Change, 42**,** 327–339, https://doi-org.offcampus.lib.washington.edu/10.1023/A:1005488920935

Wang, B., Fennel, K., Yu, L., & Gordon, C. (2020). Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico. Biogeosciences, 17, 4059-4074. Doi: 10.5194/bg-17-4059-2020

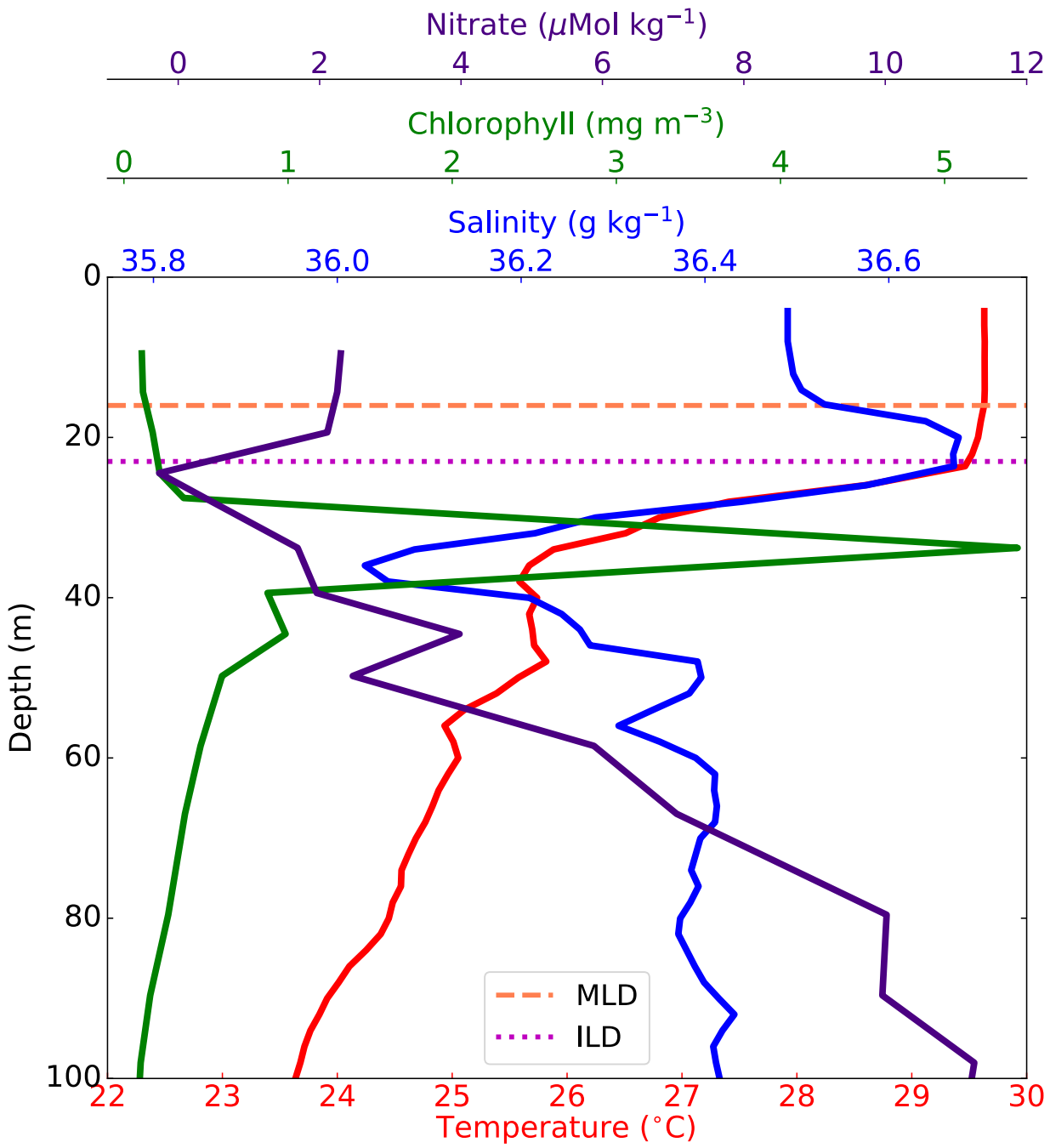

Figure 6.1. Profile showing a barrier layer, chlorophyll concentration, and nutrient distribution in the northeastern Arabian Sea. MLD=mixed layer depth; ILD = isothermal layer depth. Barrier layer is the region in between MLD and ILD. Despite significant chlorophyll peak ~35m, surface nutrients are elevated and lowest nutrient concentration is observed immediately below ILD. Barrier layer effects in this region are poorly studied.

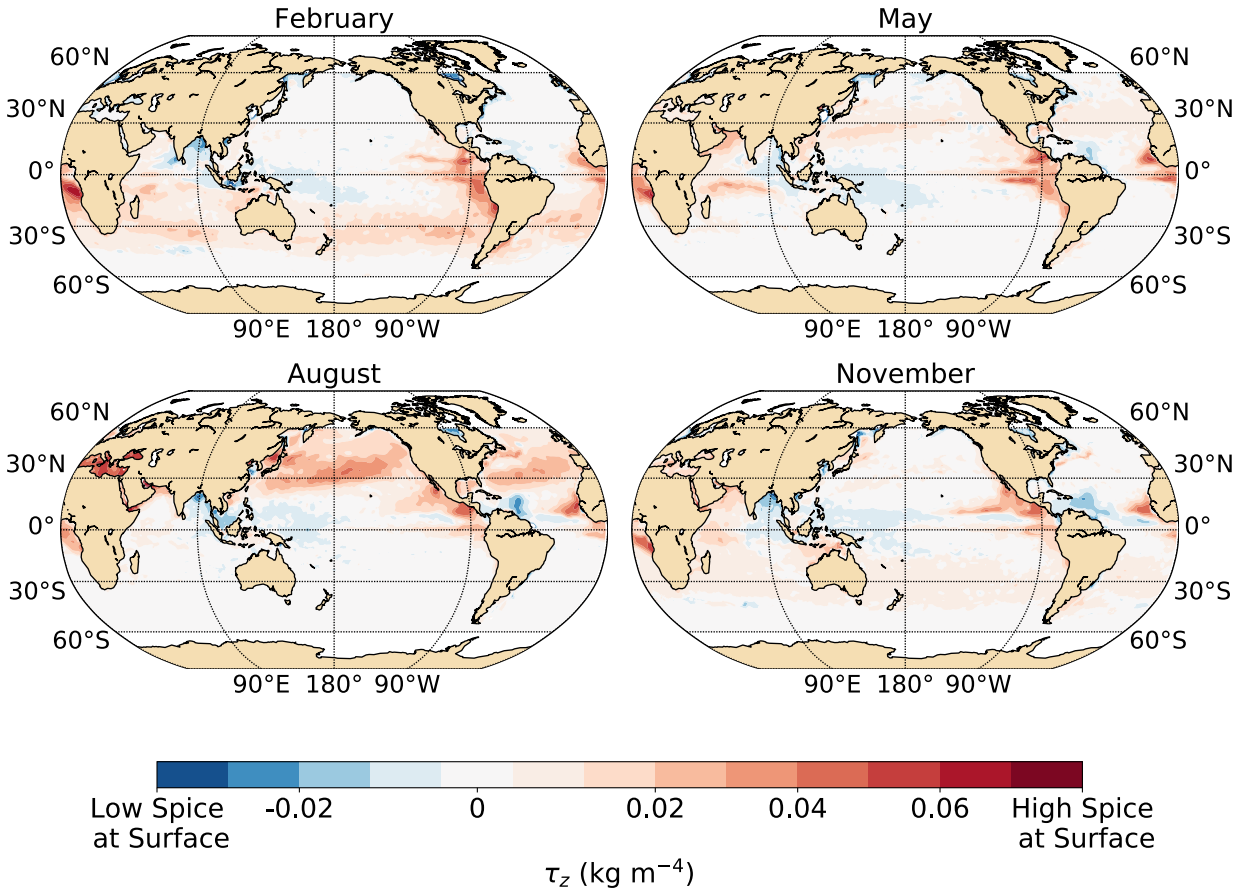

Figure 6.2. Diapycnal spiciness curvature for the top 50m of the global ocean. Blue indicates regions where we expect to find barrier layers.

APPENDIX A: CHAPTER 4 SUPPLEMENTAL MATERIALS

Coding details for Chapter 4

All code for the methods discussed here is available publicly at github.com/rechols/GMM_chlorophyll.

Principal Component Analysis (PCA) and Gaussian Mixture Model (GMM) performed using the Python *scikit-learn* toolbox (Scikitlearn.org; Pedregosa et al, 2011)

Day/night delineation performed using tools from the Skyfield module for Python (https://rhodesmill.org/skyfield/). Skyfield has tools that facilitate calculating local sunrise and sunset using latitude, longitude, and local time. For high-latitude locations, it also allows determination of whether it is always day or night for a given input.

Curve-fitting was performed using the *lmfit* module for Python (lmfit.github.io), which has two advantages for the current purposes: (1) it allows simple construction of composite curves (e.g. a Gaussian superimposed on a Sigmoid) and (2) it returns fit statistics as part of the model result, which allows for streamlined comparison of different fits.

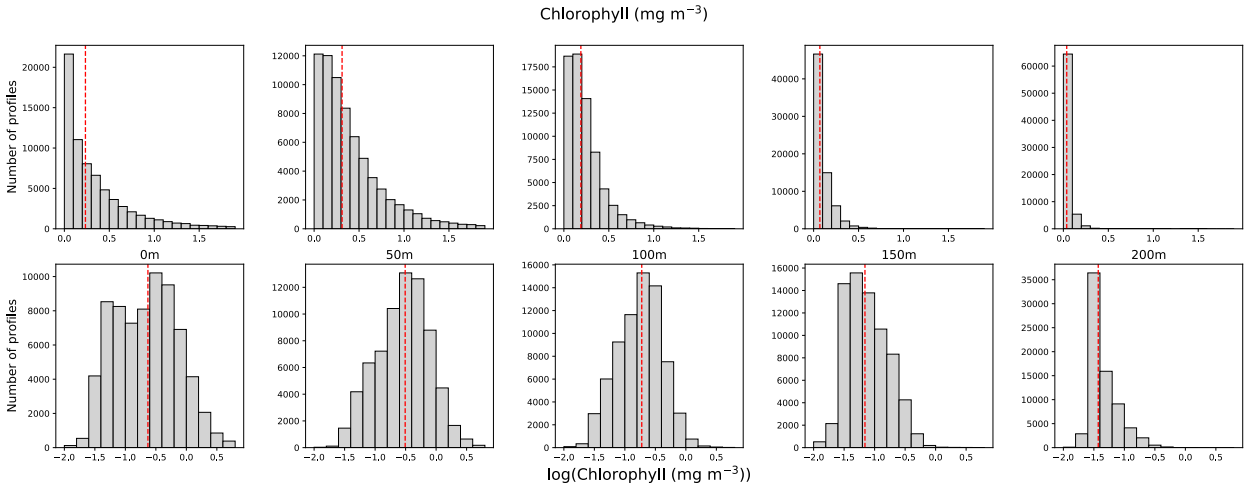

Figure S1. (a)-(e) Raw concentration of chlorophyll at (a) 0m, (b) 50m, (c) 100m, (d) 150m, and (e) 200m. (f)-(i) as for (a)-(e) but lognormalized.

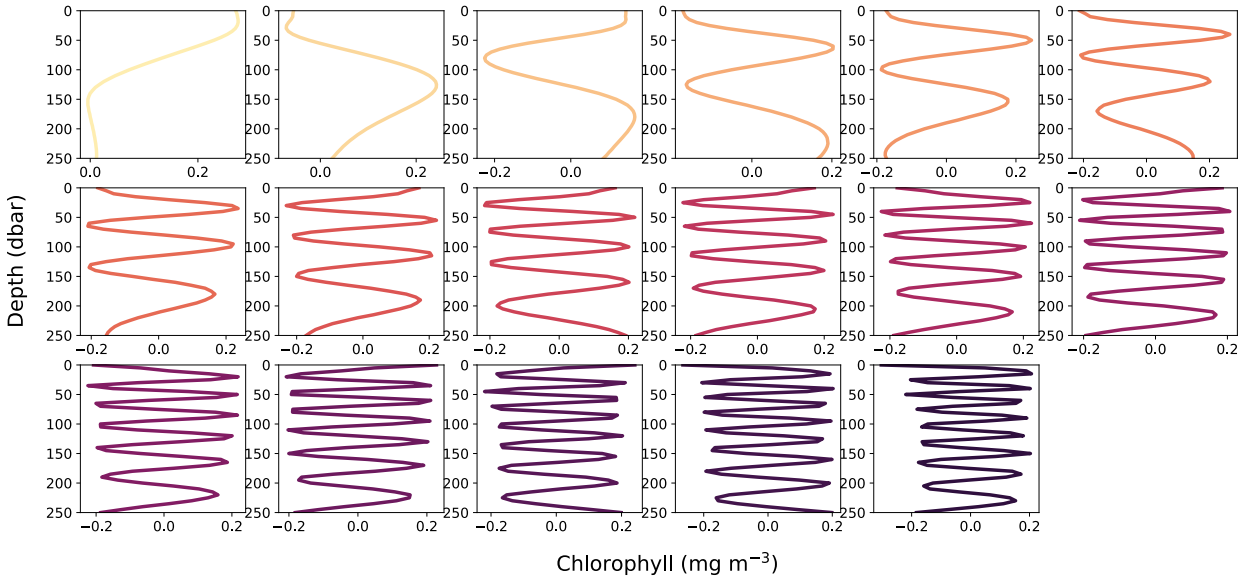

Figure S2. First 17 principal components.

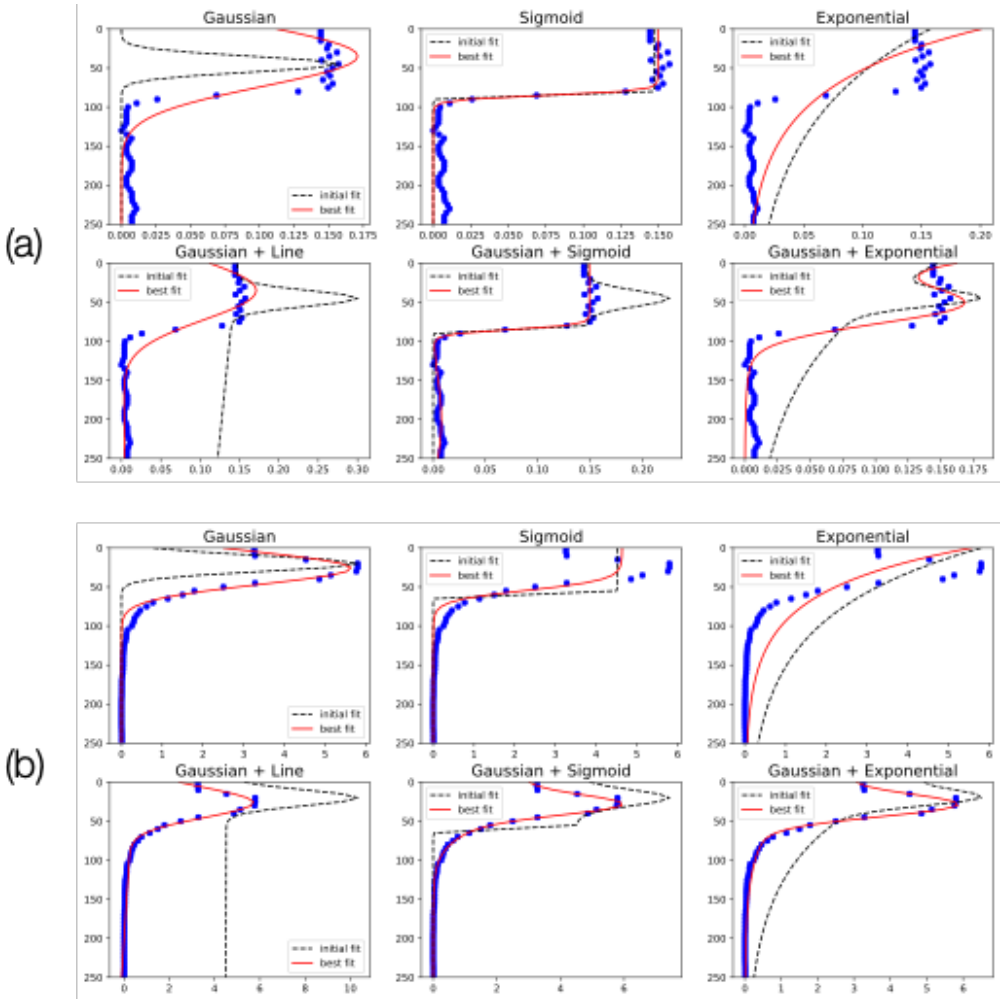

Figure S3. Two example profiles with the initial and final curve fits in each category. Of particular interest here is that the Gaussian+Sigmoid fit approximates both profiles very well, despite significant differences in overall profile characteristics.

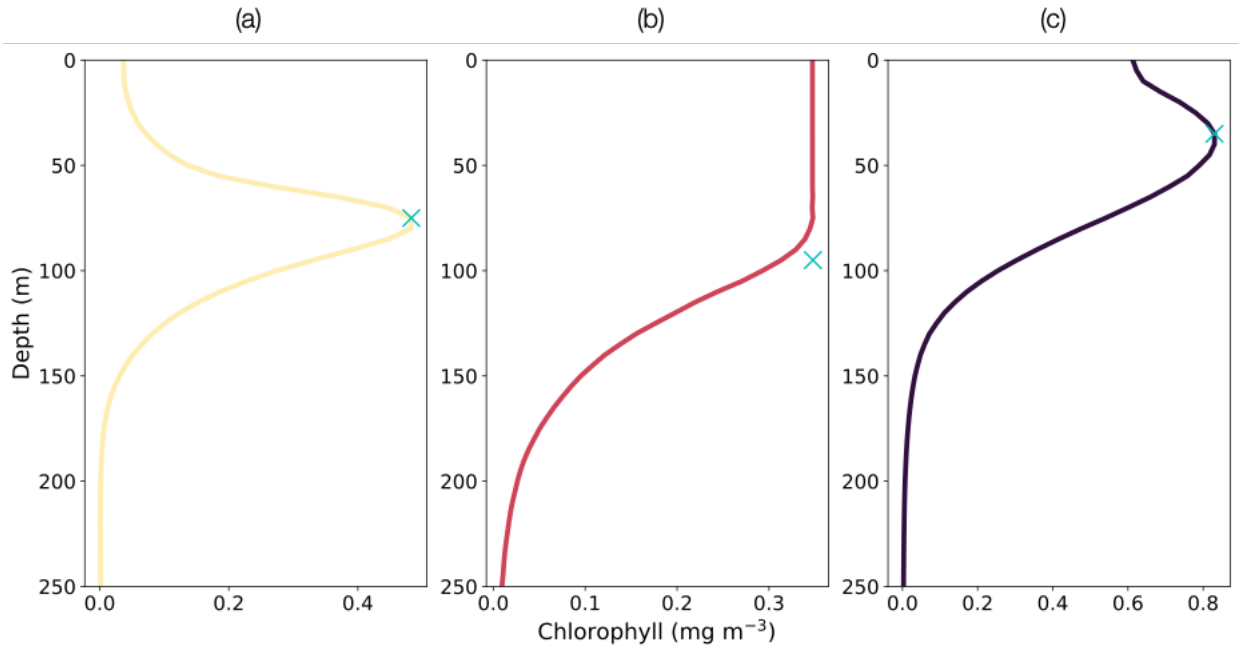

Figure S4. Sample profile shapes. Blue X on each profile indicates the depth used when calculating the chlorophyll depth (a and c) or the end of the well-mixed layer.

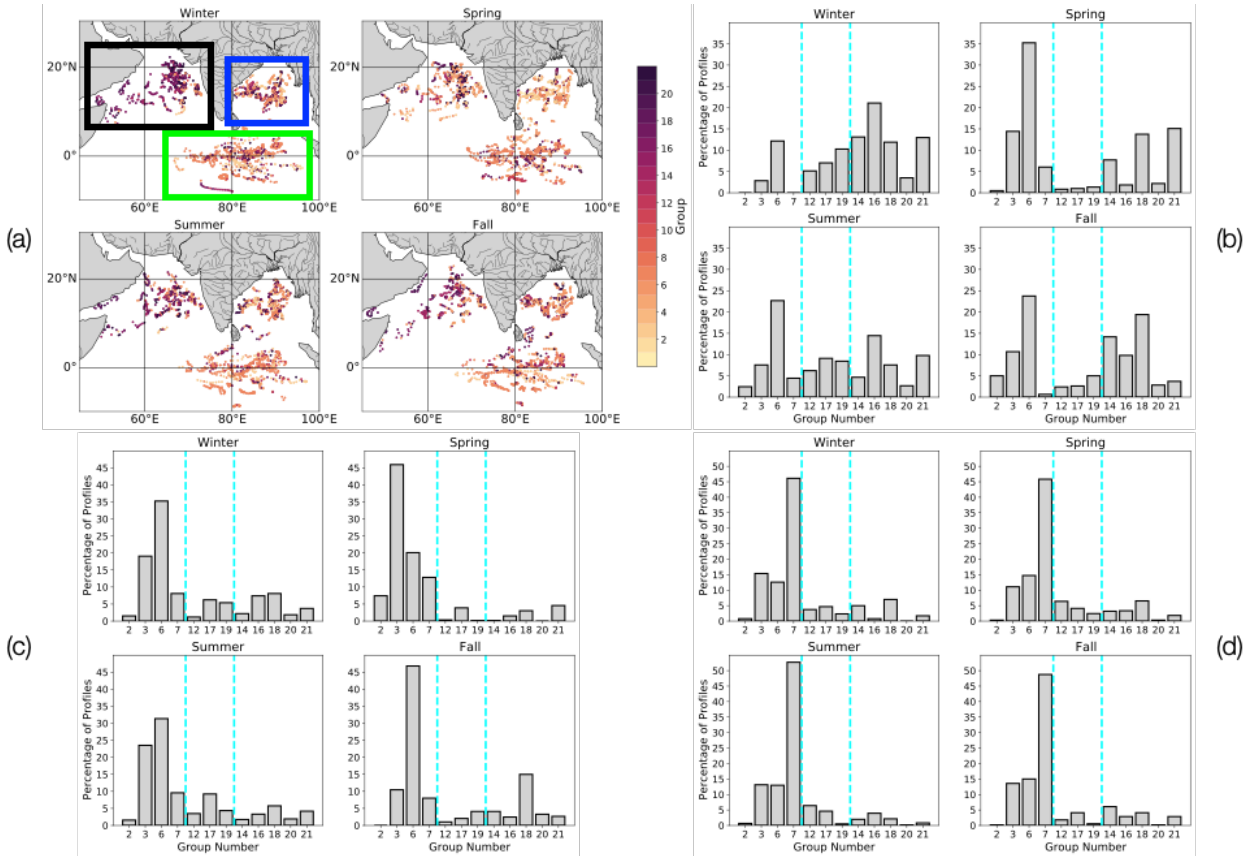

Figure S5. Seasonal patterns of cluster prevalence in the northern Indian Ocean. (a) Seasonal map showing dominant clusters. Black box indicates Arabian Sea, blue box indicates Bay of Bengal, green box indicates equatorial region. (b) Seasonal histogram for the Arabian Sea only. (c) Same as (b) but Bay of Bengal. (d) Same as (b) but for equatorial region. Dotted cyan lines in (b)-(d) indicate delineation between the three subgroups (from left to right: deep SCM, well mixed, and shallow SCM).

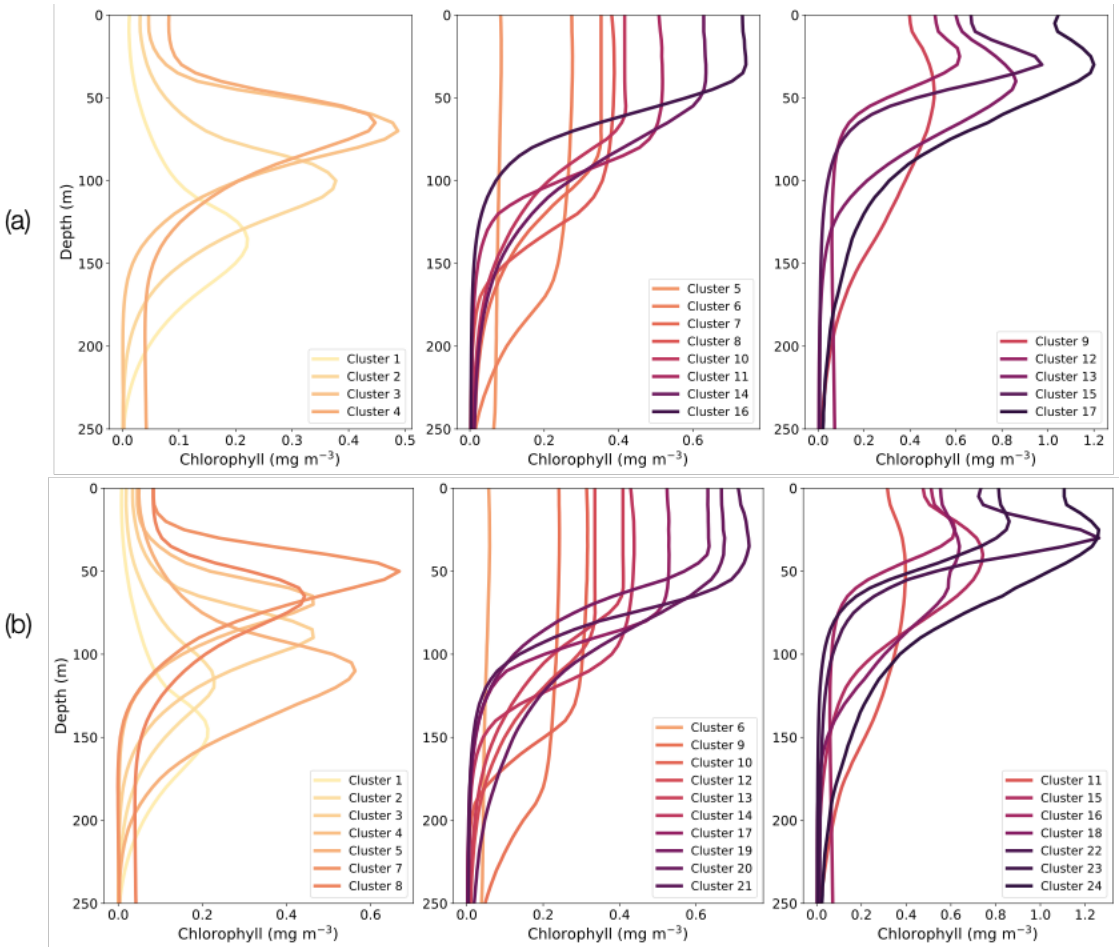

Figure S6. Weighted average profiles for (a) N=17 and (b) N=24 clusters, divided into the same subgroups as discussed in the text.

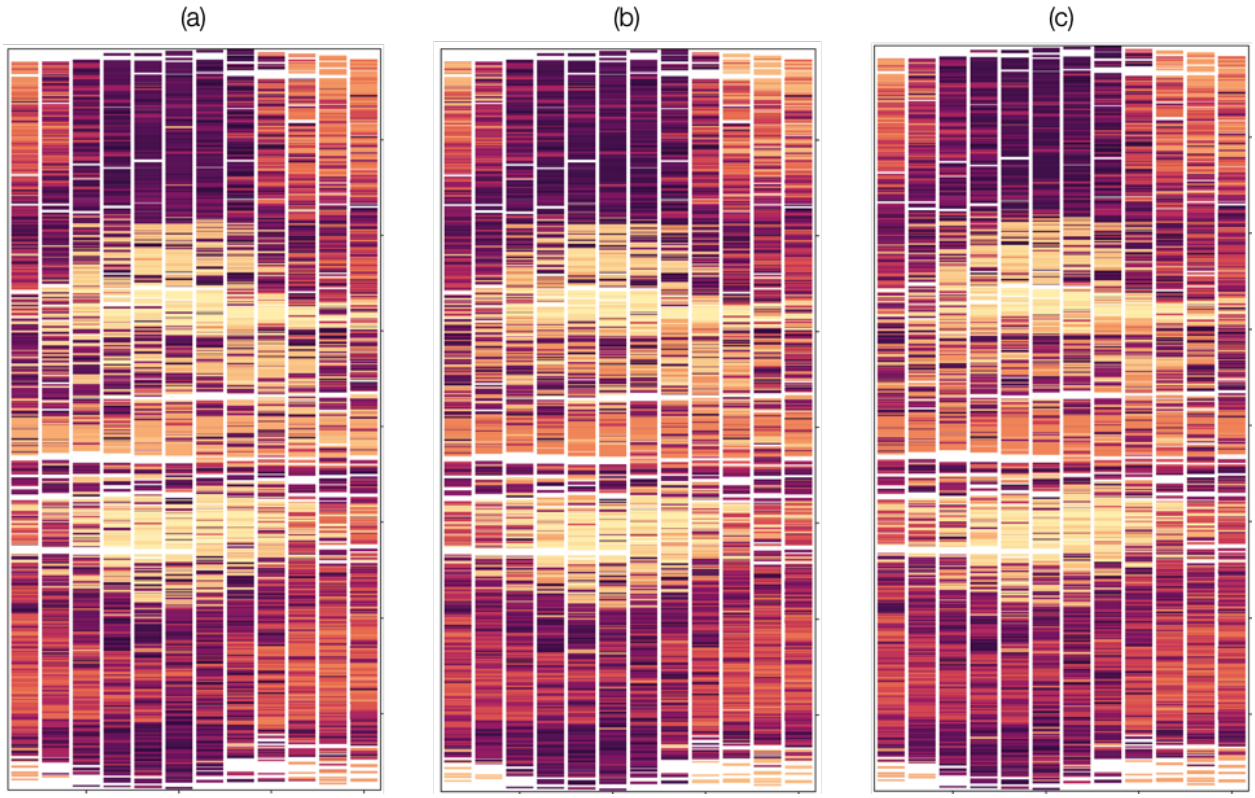

Figure S7. Latitude vs. season plots for (a) N=17; (b) N=21 (test case presented in paper); and (c) N=24 clusters. Shading as with Figure 7, where light yellow indicates the cluster with the lowest average surface chlorophyll and dark purple indicates the cluster with the highest average surface chlorophyll.

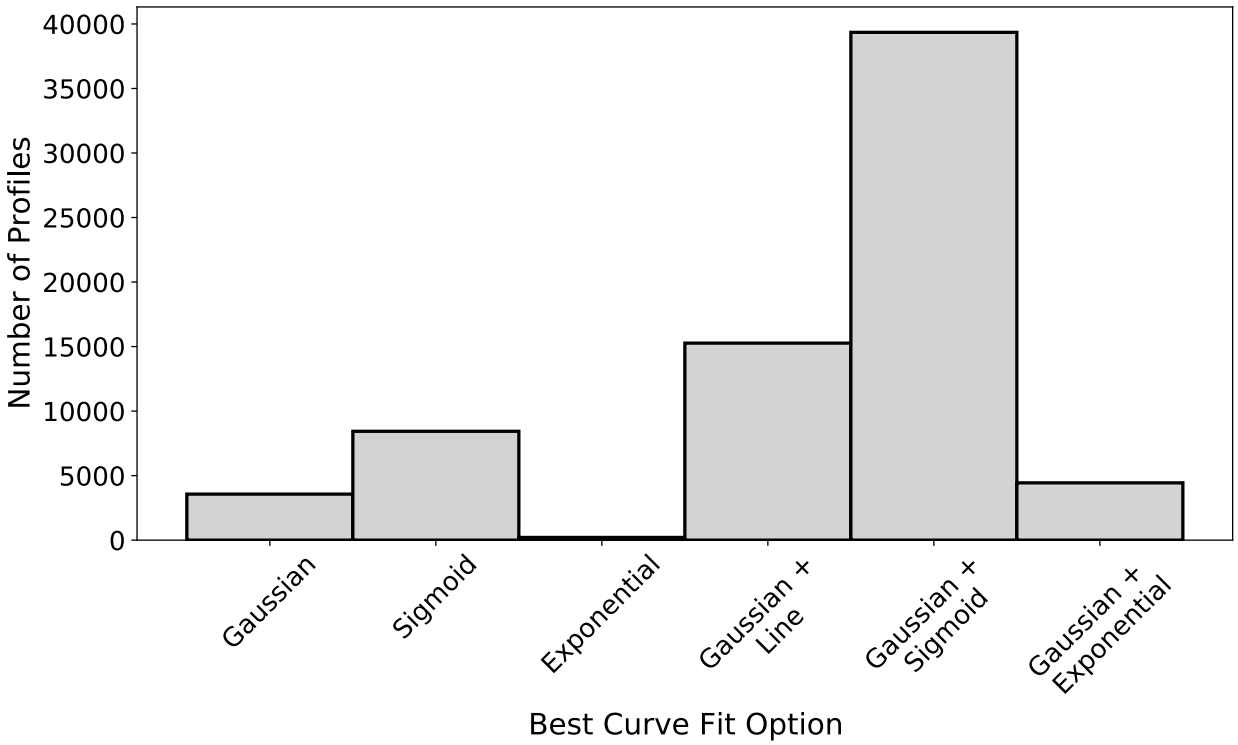

Figure S8. Distribution of most representative curve fit using minimum AIC score to determine best fit.

APPENDIX B: CHAPTER 5 SUPPLEMENTAL MATERIALS

Coding details for Chapter 5

Development of the MLR is done in Python using the Ordinary Least Squares method from the *statsmodels* package.

We perform a supervised K-nearest neighbors (KNN) classification using *KNeighborsClassifier* from *SciKit Learn* (scikit-learn.org) for Python (Pedregosa et al, 2011)

For developing the curve-fits, we use the *lmfit* module for Python (lmfit.github.io)

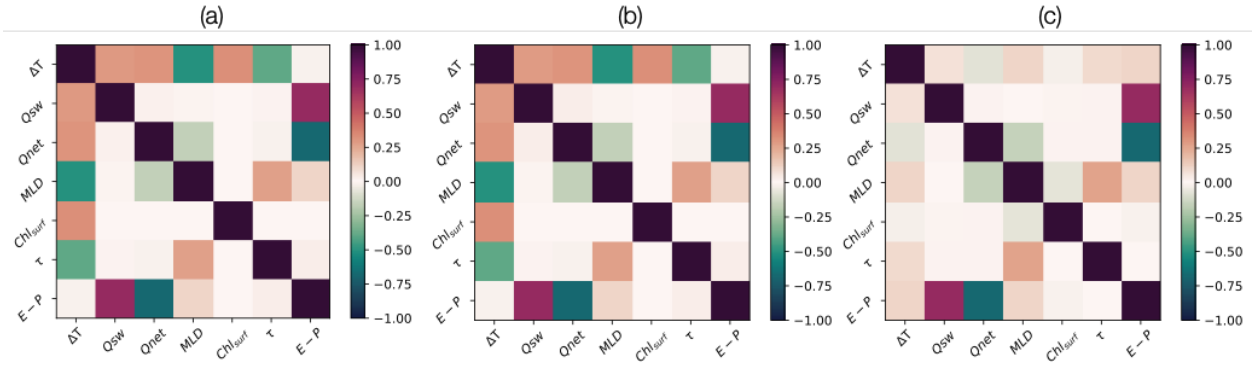

Figure S1: Heatmaps for unscaled correlations for each of the three comparison cases. (a) SS-C (same as Figure 5.4). (b) NO_SS-C. (c) SS-NO_SS.

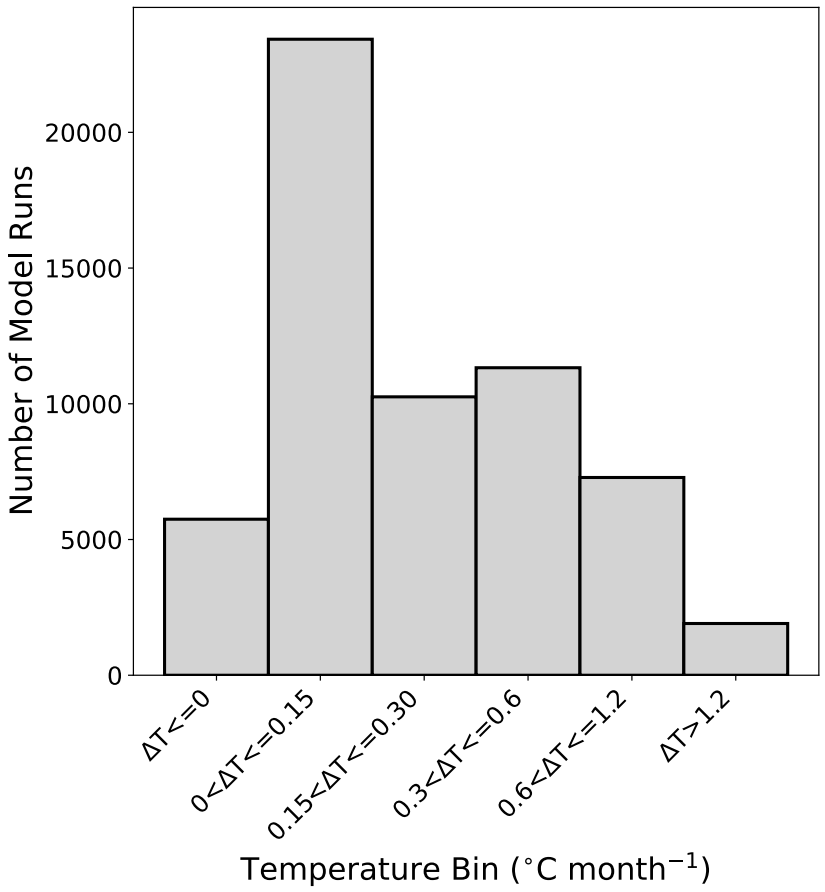

Figure S2. Histogram of number of simulation runs in each ΔT_{ML} (SS-C) bin. Positive values indicate that using the self-shading parameterization leads to warmer average mixed layer temperatures.

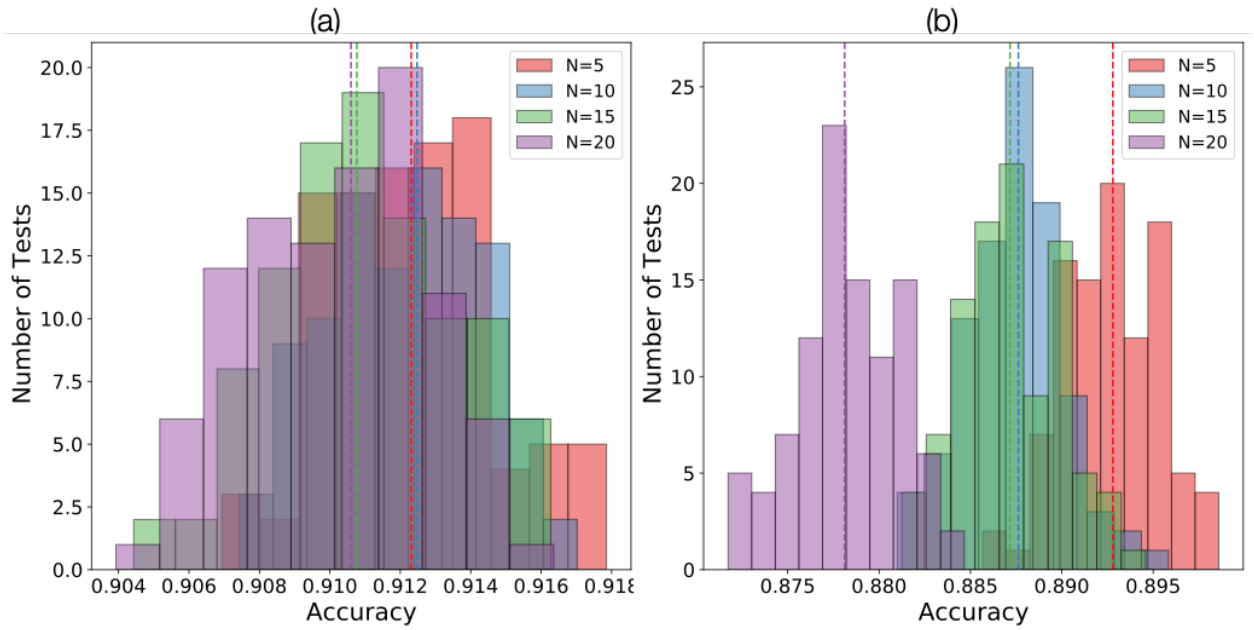

Figure S3: Model evaluation of K-Nearest Neighbors model for different numbers of neighbors for initial clustering (SS-C comparison). (a) Weighting of neighbors based on distance and (b) uniform weighting.

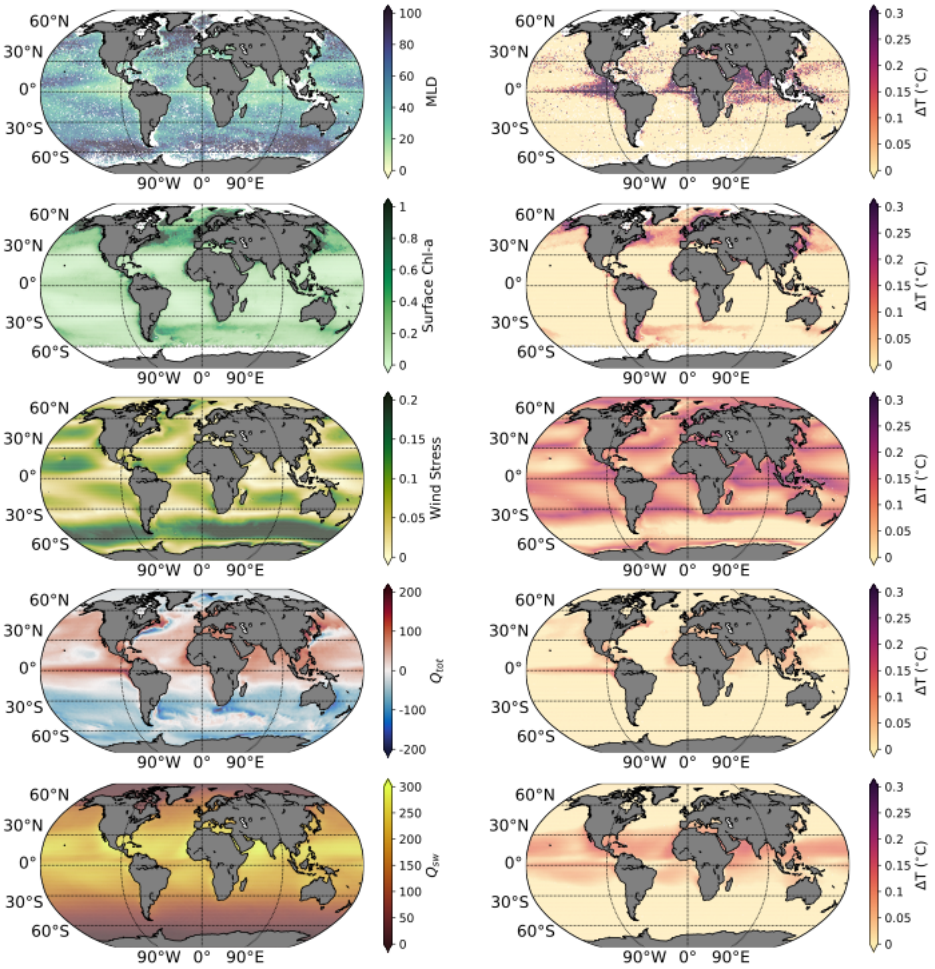

Figure S4. Same as Figure 5.5 but for April.

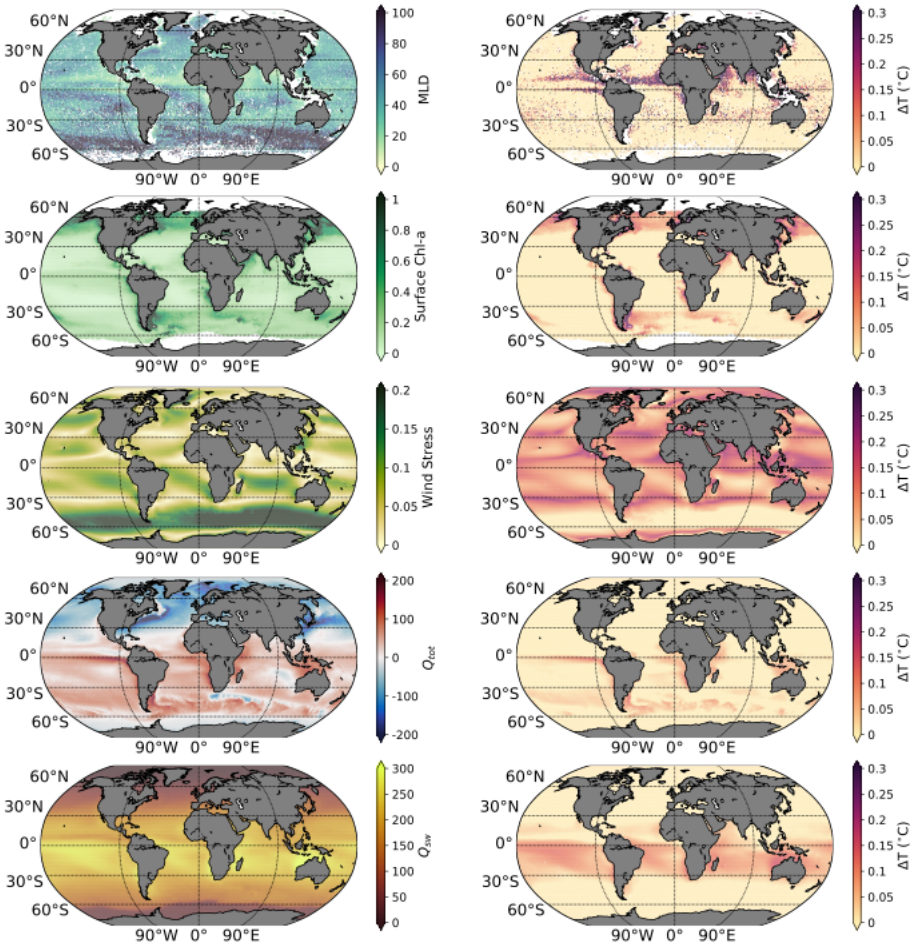

Figure S5. Same as Figure 5.5 but for October.

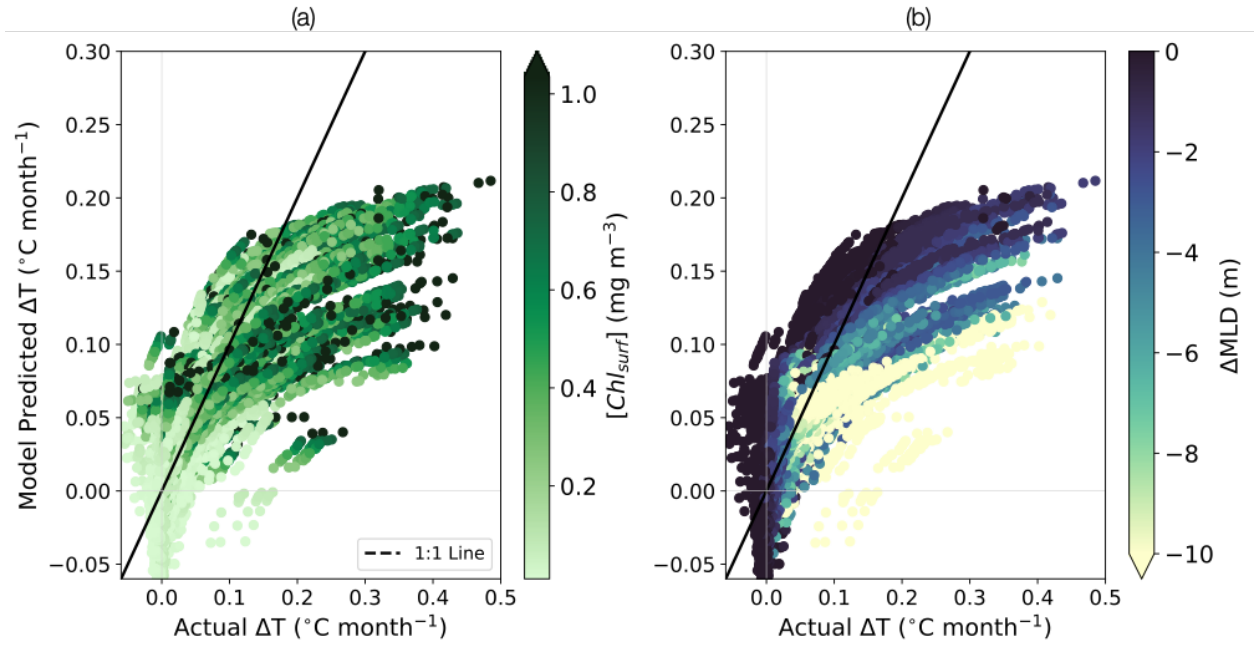

Figure S6. Difference between ΔT_{ML} (SS-C) produced by simulations (x-axis) and the value predicted by the multiple linear regression. (a) Shading indicates surface chlorophyll concentration. (b) Shading indicates difference in mixed layer depth at the end of the simulations (SS-C).

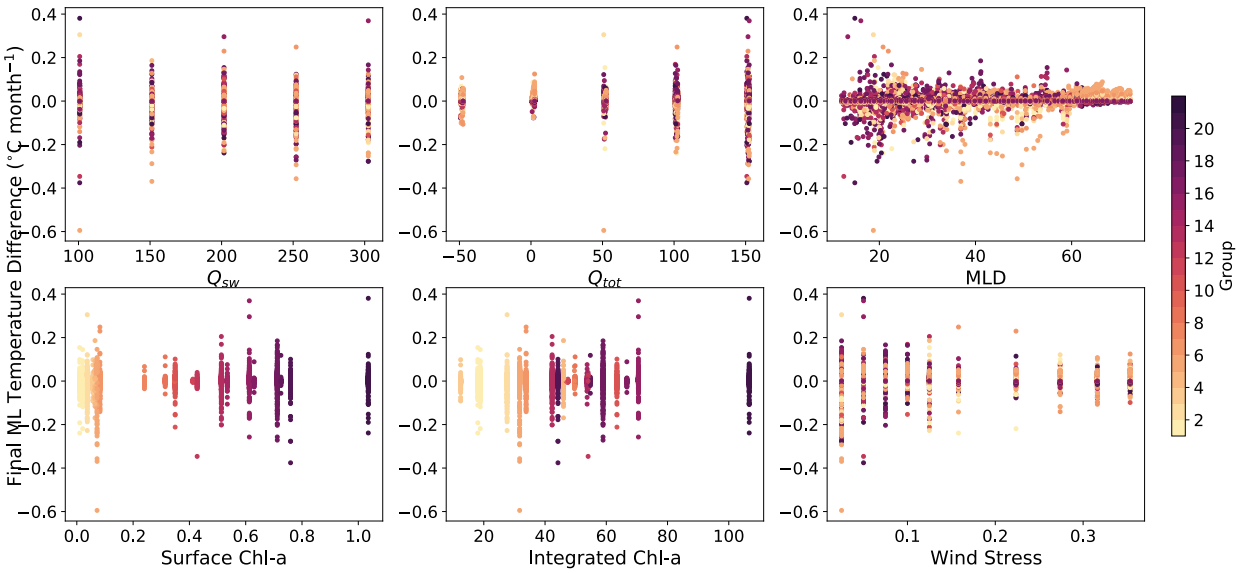

Figure S7. Scatter plot of model variables vs. final temperature mixed layer temperature differences for all simulation results (SS-NO_SS comparison). Coloring of dots corresponds to chlorophyll profiles (Group) shown in Figure 5.2 (coloring is sequenced from low surface chlorophyll in light yellow to high surface chlorophyll in dark purple).

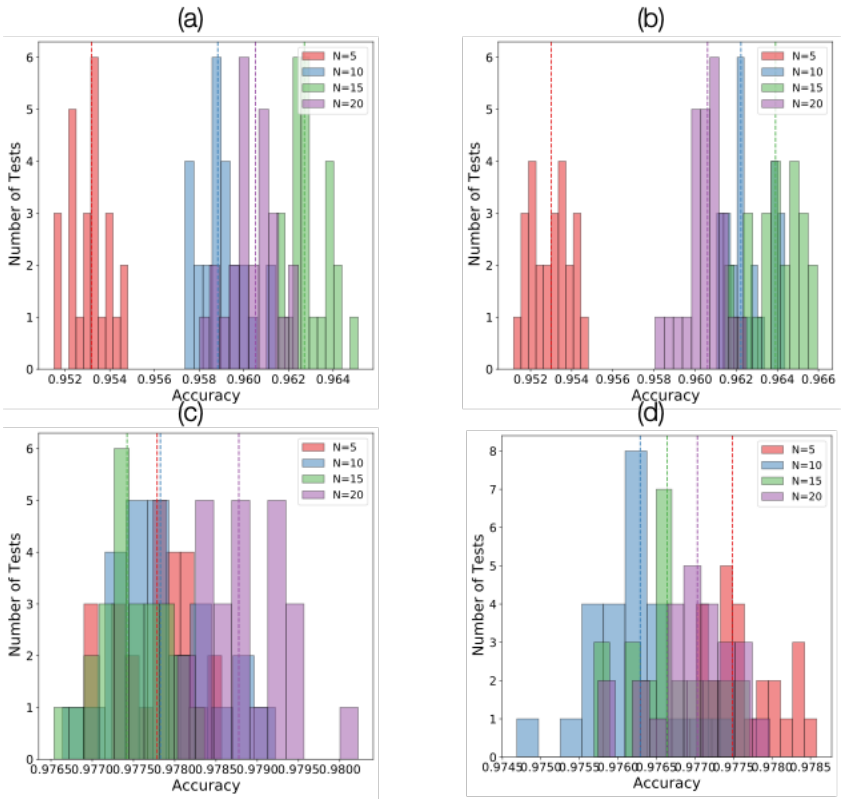

Figure S8: Model evaluation of K-Nearest Neighbors model for different numbers of neighbors for curve-fit clustering. (a) Weighting of neighbors based on distance and (b) uniform weighting for Gaussian + Line profiles. (c) and (d) Same as (a) and (b) but for Gaussian + Sigmoid profiles.

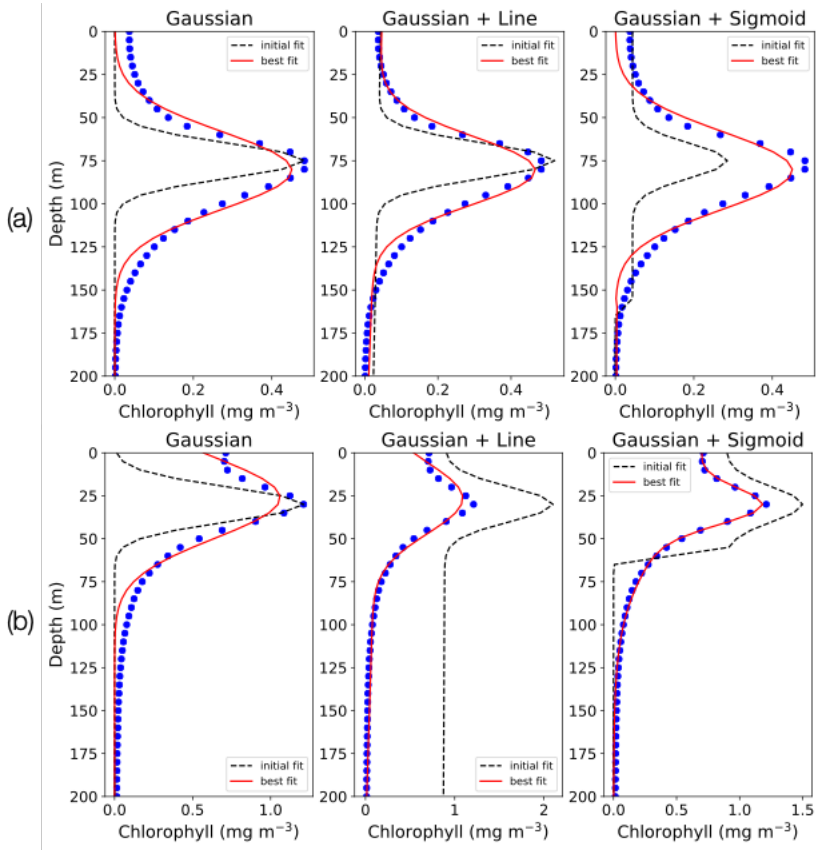

Figure S9. Example curve fits for two profiles from BGC Argo floats. (a) Best fit is a Gaussian+Line combination (Cluster 3 profile). (b) Best fit is a Gaussian + Sigmoid combination (Cluster 18 profile).

VITA

Rosalind Echols grew up in Seattle, Washington, and spent much of her childhood riding ferries, visiting the coast, and kayaking around the San Juan Islands. She discovered a love for physics and engineering in high school, and spent her first ten years after college teaching students about the power of scientific curiosity as a high school science teacher. After participating in the NOAA Teacher at Sea program, she decided to return to research and pursue a PhD in Physical Oceanography. The culmination of that endeavor is presented in this dissertation.