UNIVERSITY OF CALIFORNIA

Los Angeles

Analysis of Particulate Carbon Export in the Global Ocean, using *in situ* Observations and Machine Learning

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Atmospheric and Oceanic Sciences

by

Daniel Joseph Clements

2023

ABSTRACT OF THE DISSERTATION

Analysis of Particulate Carbon Export in the Global Ocean,
using *in situ* Observations and Machine Learning

by

Daniel Joseph Clements
Doctor of Philosophy in Atmospheric and Oceanic Sciences
University of California, Los Angeles, 2023
Professor Daniele Bianchi, Chair

The abundance and size distribution of marine organic particles are two major factors controlling biological carbon sequestration in the ocean. These quantities are the result of complex physical-biological interactions that are difficult to observe, and their spatial and temporal patterns remain uncertain. This dissertation describes our analysis of particle size distributions (PSD) and the resulting export, from a global compilation of *in situ* Underwater Vision Profiler 5 (UVP5) optical measurements.

In Chapter 2, we demostrate the ability to extrapolate sparse UVP5 observations to the global ocean from well-sampled oceanographic variables, using a machine learning algorithm. We reconstruct global maps of the biogenic PSD parameters (biovolume and slope) for particles at the base of the euphotic zone. These reconstructions reveal consistent global patterns, with high chlorophyll regions generally characterized by high particle biovolume and flatter PSD slope, i.e., a high relative abundance of large vs. small particles. The resulting negative correlations between particle biovolume and slope further suggest amplified effects on sinking particle fluxes. Our approach and estimates provide a baseline for understanding the export of organic matter from the surface ocean.

Chapter 3 describes how applying a simple empirical relationship to our reconstructions

of the PSD, we can calculate the total export. In this Chapter, we explore the seasonal and spatial patterns of carbon export. Taking advantage of the high vertical resolution of the UVP5, we quantify the export from the surface using two previously established depth horizons. We identify a larger export from the Southern Ocean than most other models of export. Similarly, we find the lower part of the euphotic zone to be dominated by heterotrophy, rather than autotrophy. Being able to reconstruct the PSD and particle flux at multiple depths allows for further exploration of the full 3-dimensional particle field.

Chapter 4 describes a full 3-D model, where depth specific export is calculated, highlighting significant deviations from idealized flux profiles and quantify the efficiency of the biological pump globally. Our results reveal the primary drivers of carbon storage and sequestration and highlight the importance of transport by diel vertical migration of marine animals. These estimates of the global particle field serve as a baseline for future model-based estimates of particulate flux, and as an independent estimate of the efficiency of organic matter storage in the ocean.

The dissertation of Daniel Joseph Clements is approved.

Jacob Bortnik

Thomas Weber

Tina Irene Treude

Daniele Bianchi, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The research presented in this dissertation is the product of collaborative work with my thesis advisor, Daniele Bianchi. I am grateful for his mentorship and guidance throughout this work. I am greatful for all my coauthors and mentors for constantly pushing me to improve my work, and for their guidance. Thomas Weber, Rainer Kiko, and Lars stemmann all provided important clarity and insights to each Chapter presented here.

I would also like to acknowledge the non-scientific support I had getting through the PhD. Firstly, and most importantly, my partner Roxana Azar, for her continued patience, support, and for constantly being my rock. I would also like to thank the many different friends and colleagues who helped me maintain sanity, and social interactions. Lastly, I would like to thank my family, both blood and chosen family for their constant patience as I rambled about the ocean. And for their comfort when I thought I was too stressed to continue.

Chapter 2 is reproduced from Clements et al. (2022a), Chapter 3 is reproduced from Clements et al. (2022b)(in review at Global biogeochemical cycles), and Chapter 4 is reproduced from Clements et al. (2023) (Not yet submmitted). Material from the Chapters that have been published are reproduced with permission from the publishers.

Each Chapter is based in part upon work supported by the National Science Foundation under Grant Number OCE-1635632 and OCE-1847687. For Chapter 2, I would like to thank my coauthors, Kelsey Bisson, and two anonymous reviewers for helpful comments. For Chapter 3, I would again like to thank my coauthors, Kelsey Bisson, and two anonymous reviewers for helpful comments. For Chapter 4 I would like to thank the attendees of the UVP user meeting 2022, for their keen insights into the work presented. I would like to thank my co-authors for their comments and feedback.

VITA

2011-2013 A.A. (Business Administration) Bakersfield College.

2013-2015 B.A. (Marine Science) University of California, Berkeley.

2017-2020 M.S. (Atmospheric and Oceanic Science), UCLA.

PUBLICATIONS

Clements, D.J. Yang, S., Weber, T., McDonnell, A. M. P., Kiko, R., Stemmann, L., & Bianchi, D. Constraining the particle size distribution of large marine particles in the global ocean with in situ optical observations and supervised learning. *Global Biogeochemical Cycles 2022*

Clements, D.J. Yang, S., Weber, T., McDonnell, A. M. P., Kiko, R., Stemmann, L., & Bianchi, D. New estimate of organic carbon export from optical measurements reveals the role of particle size distribution and export depth. *in review at Global Biogeochemical Cycles*

Clements, D.J. Yang, S., Weber, T., McDonnell, A. M. P., Kiko, R., Stemmann, L., & Bianchi, D. The fate and distribution of sinking organic matter in the global ocean. *in preperation for Nature Geosciences early 2023*

McCoy, D., Damien, P., **Clements, D.J.**, Yang, S., & Bianchi, D. Pathways of Nitrous Oxide production in the Eastern Tropical South Pacific oxygen minimum zone. *in review at Global Biogeochemical Cycles.*

CHAPTER 1

Introduction: The importance of the biological pump

The ocean absorbs inorganic carbon from the atmosphere, in the form of carbon dioxide (CO_2). This CO_2 is then taken up and utilized by phytoplankton to build their biomass, which accumulates in the water column as suspended particles. The fate of these particles, in turn, controls major oceanic biogeochemical cycles, and the ability of the ocean to sequester atmospheric CO_2 (Broecker et al., 1982; Sarmiento & Gruber, 2006a; Kwon et al., 2009). Particles can be remineralized by microbial degradation and grazing by zooplankton (Karl et al., 1984; Steinberg et al., 2008; S. L. Giering et al., 2014), aggregate and disaggregate following physical and biological interactions (Kiørboe et al., 1990; Jackson, 1990; Dilling & Alldredge, 2000; Burd & Jackson, 2009; Briggs et al., 2020) and sink out of the surface layers to transfer carbon and nutrients to the ocean interior (K. O. Buesseler et al., 2007; J. S. Turner et al., 2017; Boyd et al., 2019). Marine particles are hotspots of microbial activity and diversity (Karl et al., 1984; DeLong et al., 1993; Church et al., 2021) provide the bulk of energy that sustain mesopelagic and abyssal ecosystems (S. L. C. Giering et al., 2020), and eventually supply materials to the ocean sediment, where burial sequesters them for geological timescales (Sarmiento & Gruber, 2006a; Dunne et al., 2007). The set of processes responsible for generating, exporting, and remineralizing organic matter production are collectively known as the biological pump.

The strength and pattern of the global biological pump is dependent on the ability of organic matter to escape shallow remineralization and sink into the interior ocean, often referred to as the particle transfer efficiency (K. O. Buesseler & Boyd, 2009; Weber et al., 2016). Although organic matter is initially quite small (micrometer scale), aggregation, and coagulation, as well as repackaging of organic matter, can generate large enough particles

that sink efficiently through the water column (Boyd et al., 2019; Burd & Jackson, 2009). Disaggregation and consumption of particulate matter by animals results in a reduction of size and thus sinking speed (Cram et al., 2022). Similarly, the amount of carbon in a particular particle is approximated by the size of the organic matter (Kriest, 2002; Guidi et al., 2008; Kiko et al., 2017). Ultimately, the rates of these particles transformations, sinking speed, and carbon contents are thought to be size-dependent (Kiørboe et al., 1993; Devries et al., 2014). Thus, the particle abundance in different sizes (i.e., the Particle Size Distribution; PSD) is considered a primary determinant in controlling the ocean's biological pump.

This Chapter reviews the current understanding of the biological pump and the different methods used to assess its strength. The goal of this Chapter is to provide a brief and broad overview of current literature relevant to this work. Each individual Chapter will provide an introduction and background relevant to that specific self-contained project.

1.1 Previous estimates of export

Satellite observations have been previously shown to estimate the size distribution of organic matter in the surface ocean (Kostadinov et al., 2009, 2010a,b). However, satellite retrievals miss larger particles that more directly contribute to sinking particle export and are limited to the upper few meters of the ocean. Thus satellites provide little information on particle transformations, interactions, and fluxes from the surface ocean.

Despite their limitations, satellite retrievals of organic matter have historically been used in conjunction with ocean models to estimate the total organic matter exported from the surface ocean (Siegel et al., 2014; DeVries & Weber, 2017). Current estimates of organic matter export yield a wide range of values, seasonality, and spatial patterns, with some discrepancies depending on the methods used (Quay et al., 2020). Biogeochemical models yield a global export of 4-6 PgC/y (Siegel et al., 2014), but can reach up to 10 PgC/y when tuned to match *in situ* profiles of nutrients and other biogeochemical tracers (DeVries et al., 2017). A similar range is suggested by recent global IPCC-class Earth System Models, which produce global carbon exports from 2.4 to 12 PgC/y, with an average of 7.4 PgC/y (Séférian et al.,

2020). Data-driven estimates combining satellite-based primary production with empirical estimates of particle export ratios often result in larger fluxes (Dunne et al., 2007; Laws et al., 2011; Guidi et al., 2016). Regional particulate matter export patterns predicted by these methods also vary substantially; with some models predicting strong gradients between the high export in productive waters and muted export in oligotrophic waters (Dunne et al., 2007), while others suggest more muted variations (DeVries et al., 2017). Differences in regional export fluxes have been attributed to methodological limitations, including scarcity and variability of in situ data used to constrain models, variability in satellite-based primary production algorithms, and models not able to fully capture underlying physical and biological processes.

While quantifying the surface export is of importance, it is does not wholly reflect the ocean's ability to respire and trap carbon. The strength of remineralization in the interior ocean is often reflected through the particle transfer efficiency, or the ratio of organic matter at 1 km to the base of the euphotic zone (K. O. Buesseler & Boyd, 2009; Cram et al., 2018; Weber et al., 2016). Quantification of this efficiency is typically done with coarse vertical resolution of sediment trap observations, or by prescribing a particle flux attenuation based on historic measurements (Martin et al., 1987). Previous model based estimates, on the global scale, found a spatial pattern of export efficiency to reflect that of the surface productivity and highlights the need for more observationally constrained estimates (Weber et al., 2016; Cram et al., 2018).

Recent advances in the field of ocean optics have dramatically increased the particulate matter observations in the ocean and allow for direct quantification of the PSD. The Underwater Vision Profiler 5 (UVP5) is an optical particle counter that provides the in situ particle abundance for large sinking class particles ($80\mu m - 2.6cm$) in a given sampled volume (Picheral et al., 2010). The UVP5 consists of a camera attached to the CTD rosette which can collect images at high frequency as it is lowered in the water column. Vertical profiles of PSD from the UVP5 are commonly taken at up to 20 images per second, with downward speeds of 1 m/s as deep at 6 km (Picheral et al., 2010). Since 2008, UVP5s have been routinely deployed on oceanographic cruises, in all ocean basins and seasons. These PSD's can

then be combined with empirical relationships relating particle size and abundance to sinking fluxes (Kriest, 2002; Guidi et al., 2008; Kiko et al., 2017), which can be tuned against a global data set of sediment trap and thorium-derived particle flux observations (K. M. Bisson et al., 2018).

1.2 Science questions and objectives

This dissertation will investigate the ocean's biological pump and is centrally motivated on understanding the main processes affecting the climatological distribution of the resulting export.

Specifically, Chapter 2 and 3 are focused on the export from the surface ocean, specifically addressing the questions:

- How does the PSD vary globally and temporally?
- What is the resulting global carbon flux from the PSD?
- How is the export pattern determined by export horizon?
- What factors in the ocean control the size distribution of organic matter, and the resulting export?

The remainder of the dissertation will take a deeper look into this export, specifically examining what happens to organic matter below the surface of the ocean. I quantify the depth dependence on this organic matter, and address the following motivating questions?

- What is the transfer efficiency of organic matter, and how does this vary regionally?
- What are the primary controls on the pattern of export removal from the sinking pool?
- How much carbon makes it into the deep ocean, and is sequestered over long timescale?

1.3 Outline

In this first part of the dissertation, I have outlined the previous estimates of organic carbon export from the surface ocean, and the various approaches and considerations that need to be made. I have similarly highlighted the importance of a new estimate, and the overarching principals driving the rest of this dissertation. In general, each Chapter is focused on using new data and methods to understand a poorly understood, but central process in ocean biogeochmistry.

The first two Chapters take a classic approach from the surface ocean providing an estimate of what "leaves the surface" ocean which has many implication ranging from ecological to climatic importance. The last part of the dissertation changes focus to what happens below the surface ocean, this final step increases the model complexity but allows for a truly data-driven exploration into subsurface particle dynamics.

In Chapter 2, I investigate the particle size spectra from the surface ocean. There are many size-dependent reactions that can occur at the interface of large organic particles, and using a few simple assumptions I am able to fully quantify the abundance of organic matter. This Chapter employs use of machine learning model to globally extrapolate sparsely measured data into a global monthly climatology. In this I find a dynamic seasonal cycle of the particulate matter, additionally I explore some important de-coupling of the two PSD parameters.

In Chapter 3, I extend the particle size spectra into a discussion about the particle export itself. I use a size-based relationship to build a model of export from two different depth horizons; the euphotic zone and the maximum mixed layer. The resulting export from each horizon should provide a quantification of carbon export for either an ecological approach or a climatic approach. I present these resulting carbon fluxes, and identify significant deviations from previous models, and highlight the importance is distinguishing these depths.

In Chapter 4, I copy the same general model from the previous two Chapters, and analyze the changes in carbon export with depth. In this Chapter I present the resulting sub-surface calculations of particle export and describe the pattern of attenuation seen. These results

challenge the idea that a simple depth dependent relationship can be drawn to estimate the flux at depth. These results are the first of its type, using a size fractionated export model to understand the interior ocean particle dynamics, and serves as a framework for future data-driven exploration and as a baseline for large scale ocean models.

Finally, in Chapter 5, I summarize the advances presented in this dissertation. I highlight the importance and need for continued efforts, and highlight the ongoing, and future work inspired by the work presented here.

CHAPTER 2

Constraining the particle size distribution of large marine particles in the global ocean with *in situ* optical observations and supervised learning

2.1 Introduction

Throughout the ocean surface, autotrophic organisms fix CO_2 and inorganic nutrients to produce organic matter, which accumulates in the water column as suspended particles (Falkowski et al., 1998). The fate of these particles in turn controls major oceanic biogeochemical cycles, and the ability of the ocean to sequester atmospheric CO_2 (Broecker et al., 1982; Sarmiento & Gruber, 2006b; Kwon et al., 2009). Particles can be remineralized by microbial degradation and grazing by zooplankton(Karl et al., 1988; Steinberg et al., 2008; S. L. Giering et al., 2014), aggregate and disaggregate following physical and biological interactions (Jackson, 1990; Kiørboe et al., 1990; Dilling & Alldredge, 2000; Burd & Jackson, 2009; Briggs et al., 2020), and sink out of the surface layers to transfer carbon and nutrients to the ocean interior (K. O. Buesseler et al., 2007; J. T. Turner, 2015; Boyd et al., 2019). Marine particles are hotspots of microbial activity and diversity (Karl et al., 1984; DeLong et al., 1993; Church et al., 2021), provide the bulk of energy that sustain mesopelagic and abyssal ecosystems (Burd et al., 2010; S. L. C. Giering et al., 2020), and eventually supply materials to the ocean sediment, where burial sequesters them for geological timescales (Sarmiento & Gruber, 2006b; Dunne et al., 2007).

Most processes that drive particle cycles in the ocean depend on particle size (Stemmann & Boss, 2012). To first order, size controls particle elemental composition (A. L. Alldredge

& Gotschalk, 1988), aggregation and disaggregation rates (Burd & Jackson, 2009; Briggs et al., 2020), and the ability of particles to sink (A. L. Alldredge & Gotschalk, 1988; McDonnell & Buesseler, 2010; B. Cael et al., 2021), thus providing a first order influence on the ocean's biological pump (Boyd et al., 2019). Furthermore, size-dependent properties such as particle volume and surface area affect interactions with microorganisms, including colonization, metabolism, and particle degradation (Jackson, 1989; Kiørboe et al., 2002; Nguyen et al., 2022; Bianchi et al., 2018), and coupling with seawater chemistry via regeneration of elements (Broecker et al., 1982; Sarmiento & Gruber, 2006b), adsorption and scavenging processes (Turekian, 1977; Ohnemus et al., 2019).

Therefore, the abundance of particles of different sizes, i.e., the particle size distribution (PSD) is a primary determinant of biogeochemical and ecological interactions, and retains important information on particle dynamics (Stemmann & Boss, 2012). The size distribution of marine particles can thus reveal processes by which biogenic matter is generated and removed. Aggregation and coagulation, as well as repackaging by marine organisms (e.g., by filter feeding, formation of fecal pellets, and sinking carcasses) lead to increase in the size of particles, and hence of their sinking velocity and ability to escape remineralization in the upper ocean (J. T. Turner, 2015; Burd & Jackson, 2009). Conversely, disaggregation and consumption by microorganisms tend to reduce the size of particles and their sinking speed, facilitating shallow remineralization (Karl et al., 1988; Goldthwait et al., 2005; Briggs et al., 2020).

Particle dynamics is also central to Earth System Models used to describe ocean biogeochemistry and the carbon cycle, and project their future change (Bopp et al., 2013; Kwiatkowski et al., 2020). While current ocean biogeochemical models include at most few size classes (Séférian et al., 2020), explicit representation of PSD has emerged as a powerful and promising approach to mechanistically represent size-dependent processes and their influence on elemental cycles, carbon sequestration, and ecological interactions in the ocean (Kriest & Evans, 1999; Stemmann et al., 2004b; Gehlen et al., 2006; Burd & Jackson, 2009; Devries et al., 2014; Cram et al., 2018; Weber & Bianchi, 2020; Omand et al., 2020; Nguyen et al., 2022). However, quantifying the large-scale abundance, distribution, and size structure

of marine particles has been historically difficult.

Satellite-based observations allow an estimate of the PSD in the surface ocean, for particle size ranges that typically include phytoplankton and small, slowly sinking particles (Kostadinov et al., 2009, 2010b,a). This work often hinges on the assumption that a power law distribution can describe the entire size spectra (Bader, 1970; Sheldon et al., 1972); however deviations from a power law have been documented (Jonasz & Fournier, 1996; Organelli et al., 2020; R. A. Reynolds et al., 2010, 2016), and likely retain information on particle cycling processes ranging from primary production to grazing and sinking (Guidi et al., 2009; Stemmann & Boss, 2012; Huete-Ortega et al., 2014; R. A. Reynolds & Stramski, 2021a). Furthermore, satellite retrievals miss larger particles that more directly contribute to sinking particle export, and are limited to the upper few tens of meters of the ocean, thus providing little direct information on particle transformations, interactions, and fluxes in subsurface layers. Despite the limitations, satellite-based PSD estimates have proven helpful to constrain models of the ocean's biological pump (Siegel et al., 2014; DeVries & Weber, 2017).

Recent advances in ocean optical observations enable direct determination of *in situ* PSD throughout the water column (Stemmann & Boss, 2012; Boss et al., 2015; Lombard et al., 2019). The Underwater Vision Profiler 5 (UVP5) is an optical particle counter that provides the *in situ* particle abundance for relatively large particles (80 μm - 2.6 cm) in a given sampled volume (Picheral et al., 2010). The UVP5 consists of a camera attached to the CTD rosette, and is able to collect images at high frequency as it is lowered in the water column. Vertical profiles of PSD from the UVP5 are commonly taken at up to 20 images per second, with downward speeds of 1 m s^{-1}, as deep at 6 km (Picheral et al., 2010). Since 2008, UVP5s have been routinely deployed on oceanographic cruises, in all ocean basins (Kiko et al., 2021).

Because UVP5 instruments observe a range of sizes that includes marine aggregates that can rapidly sink, they are especially helpful for characterizing patterns and fate of particulate organic matter and sinking carbon. Although the UVP5 cannot determine the organic fraction or carbon content of the particulate matter, we assume this can be reasonably

well approximated by empirical relationships (A. L. Alldredge & Gotschalk, 1988; Kriest, 2002; Guidi et al., 2008; Kiko et al., 2017). Prior studies have utilized UVP5 observations, together with this assumption, to shed light on the ocean's biological pump. For example, Guidi et al. (2008) showed that PSD observations from UVP5 can be combined with sediment trap data to estimate sinking carbon fluxes. A similar approach was later used to estimate regional carbon fluxes (Forest et al., 2012; Guidi et al., 2016; Kiko et al., 2017), as well as regional patterns of particle transfer efficiency and deep carbon sequestration (Guidi et al., 2015). Recently, the study by Cram et al. (2018) combined UVP5 observations taken along a meridional section in the Pacific Ocean and satellite-based surface chlorophyll to reconstruct global PSD and drive a model of marine particle dynamics. While these studies demonstrate the potential of UVP5 observations for regional and global investigations, they are based on relatively small data sets, which limits the robustness of extrapolations to the entire ocean.

In this study, we take advantage of the rapid growth of UVP5 observations and employ a machine learning approach to reconstruct global patterns of PSD in the upper ocean, and investigate their characteristics and drivers. Specifically, we train a supervised machine learning algorithm to reconstruct PSD from relatively sparse UVP5 observations and well-sampled oceanographic variables. By comparing patterns in PSD with environmental drivers, we further gain insight into the potential mechanisms responsible for shaping the surface ocean's PSD and its variability.

The rest of the chapter is organized as follows. Section 2 describes the machine-learning approach used to globally extrapolate PSD. Section 3 presents the reconstructions of particle distributions and compares These results to previous studies, discussing the uncertainties and caveats inherent to this approach. Section 4 summarizes the main findings and discusses future directions.

2.2 Methods

The size distribution of marine particles is the result of complex physical-biogeochemical interactions (Sheldon et al., 1972). Although each individual process may leave a specific

signature on the PSD (Sheldon et al., 1977; Kiørboe et al., 1993; Huete-Ortega et al., 2014; R. A. Reynolds & Stramski, 2021a), to first order, the PSD can be quantitatively described by a power law (Bader, 1970; Sheldon et al., 1972) over a relatively broad size range (from micrometers to centimeters) that encompasses observations of marine aggregates with optical instruments, including UVP5 (Stemmann & Boss, 2012). Thus, following the power law assumption, we model observed PSD as:

$$n(s) = n_0 \cdot s^{-\beta}, \tag{2.1}$$

where s is the particle equivalent spherical diameter, or size, and, after defining an arbitrarily small size increment ds, $n(s)ds$ the number of particles in the size range $[s, s+ds]$. This power law approximation depends on two parameters: the power law constant n_0 (i.e., the size-independent coefficient often referred to as the power law "intercept"), and the slope β (the exponent for size-dependence). The power law constant of the PSD represents the number of particles at an arbitrary reference size, and the slope encapsulates the relative proportion between small and large particles. For a given slope, increasing the power law constant proportionally increases the total number of particles. Conversely, for a given power law constant, increasing the slope (i.e., making the spectrum "steeper") increases the proportion of small particles, while decreasing the slope (i.e., making the spectrum "flatter") increases the proportion of large particles. Relatively small changes in the slope can thus result in significant changes in the size partitioning of particles and in quantities that depend on this partitioning, such as the total particle volume and surface area.

While generally accurate to first order over selected size ranges, the power law assumption encapsulated in Equation 1 has been shown to miss important deviations that often characterize in situ PSD (Jonasz & Fournier, 1996; R. A. Reynolds & Stramski, 2021a). These include local maxima at specific sizes that reflect dominance of plankton species in productive regions, in particular in the micrometer range of phytoplankton cells (Cavender-Bares et al., 2001; Karp-Boss et al., 2007; Huete-Ortega et al., 2014; Organelli et al., 2020; Runyan et al., 2020), or the influence of mineral phases from riverine inputs and exchange

with sediment in coastal waters (R. A. Reynolds et al., 2010, 2016). These deviations in turn may hold important information on the processes and dynamics of particle cycling such as production, removal, aggregation, and disaggregation (Sheldon et al., 1977; Kiørboe et al., 1993; Huete-Ortega et al., 2014; R. A. Reynolds & Stramski, 2021a; Briggs et al., 2020), or zooplankton grazing dynamics (Schartau et al., 2010; Moscoso et al., 2022). Accordingly, alternative methods have been proposed to approximate PSD observations, including sums of log-normal functions (Jonasz & Fournier, 1996) and non-parameteric approaches based on cumulative PSD (R. A. Reynolds & Stramski, 2021a). Nevertheless, the power law assumption has proven valuable and accurate as a first order description of marine particle size spectra, in particular for the study of size-dependent processes such as particle sinking and transfer to depth (Guidi et al., 2009; Stemmann & Boss, 2012; Roullier et al., 2014; Guidi et al., 2016).

Here, we use UVP5 observations to reconstruct PSDs under the power law assumption (i.e., n_0 and β) at the base of the euphotic zone, by fitting Equation 2.1 to observed normalized particle concentrations, i.e., the number of particles per unit volume, divided by the width Δs of each specific size bin considered, thus providing a discrete approximation to $n(s)$. we then extrapolate the sparse UVP5 observations to a global grid, by training a supervised learning algorithm to predict spatially-varying PSD parameters from well-sampled environmental predictors. we exploit the three-dimensional nature of UVP5 observations to perform these calculations at a varying base of the euphotic zone, here defined by the 1% light level according to Morel et al. (2007), rather than a single depth. The steps used to reconstruct global PSD from UVP5 observations are illustrated in the workflow schematic in Fig. 2.1, and are discussed in the following sections.

2.2.1 Reconstructions of particle size spectra from UVP5 data

we use observations from a new compilation of UVP5 measurements spanning the global ocean (Kiko et al., 2021). The data set consists of over 6700 profiles from 119 cruises, collected from 2008 to 2020 (Fig. 2.2). These observations provide robust particle counts for

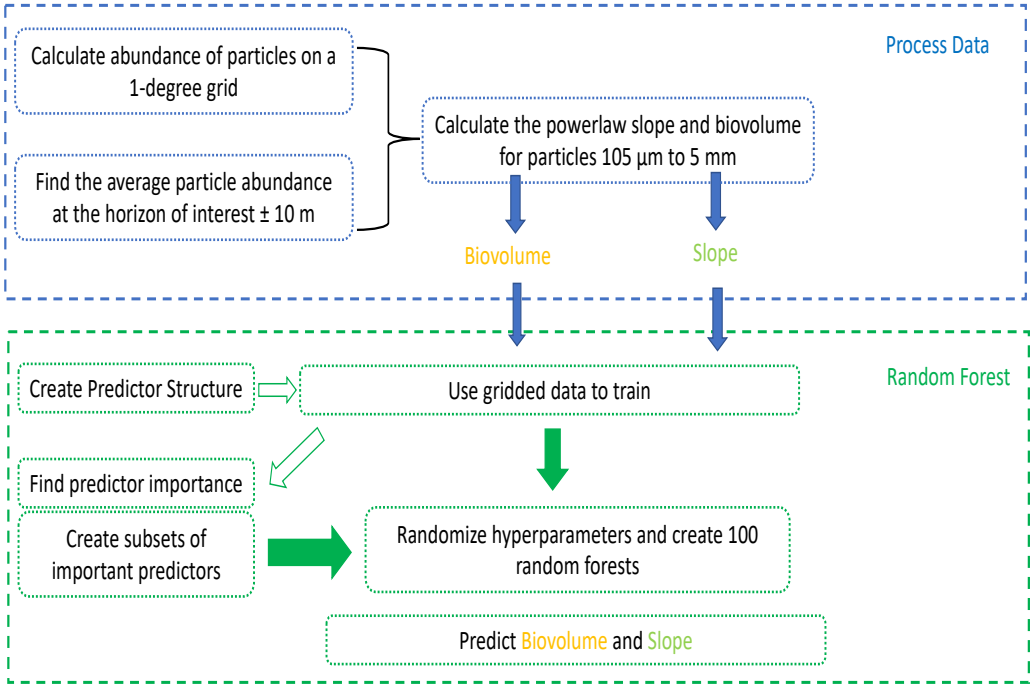

Figure 2.1: Schematic diagram illustrating the general workflow of processing UVP5 observations into a global PSD dataset. Observations are ensembled onto a normal 1 degree grid, with observation representing an average of a 20 meter vertical bin about the export horizon. PSD observations (power law slope and biovolume) are calculated for the 105 μm to 5 mm size range. The PSD slope and biovolume are globally extrapolated using a bagged Random Forest algorithm.

the 105 μm - 5 mm size range at each location and depth. Under the power law assumption (Equation 2.1), the two parameters n_0 and β are needed to capture the PSD (Bader, 1970; Sheldon et al., 1972; Stemmann et al., 2004a; Stemmann & Boss, 2012; Devries et al., 2014).

we calculate the power law slope β by fitting a linear least-squares regression through the log-transformed normalized particle abundance as a function of the log-transformed size. we then calculate the observed particle volume, hereafter referred to as the biovolume(BV) since most particles sampled by the UVP5 are biogenic or organic in nature, by multiplying the volume of a particle of a given size s by the observed size distribution $n(s)$, and integrating over the full size range:

$$BV = \int_{s_{min}}^{s_{max}} n(s) \cdot \frac{\pi}{6} \cdot s^3 \, ds. \tag{2.2}$$

In practice, the continuous integral is approximated by a summation over all size bins in which the UVP5 observations are discretized.

Under the power law assumption, the biovolume can also be expressed analytically as a function of the slope and power law constant, by substituting Equation 2.1 into Equation 2.2:

$$BV = \int_{s_{min}}^{s_{max}} n_0 \cdot s^{-\beta} \cdot \frac{\pi}{6} \cdot s^3 \, ds = \int_{s_{min}}^{s_{max}} \frac{\pi}{6} \cdot n_0 \cdot s^{3-\beta} \, ds = \frac{\pi}{6} \cdot n_0 \cdot \left(\frac{s_{max}^{4-\beta}}{4-\beta} - \frac{s_{min}^{4-\beta}}{4-\beta} \right). \tag{2.3}$$

By fixing the size range, i.e., the minimum and maximum particle size that can be robustly derived from UVP5 instruments (s_{min} and s_{max} respectively), we solve Equation 2.3 for the power law constant n_0 as a function of the PSD slope and the observed biovolume:

$$n_0 = \frac{6 \cdot BV}{\pi} \cdot \left(\frac{s_{max}^{4-\beta}}{4-\beta} - \frac{s_{min}^{4-\beta}}{4-\beta} \right)^{-1}. \tag{2.4}$$

we set the minimum and maximum size for this equation to the same values used to estimate the slope and biovolume from UVP5 observations. we use a minimum size s_{min}=105 μm to avoid a potential slight instrument bias in the lowest size classes. we set the maximum size to s_{max}=5 mm, which corresponds to the size where zooplankton start to dominate the biovolume at a variety of locations sampled by UVP5 (Forest et al., 2012; Stemmann, Youngbluth, et al., 2008; Stemmann & Boss, 2012).

we coarsen the temporal and spatial resolution of the UVP5 profiles by binning them onto the standard monthly 1 degree-resolution grid of the World Ocean Atlas (H. Garcia et al., 2018; H. E. Garcia et al., 2019). That is, we combine multiple profiles in a given grid cell and month together, thus reducing variability due to the noisy and episodic nature of particle observations. we opted for a spatial resolution of 1-degree to allow representation of spatial variability associated to major oceanographic features relevant to carbon export, such as

large-scale fronts, upwelling systems, coastal to open-ocean transitions, marginal seas, and shelves. Furthermore, a 1-degree resolution enables us to utilize the native resolution of many predictors with minimal post-processing, e.g., hydrographic properties and nutrients that are defined on the World Ocean Atlas grid, a *de facto* standard in ocean biogeochemistry. we also combine all observations within a 20 meter-thick depth bin around each chosen depth horizon, to further smooth out small-scale vertical variability, and to increase the significance of particle counts, especially for the largest sizes. To reconstruct global PSDs, we calculate slope and biovolume at each location, at the given depth horizon, using the gridded observations, and assume that these averages are representative of the climatological monthly PSD in each grid cell.

Although the gridding procedure reduces noise and data patchiness in many well-sampled regions, a significant proportion of grid cells only contains a single profile ($\sim$ 45%). As a further quality check, we test the assumption that a power law distribution is a good approximation for the observed PSD, while recognizing that significant deviations from a power law often exist, as discussed in Section 2. For each grid cell with observations, we place an objective goodness of fit threshold to determine the robustness of the power law fit. If a power law fit has a Pearson correlation coefficient R^2 of less than 0.9, we remove the data point, as it likely does not closely follow a power law distribution. This quality control step removes less than 1% of data, with the majority of retained profiles showing correlations near or greater than R^2=0.98 (Supplementary Information Fig. A.1). Data points removed, and points with lower correlations, appear to be found in few coastal regions such as the tropical Atlantic under the influence of the Amazon River plume, the Arctic shelf, and productive waters in the subpolar North Atlantic and near Antarctica (Fig. S1), and likely reflect a combination of allochthonous mineral particle supply and highly productive or variable conditions (R. A. Reynolds et al., 2010, 2016; R. A. Reynolds & Stramski, 2021b). The results of this quality check indicate that, for the size range considered here (105 μm to 5 mm), the majority of open ocean UVP5 observations can be well approximated by power laws, in line with previous work (Guidi et al., 2009; Stemmann & Boss, 2012; Roullier et al., 2014; Cram et al., 2018). The final processed UVP5 observation data set contains 2,034

gridded observations at the export horizon, which together cover slightly less than 10% of the ocean surface. Figure 2.2 shows the spatial and temporal resolution of the final gridded data set, and an example of the observed PSD from UVP5 with the corresponding power law fit.

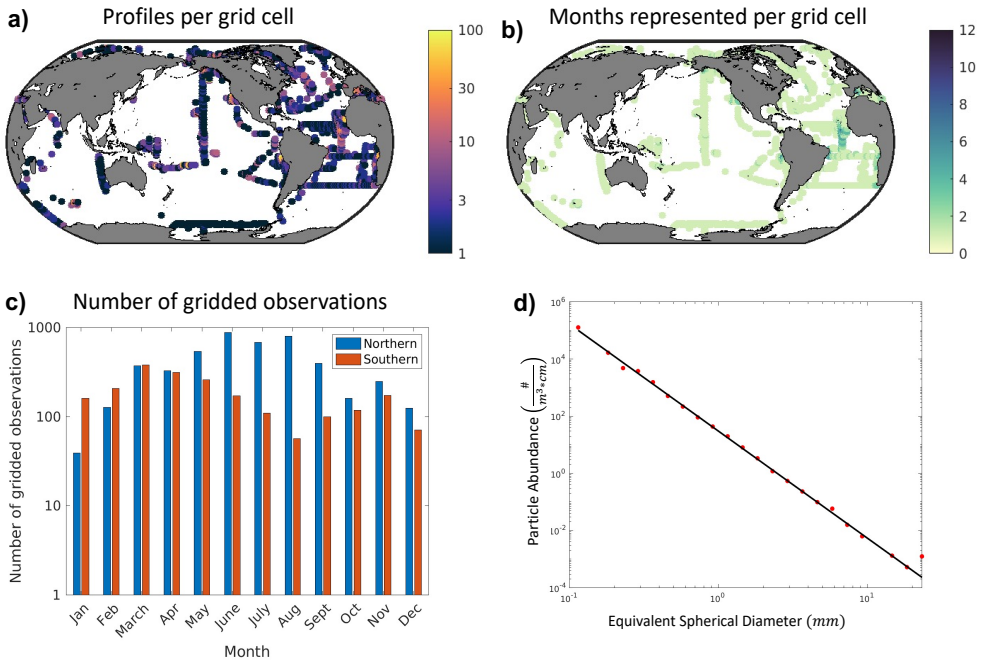

Figure 2.2: Global distribution of the UVP5 observations used in this study. (a) Number of profiles per one-degree resolution grid cell. (b) Number of months represented in each grid cell. (c) A bar plot showing the monthly distribution of observations for the Northern Hemisphere (blue) and the southern Hemisphere (red). (d) Typical particle size distribution sampled by the UVP5, plotted on a logrithmic scale. The red dots indicate actual observations, and the black line the linear fit ($R^2 = 0.99$).

2.2.1.1 Training and evaluating a Random Forest model

Monthly flux reconstructions require extrapolation of PSD parameters to the whole ocean on monthly time scales. we use a bagged Random Forest (RF) algorithm (the "fitrensemble" function in MATLAB) to reconstruct climatological PSD slope and biovolume globally, following an approach similar to Yang et al. (2020). A RF deploys a decision tree learning

scheme to solve a regression equation iteratively, and reports the ensemble average. Using a RF, each individual decision tree is trained on a subset of the available data, with a subset of predictors, but the power of the method emerges when considering the ensemble average. The RF is able to learn statistical relationships between target variables (here, UVP5-derived slope and biovolume) and a series of predictors (here, environmental variables), to make reconstructions that minimize the error between predicted and observed data. Because a RF is highly non-linear, it runs the risk of overfitting the data, producing solutions with low error, but also limited predictive power outside of the training data set. To mitigate the risk of overfitting, the RF does not use all data points for training. Instead, a bootstrapped sample ($\sim 70\%$) of the data is selected for each tree in the forest. The skill of the final regression is determined by finding the error between the model and the data that was not used for training, i.e., the so-called "out-of-bag" data.

The rank of predictors is given by the out-of-bag error coupled with an internally derived measure of importance, using a so-called "recursive feature elimination" approach. A recursive feature elimination systematically removes the least important predictor and records the out-of-bag error to describe the contribution of each predictor to the final solution. When there is relatively no change in the out-of-bag error for every additional predictor, these predictors are considered not important for the RF (Supplementary Fig. A.2). we determine statistical importance in order to establish a reduced set of predictors, reducing the risk of over-fitting while not losing predictive power. When interpreting the RF results, we apply qualitative understanding of the predictors combined with the recursive feature elimination to determine if a predictor should be included in the final regression or if it should be excluded.

2.2.1.2 Environmental Predictors

The RF algorithm relies on a set of predictors and target data at the resolution of the desired reconstruction. In this case, we use climatological monthly predictors at 1-degree spatial resolution. we include a variety of predictors that are globally sampled and

could be mechanistically related to particle production in the surface ocean, ranging from physical variables (e.g., temperature and salinity) to ecosystem-level quantities (e.g., primary production, euphotic zone depth). we also include as a predictor the standard deviation of the primary production, using it as a proxy for intermittency and sub-seasonal variability. A list of all predictors is shown in Table 2.1.

Some of these predictors are obtained from satellite products at high spatial and temporal resolution (e.g., surface chlorophyll and vertically integrated net primary production), and include missing values caused by the presence of clouds or sea-ice. For these variables, we first average observations into monthly climatologies, then replace missing data by using a temporal interpolation followed by a spherical interpolation using the algorithm by D'Errico (2016) as described in Yang et al. (2020). To avoid excessive extrapolation in high latitude regions in wintertime, only points with at least 8 months of satellite observations are used for the final reconstruction, following the approach of Siegel et al. (2014).

we apply an additional processing step to primary production to further minimize artifacts caused by the interpolation and gap-filling algorithms at high latitudes, and to provide a smooth transition towards the areas were light limitation (caused by deep mixing and polar night) is likely to significantly reduce production of new organic particles. To this end, we adopt a criterion based on Sverdrup's critical depth (Sverdrup, 1953) to determine regions where light likely becomes too limiting to support photosynthesis, based on the method of Siegel et al. (2002). we adopt this framework because of its simple applicability as a proxy for light limitation, recognizing that more complex interpretations exist on the interplay between light, vertical mixing, and grazing in limiting primary production at high latitudes (Behrenfeld & Boss, 2014). Accordingly, we calculate the Sverdrup critical depth based on climatological chlorophyll concentration and incident shortwave radiation (Siegel et al., 2002). When the mixed layer depth exceeds this critical depth, we assume that phytoplankton spend too much of their life cycle in light-limited layers, making net productivity negligible. Net primary production is thus set to zero at all points where, in a given month, the mixed layer depth exceeds the critical depth, before interpolating. Similarly, we restrict these reconstructions of high latitude PSD based on seasonal sea ice cover

from ERA5 reanalysis, recognizing that regions with over 30% sea ice coverage, by area, cannot be captured robustly by this approach. Because we limit the final calculations to regions with at least 8 continuous months of satellite data (Siegel et al., 2014), thus removing significant portions of the polar regions in wintertime, these steps do not significantly affect thefinal reconstructions.

we use two different depth-dependent averaging procedures to generate two-dimensional predictor fields from three-dimensional variables, such as temperature. we generate a "surface" predictor by taking the average of the variable over the mixed layer, and a "sub-surface" predictor by taking the average from the base of the mixed layer to 100 m below it. For surface-only variables (e.g., chlorophyll, net primary production) and nutrients we also include predictors that quantify the change of the variable over time, because time variability (e.g., blooms in chlorophyll) could also be related to export flux. In practice, we calculate the time derivative of each variable by taking the difference between the month of observation and the prior month. we refer to these depth- and time-change variables as "variations" in Table 2.1. we test the significance of each predictor, including vertical and time variations, with the recursive feature elimination. Finally, we group predictors into different categories, with variations for selected variables (Table 2.1). If a predictor is in the "universal" category in Table 1, it is always included in all RF realizations. For all other categories, only one predictor is randomly chosen for each realization, but if a predictor is chosen, all variations are included too. After processing, all predictors consist of monthly climatological two-dimensional fields.

The global data sets used as predictors are characterized by variable degrees of uncertainty and errors. Typically, for the climatological mapped fields used here (e.g., from the World Ocean Atlas) much of the final uncertainty does not depend on specific instrumental errors, but rather on the interpolation method and the temporal averaging steps adopted. we refer readers to the specific references for details on the uncertainty of each product (Table 1); further details on biases in satellite products can be found in K. Bisson et al. (2021). we do not make any specific attempt to directly propagate errors in the predictors through these reconstructions. Rather, we rely on the standard deviation of the large ensemble (100) of

Table 2.1: Variables used to predict PSD parameters, variations (i.e., vertical or temporal changes) and data sources. The categories are organized based on predictor type, where universal predictors are used in every Random Forest realization.

Category	Variable	Variations	Source
Universal	Topography		N.G.D.C (2006)
	Temperature below MLD	Time Derivative	Locarnini et al. (2019)
	Chlorophyll	Time Derivative	NASA G.S.F.C (2014)
	Oxygen	ML/ ML+100m Time Derivative	H. E. Garcia et al. (2019)
	Shortwave Radiation	Time Derivative	Copernicus C.C.S (2017)
	Nitrate	ML/ ML+100m Time Derivative	H. Garcia et al. (2018)
	Phosphate	ML/ ML+100m Time Derivative	H. Garcia et al. (2018)
	Salinity	ML/ ML+100m	Zweng et al. (2019)
Mixed Layer	Mixed Layer	Time Derivative	Johnson et al. (2012)
	Mixed Layer	Time Derivative	de Boyer Montégut et al. (2004)
Primary Production	Eppley VGPM	Time Derivative Standard Deviation	Antoine & Morel (1996)
	VGPM	Time Derivative Standard Deviation	Behrenfeld & Falkowski (1997)
	CBPM	Time Derivative Standard Deviation	Westberry et al. (2008)
	CAFE	Time Derivative	Silsbe et al. (2016)
Euphotic Zone Depth	VGPM		Morel et al. (2007)
	CBPM		Morel et al. (2007)
Iron	Soluble Iron	Time Derivative	Hamilton et al. (2019)
	Labile Iron	Time Derivative	Myriokefalitakis et al. (2018)

RF realizations as a measure of uncertainty. Furthermore, the use in each RF of random permutation of the predictors, and, when possible, different data sets for the same variable, likely reduces some of the biases inherent to any specific data product.

The predictors are used to reconstruct PSD slope and total particle biovolume at the climatological euphotic zone depth. Each prediction is based on the ensemble average of 100 RF realizations with variable hyper-parameters (the number of trees and their complexity), with the inter-model spread representing the error. Each RF realization uses a total of 29 predictors randomly chosen from the categories listed in Table 2.1. By generating an ensemble of 100 RFs for each reconstruction, with varying hyper-parameters and predictors, we reduce biases and overfitting, making the results robust with respect to parameter tuning and the choice of different observational products. Thus, these reconstructions are not the result of tuning the hyper-parameters, or choosing only the best predictors. we evaluate the overall robustness of the predictions by reporting goodness-of-fit statistics that include the correlation coefficient, the root mean square error (RMSE), and the average bias, calculated by comparing predictions to *in situ* data.

2.3 Results and Discussion

2.3.1 Particle size distribution reconstructions

Figs. 2.3 and 2.4 show the global reconstructions of PSD biovolume and slope. This reconstruction method is able to capture most of the variability of the UVP5 observations, and robustly reproduce the gridded measurements, with global average values of 0.6 ppm for biovolume (r^2=0.91) and 3.9 for slope (r^2=0.86) when considering the entire data set. Observations that are not used in the training (out-of-bag) provide a more stringent test for the method's robustness. As shown in Figs. 2.3d and 2.4d, these out-of-bag observations are also robustly predicted, with a RMSE of 2.1 ppm for biovolume (r^2=0.74) and 0.33 for slope (r^2=0.68). Relative to both the full data set and the out-of-bag observations, these reconstructions show negligible biases. That is, there is an overall compensation between data

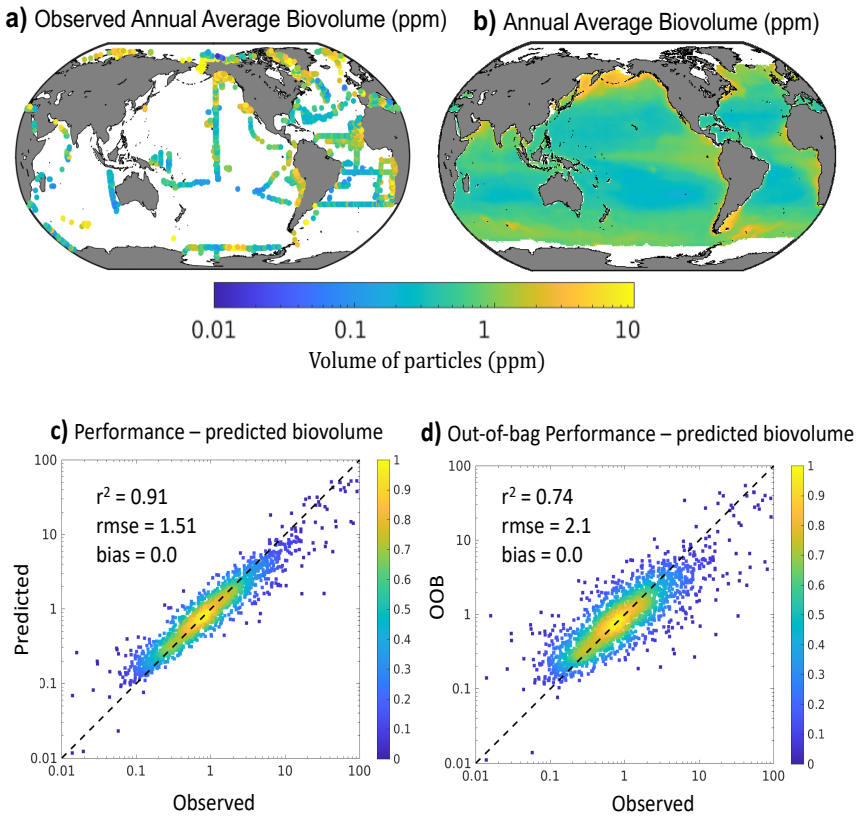

Figure 2.3: Observed and reconstructed particle biovolume (in parts per million, ppm) at the base of the euphotic zone. (a) Observed average biovolume. (b) Annual mean biovolume reconstructions. (c) Performance of the RF reconstruction shown as density scatter plots of predicted vs. observed biovolume (colors indicate the normalized density of observations at each point). (d) Same as (c), but using out-of-bag (OOB) predictions, i.e., predictions vs. observations withheld from training. Annotations in (b) and (c) show the coefficient of determination (r^2), the rmse, and the global bias.

points where this method overestimates observations, and data points where this method underestimates them.

While most observations are generally accurately reproduced, there remains a degree of uncertainty in the reconstructions, as shown by the scatter around the one-to-one line in Figs. 2.3c,d and 2.4c,d. Some of this remaining uncertainty could be explained by the episodic nature and patchiness of particle production transport, and export, and by factors not cap-

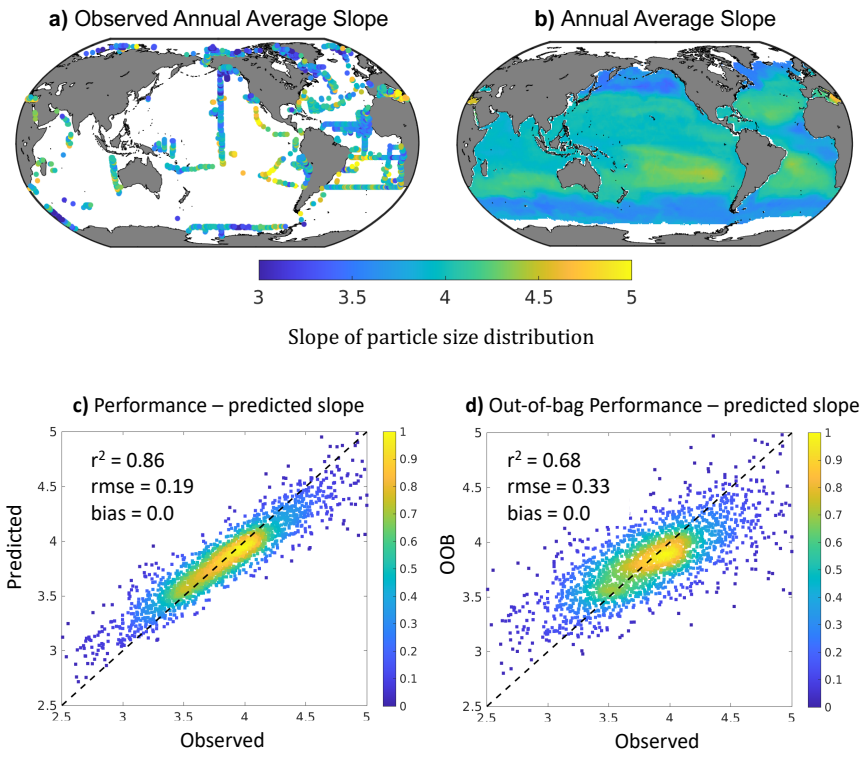

Figure 2.4: Observed and reconstructed PSD slope at the base of the euphotic zone. (a) Observed average PSD slope. (b) Annual mean PSD slope reconstructions (c) Performance of the RF reconstruction shown as density scatter plots of predicted vs. observed particulate slope (colors indicate the normalized density of observations at each point). (d) Same as (c), but using out-of-bag (OOB) predictions, i.e., predictions vs. observations withheld from training. Annotations in (b) and (c) show the coefficient of determination (r^2), the rmse, and the global bias.

tured by the chosen climatological predictors. This method operates under the assumption that the input data (i.e., the UVP5 observations) consists of monthly climatological averages, rather than instantaneous snapshots. By ensembling *in situ* UVP5 measurements into 2,034 monthly data points, we reduce part of the episodic nature and patchiness of these observations; however variability is likely to still persist in the gridded data. Finally, while the mean bias is zero, the reconstructions show a slight underestimate of extreme values at

both the high and low range of the observations, i.e., these reconstructions have a slightly reduced range compared to observations (Figs. 2.3c,d and 2.4c,d). This slightly reduced range in the reconstructions is typical for bagged ensemble ML methods such as the RF used here, which results in a limited ability to extrapolate data and tends to smooth out extreme values (Zhang & Lu, 2012). Spatially, this bias would result in an overestimate of biovolume in the subtropics, and an underestimate in mid to high latitude regions, and vice versa for the PSD slope. we discuss the consequences of this potential range reduction in Section 2.3.5.

2.3.2 Global patterns in particle size distribution

These reconstructions of the PSD for the time frame 2008 to 2020, reveal high biovolume in productive regions such as high latitudes, coastal waters, and upwelling systems, and low biovolume in the oligotrophic subtropical gyres (Fig. 2.3b and Supplementary Fig A.3). PSD slopes show a nearly opposite pattern, with smaller slopes (i.e., "flatter" PSD) in more productive regions, and larger slopes (i.e., "steeper" PSD) in oligotrophic waters (Fig. 2.4b and Supplementary Fig A.4), although with somewhat less pronounced variations compared to biovolume. Consistent with this, we find that slope and biovolume are negatively correlated ($r^2 = 0.4, p < 0.01$ Fig. 2.5a,b). Spatial patterns in biovolume and slope roughly follow the distribution of satellite-derived primary chlorophyll and primary production estimates, suggesting that phytoplankton and photosynthesis exert a strong control on total abundance of particles in any given region (Kostadinov et al., 2009, 2017; Cram et al., 2018). Accordingly, we find a positive correlation between biovolume and surface chlorophyll (R=0.49 for observations, and R=0.68 for reconstructions, both with p¡0.01, Fig. 2.5a,b) and a negative correlation for slope (R=-0.18 for observations, and R=-0.37 for reconstructions, both with p¡0.01, Fig. 2.5c,d)).

The negative correlation between particle biovolume and slope ($R = -0.40, -0.64$ Fig. 2.5e,f) indicates that particle-rich regions (higher biovolume) are also characterized by an excess of large particles over small particles (i.e., flatter slope), relative to average oceanic

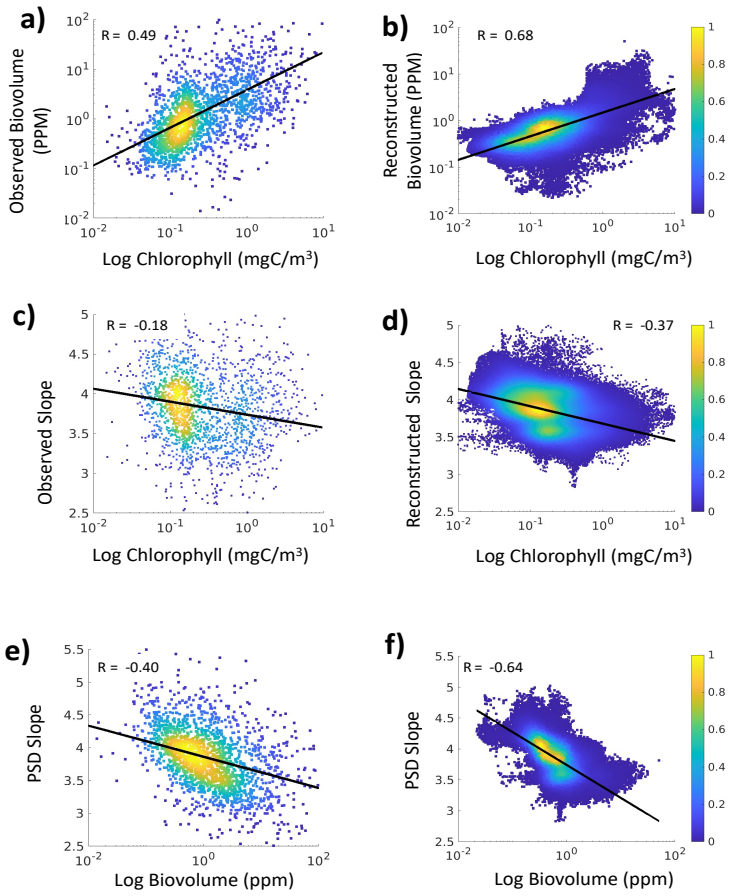

Figure 2.5: Relationships between PSD parameters and surface chlorophyll. (a,b) Relationship between PSD slope and chlorophyll for (a) observed and (b) predicted data. (c,d) Relationship between particle biovolume and chlorophyll for (c) observed and (d) predicted data. (e,f) Relationships between PSD slope and particle biovolume, for (e) observed and (f) predicted data. The black line in each panel shows a linear fit between the two variables, and R is the Pearson's correlation coefficient. All fits are significant to the 0.01 p-value.

conditions. Such co-variation between slope and biovolume is likely to be important for processes that depend both on the abundance of particles and their size, such as sinking particle fluxes (Guidi et al., 2008; Stemmann & Boss, 2012). While this pattern of correlations holds true for most regions, we find few significant exceptions where the PSD slope and biovolume do not co-vary as closely as expected. For example, in the North Pacific subpolar gyre, flatter slopes are found in the open ocean (Fig. 2.4b), in particular close to the subpolar-

subtropical transition, while the highest biovolumes are found closer to the coast and in marginal seas. Similarly, slopes in coastal upwelling systems, such as the California Current and the Arabian Sea upwelling, are not as flat as the high biovolumes would suggest. we also find relatively flatter slopes in the North Pacific subtropical gyre as compared to other oligotrophic regions.

These patterns suggest that while the partitioning between large and small particles typically reflects the strength of primary production, as previously noted (Stemmann et al., 2002; Stemmann, Youngbluth, et al., 2008), there are regions where the dynamics are more complex. Coastal upwelling regions are generally productive and exhibit high export (Bishop et al., 2016). However, according to this reconstruction, the California Current exhibits steeper slopes than expected, nearly matching the North Pacific subtropical gyre. It is possible that, in the coastal region, slopes are higher because of an increased number of large phytoplankton (Kostadinov et al., 2010a). Diatoms observed by the UVP5 could artificially inflate the particle abundance in the smaller size ranges, resulting in a lower slope. This could also be caused by reduced surface aggregation or effective disaggregation of particles, or less efficient surface remineralization, which tends to reduce small particles faster than large ones. Conversely, relative to other oligotrophic gyres, the North Pacific subtropical gyre may be characterized by somewhat larger phytoplankton cells, increased surface aggregation, and reduced disaggregation, or more efficient remineralization, especially because of the deep euphotic zone present in the region.

2.3.3 Seasonal variability in particle size distribution

The seasonal dynamics of biovolume and slope confirms the general anti-correlation of these two variables, and reveals significant seasonal cycles, with maximum biovolume and minimum slope generally found in spring, and minimum biovolume and maximum slope in late fall to winter (Fig. 2.6). Similar to the spatial distribution, we find significant deviations from the general anti-correlation between biovolume and slope. For example, in the North Atlantic, the peak in biovolume (May) precedes the minimum in slope (July). In some of the

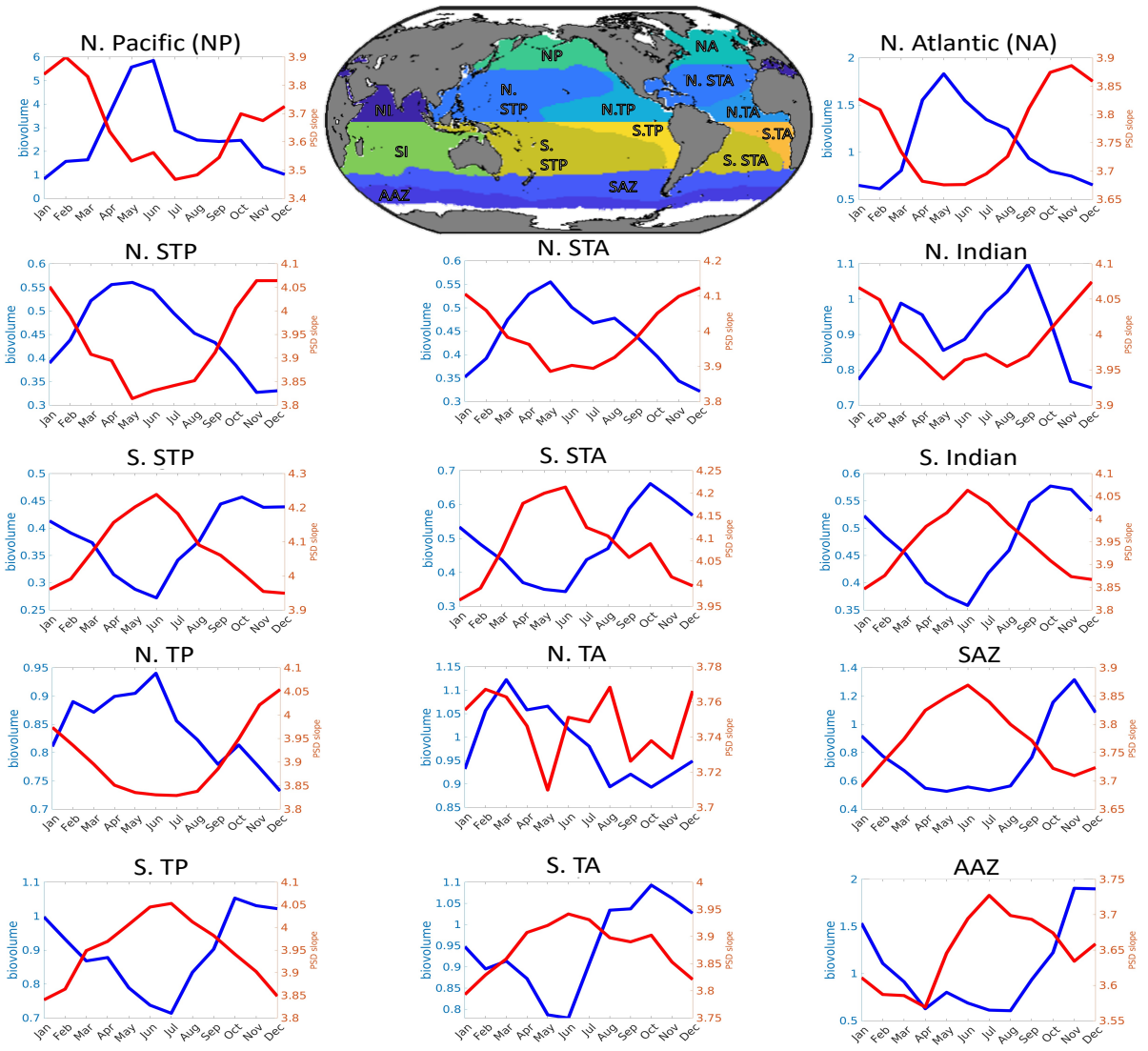

Figure 2.6: Annual seasonal cycle of particle biovolume (blue lines, in ppm) and slope (red lines) from the Random Forest reconstructions. Each seasonal cycle is from the euphotic zone for the regions specified on the map (top).

tropical regions (e.g., in the North Pacific and North Atlantic) the anti-correlation is also less robust, with periods of several months where biovolume and slope increase or decrease simultaneously. Several low biomass regions (e.g., the North Subtropical Atlantic, the South Subtropical and Tropical Pacific, the South Indian) display seasonal cycles in slope (and to a lesser extent biovolume) that are remarkably symmetrical across the months of June or December, suggesting a possible influence of the annual cycle of insolation. Other regions,

however (e.g., North Atlantic, South Subtropical Atlantic, Antarctic Zone), display more marked asymmetries, possibly reflecting seasonal variations in mixing (Behrenfeld & Boss, 2014).

In general, regions that show higher total biovolume and lower slopes also display higher seasonality. High latitude regions are characterized by large biovolume and flatter slopes, following the pattern of productivity for these waters. Conversely subtropical regions characterized by low biovolume also exhibit low seasonal variability. The interplay between variations in slope and biovolume could have important consequences for quantities that directly depend on PSD properties, such as sinking particle fluxes, which are enhanced by high biovolume and a relative abundance of large vs. small particles (i.e., flatter slopes). Accordingly, we expect higher seasonality in sinking particle fluxes in regions of strong anti-correlation between biovolume and slope, relative to regions where these quantities tend to co-vary. Likewise, we expect high seasonality in biovolume and slope to drive strong variations in sinking particle fluxes over the course of the year, for example in high latitude regions.

While several of the patterns discussed in this section may deserve further attention, validation with in situ data would require a greater number of observations than currently available. Furthermore, consideration should be given to biases in predictors, such as satellite products (K. Bisson et al., 2021), that could introduce spurious patterns in biovolume and slope. Specific caveats related to the seasonality of these reconstructions are discussed further in Section 3.5.

2.3.4 Empirical Drivers of PSD

A recursive feature elimination indicates that multiple variables are required for a robust reconstruction of PSD, as each one increases the ability of the reconstruction to explain observations (Supplementary Fig. A.2). Among the important features, we highlight chlorophyll, mixed layer depth, and oxygen, although each has a somewhat different importance for explaining biovolume and slope variability. Interpretation of these rankings should be

done with care because of the statistical nature of the RF algorithm. However, while a mechanistic understanding of PSD patterns cannot be directly tied to these rankings, highlighted predictors can provide insights into the role of different processes that may be affecting PSDs.

we find that biovolume at the base of the euphotic zone correlates positively and significantly with chlorophyll (R=0.49 for observations, and R=0.68 for reconstructions, both with p¡0.01, Fig 2.5a,b). This is not surprising, since chlorophyll is an indicator of phytoplankton, the main source of organic matter and sinking particles in the ocean (Stemmann et al., 2002). However, we find that chlorophyll is not as strong a predictor of slope, when the whole ocean is considered (R=-0.18 for observations, and R=-0.37 for reconstructions, with p¡0.01, Fig 2.5c,d), and that additional predictors are needed for robust slope reconstructions. This result reflects previous findings based on UVP5 observations along a meridional section in the Pacific Ocean (Cram et al., 2018). Slope reconstructions also reveal a significant predictive power for subsurface oxygen. Previous work suggests a connection between oxygen and total particle concentration (Roullier et al., 2014), whereby particle concentrations increase as oxygen decreases. Oxygen is a proxy of respiration in the water column, which in turn reflects the characteristics of both the surface community that drives export, and of the subsurface community responsible for this respiration (Sarmiento & Gruber, 2006a). we note that the PSD slope is an emergent property that reflects the interaction of physical and biological processes that are still poorly understood.

Spatial patterns in slope and biovolume share several features with estimates of particulate backscattering, and phytoplankton size spectra and composition from observations and models (Kostadinov et al., 2009; Roy et al., 2013; Barton et al., 2013; Ward et al., 2014). Regions with higher biovolume and flatter slope are dominated by larger phytoplankton, while the subtropics, with lower biovolume and steeper slope, are dominated by smaller phytoplankton (Kostadinov et al., 2009; Mouw et al., 2017). The composition and size structure of phytoplankton can be linked mechanistically to the size of particles and aggregates in the upper ocean (Jackson, 1990; Kiørboe et al., 1990; Burd & Jackson, 2009). Large cells, for example chain-forming diatoms, can more easily aggregate to form large phytodetritus particles. More indirectly, phytoplankton composition and size structure exert an important

control on the size structure of zooplankton and the upper ocean food web, thus of affecting the abundance and size structure of fecal pellets and other aggregates that are the byproduct of zooplankton feeding (J. T. Turner, 2015).

Phytoplankton functional groups (e.g., Mouw et al. (2017)) and abundance should be considered as important controlling factors on both biovolume and slope (Guidi et al., 2009; Stemmann et al., 2002), and could be used as predictors alongside other physical and biogeochemical variables. However, methodological shortcomings and disagreement between different approaches (such as satellite based retrievals) currently limit the applicability of these datasets—something that may be mitigated by future advances. It is also likely that information related to phytoplankton composition and size structure retrieved from satellite implicitly enters the RF regression via relationships with environmental predictors such as satellite retrieved surface chlorophyll and temperature (Kostadinov et al., 2017; Mouw et al., 2017).

2.3.5 Caveats to this approach

While the global data set of UVP5 observation enables robust global reconstruction of PSD properties, there remain sources of uncertainty and inherent limitations that could affect these estimates and call for further work. First, expanding the coverage of observations with UVP5 and similar instruments, in particular in under-sampled regions characterized by large variability, such as coastal and high latitude regions, would improve the robustness of these estimates, and shed additional light on regional particle size distribution patterns not captured by previous work. Regional correlations between environmental properties and PSD may not be well captured by extrapolation with a RF algorithm trained on data from different regions, especially when non-linear relationships between variables are important.

These reconstructions also rely on a two-parameter power law approximation to describe the observed PSD. While this is a common and useful assumption for a first order description of PSD in the ocean (Bader, 1970; Sheldon et al., 1972; Stemmann & Boss, 2012), significant deviations from a power law have been reported (Organelli et al., 2020;

R. A. Reynolds & Stramski, 2021b), in particular in the micrometer range mostly representative of phytoplankton (Cavender-Bares et al., 2001; Huete-Ortega et al., 2014) and for productive coastal waters and regions affected by mineral inputs from rivers and sedimentary exchanges ((R. A. Reynolds et al., 2010, 2016). While data tests suggest that a power law assumption is globally robust for the range of sizes sampled by UVP5 instruments (Fig. S1), other statistical models, e.g., based on different distributions (Jonasz & Fournier, 1996) or non-parameteric descriptors (R. A. Reynolds & Stramski, 2021b) may be more appropriate, and could result in somewhat different patterns of PSD and biovolume globally. we also suggest that new work could exploit deviations from power laws to shed further light on PSD patterns and particle cycling mechanisms (Weber & Bianchi, 2020).

Furthermore, we do not test how well reconstructed PSD slope translates to particles smaller or larger than the range robustly sampled by the UVP5, which may be possible by combining UVP5 observations with other optical instruments (Karp-Boss et al., 2007; R. A. Reynolds et al., 2010; Stemmann & Boss, 2012; Boss et al., 2015; Lombard et al., 2019). Lastly, these results assume that particles observed by UVP5 instruments are largely biogenic. While this assumption generally holds for open ocean regions (Sheldon et al., 1972), care should be taken when applying these reconstructions in coastal regions where inputs of mineral particles from rivers and sediment may be important (R. A. Reynolds et al., 2010, 2016).

Supervised learning methods are only as reliable as the data used for training, and it is possible that some of the patterns in these reconstructions may be influenced by biases in the predictors utilized. However, it should be noted that machine learning algorithms, provided that sufficient data is available for training, will learn patterns from predictors only if they allow to robustly reconstruct the observations. Therefore, even biased predictors could lead to accurate predictions. However, this may also introduce spurious correlations between predictors and predicted variables. This is especially the case for remotely sensed variables, which have inherent seasonal biases that could limit the ability to reconstruct seasonal cycles and interpret correlations, and have a greater inherent error compared to other features used for the reconstruction (i.e., temperature) (K. M. Bisson et al., 2021). Therefore, continued

work on improving satellite reconstructions of surface chlorophyll, net primary production, and other remotely-sensed variables, in particular at high latitudes, would help improve the robustness of these methods. Reducing biases in satellite products is an urgent undertaking (K. M. Bisson et al., 2021).

Similar to previous work (Siegel et al., 2014; DeVries & Weber, 2017), limiting this reconstruction to regions with more complete satellite coverage leads us to underestimate export in polar regions at times of the year characterized by significant sea ice coverage and pervasive light limitation (Siegel et al., 2002). Although based on *in situ* measurements some level of particle production and export is likely to occur in these regions and times of the year (e.g., Lowry et al. (2018); K. Bisson & Cael (2021); Hague & Vichi (2021)), we lack both remote sensing and UVP5 observations that would allow robust estimates of particle abundance and size under such conditions. Future work should be devoted to closing these gaps.

Some variables that are known to be mechanistically linked to particle production are not considered important by the Random Forest method. For example silicate, which could serve as a proxy for diatom biomass or production, did not significantly reduce the error when included, and thus were excluded from the final reconstructions (Supplemental Figure. A.2). It is possible that this Random Forest method is biased to select only few of highly correlated variables, even if other features are mechanistically important (Nicodemus et al., 2010).

Lastly, different machine learning approaches are likely characterized by different biases. Here, we note a slight underestimate of extreme values in reconstructed PSD properties, which may affect the reconstructed variability in particle size spectra (Zhang & Lu, 2012). Different machine learning methods (i.e. Artificial Neural Networks, Boosted Forests, etc.) have been used to reconstruct particulate matter in the surface ocean (Liu et al., 2021). Adoption of additional machine learning algorithms in conjunction with increased data coverage may eventually reduce errors. Additionally, increasing number of measurements, more detailed analyses of particle size spectra distribution, including at time-series stations, and spatial clustering techniques, may allow reconstruction of interannual variability (Gregor &

Gruber, 2021).

2.4 Conclusions

In this chapter, we provide a new, data-constrained estimate of particle size spectra based on global UVP5 observations obtained between 2008 and 2020. It captures regional and seasonal variability in observed PSD properties, and demonstrates the ability of statistical machine learning methods to extrapolate these quantities globally. These global PSD reconstructions can in turn shed light on processes that depend on, or are reflected by, particle abundance and size distribution, including sinking particle fluxes (Guidi et al., 2008), aggregation, disaggregation and degradation of organic particles (Burd & Jackson, 2009; Briggs et al., 2020), interactions with microbial communities (DeLong et al., 1993; Church et al., 2021) and migrating animals (Cram et al., 2022), and chemical exchange between particles and seawater, including element scavenging (Ohnemus et al., 2019).

The statistical nature of the machine learning approach does not directly reveal mechanisms behind particle abundance and size structure. However, we are able to highlight spatially coherent patterns, and the seasonal variability of particle abundance and size structure. Specifically, we show that the total particle biovolume and the PSD slope are characterized by similar but inverse patterns, with regions of high particle biovolume generally characterized by flatter slopes, i.e., relatively more abundant large particles. Similarly, the seasonal cycle of the particle slope and biovolume are inversely correlated over time through most of the ocean. Importantly, because of this anti-correlation, biovolume and slope variations would act synergistically on sinking particle fluxes, by enhancing them in region of higher biovolume and flatter slope, and reducing them in regions of low biovolume and steeper slope. we also show that biovolume and slope tend to correlate with observed sea surface chlorophyll and other biogeochemical variables. Specifically, regions of high chlorophyll tend to be characterized by higher particle biovolume and flatter slope, highlighting the important role for primary production and phytoplankton size structure for particle abundance and size distribution at the lower limit of the euphotic zone. These findings in turn support obser-

vational programs with the goal of linking remotely sensed surface variables to subsurface properties and biological processes (Siegel et al., 2021).

UVP5 and other optical observations are not limited to the surface ocean, but are generally highly resolved in the vertical direction, thus enabling fully three-dimensional reconstructions of PSD. This could allow a closer investigation of particle dynamics in the water column, a better diagnosis of the processes that cause deviations of PSD from a simple power law, and enable three-dimensional reconstructions of size-dependent processes such as particle sinking fluxes. Enhanced deployments of UVPs—also on Argo floats (Picheral et al., 2022)—combined with the approaches developed in this paper could also enable decadal or even annual estimates of global PSD and particle flux through the water column. Ultimately, a three-dimensional view of particle abundance and size distribution in the ocean would shed light on an essential component of ocean biogeochemistry and ecosystem, and inform new models of the ocean's biological pump.

CHAPTER 3

New estimate of organic carbon export from optical measurements reveals the role of particle size distribution and export depth

3.1 Introduction

At the ocean surface, primary production and other biogeochemical processes interact to form organic particles that drive the ocean's biological pump (Volk & Hoffert, 1985; Honjo et al., 2008; J. T. Turner, 2015; Siegel et al., 2022). Aggregation and sinking of particulate organic matter stores inorganic carbon and nutrients in the deep ocean for timescales ranging from decades to centuries (DeVries et al., 2012; Boyd et al., 2019), reducing surface carbon concentrations and leading to a decrease in atmospheric CO_2 (Kwon et al., 2009). Sinking particles provide organic matter sustaining the deep ocean biosphere (Robinson et al., 2010) and shape the ocean's microbiome (Karl et al., 1984; Fontanez et al., 2015; Bianchi et al., 2018).

Export of particulate organic matter results from the interaction of complex physical and biological processes (J. T. Turner, 2015; Boyd et al., 2019; Siegel et al., 2022). Gravitational settling of particles denser than seawater, including fecal pellets, phytodetritus, and heterogeneous aggregates, is thought to be the primary export mechanism, contributing to about 60% of the total carbon export, and more than half of the carbon storage in the deep ocean (Boyd et al., 2019). Other export processes, such as organic matter transport and repackaging by vertically migrating organisms (Longhurst et al., 1990; Steinberg et al., 2000; Bianchi et al., 2013) and physical injection (Carlson et al., 1994; Omand, D'Asaro, et al.,

2015; Stukel et al., 2017; Dall'Olmo et al., 2016), make up the remainder (Boyd et al., 2019). The importance of large sinking particles in driving the export flux has been well described (Honjo et al., 2008; A. L. Alldredge & Gotschalk, 1988; J. T. Turner, 2015), although several studies have also highlighted the importance of smaller aggregates (Alonso-González et al., 2010; Durkin et al., 2015; Kiko et al., 2017; Richardson, 2019).

Export fluxes can be quantified at different depth horizons, with the euphotic zone and mixed layer depths as common choices, underlying competing interpretations: export from the euphotic zone provides an ecosystem-level viewpoint, while export from the mixed layer provides an estimate of long-term carbon storage. Observational and model-based estimates generally evaluate export at the base of the euphotic zone, as defined by the 1% or 0.1% light levels (K. O. Buesseler & Boyd, 2009; Siegel et al., 2014; K. O. Buesseler et al., 2020). On annual timescales or longer, organic carbon export balances net community production (Emerson, 2013), and, since synthesis of new particles is greatly reduced below the euphotic zone, it also provides an upper limit to the energy that can fuel subsurface ecosystems. Meanwhile, carbon exported below the maximum mixed layer depth is removed from contact with the atmosphere for timescales longer than a year, and thus is relevant for ocean carbon sequestration. Recent work with a global biogeochemical model indicates that the magnitude and patterns of carbon export are sensitive to the choice of depth horizon (Palevsky & Doney, 2018). However, tests of this sensitivity based on global observations are missing.

Because of its central role in ocean biogeochemistry, the global particle export has received significant attention, resulting in a wide range of estimates – from less than 3 to more than 10 PgC y^{-1} (Henson et al., 2011; Siegel et al., 2014; DeVries & Weber, 2017; Dunne et al., 2007), with some of the discrepancies depending on the methods used (Quay et al., 2020). Biogeochemical models yield a global export of 4-6 PgC y^{-1} when tuned to match particle flux observations (Siegel et al., 2014), but can reach up to 10 PgC y^{-1} when tuned to match *in situ* profiles of nutrients and other biogeochemical tracers (DeVries & Weber, 2017). A similar range is suggested by recent global IPCC-class Earth System Models, which produce global carbon exports from 2.4 to 12 PgC y^{-1}, with an average of 7.4 PgC y^{-1} (Séférian et al., 2020). Data-driven estimates that combine satellite-based primary production with

empirical measures of particle export ratios often result in fluxes near the upper range (Dunne et al., 2007; Laws et al., 2011; Guidi et al., 2015), with some exceptions (Henson et al., 2011).

A global export of around 10 PgC y^{-1} is comparable to biogeochemical estimates of annual net community production in the mixed layer (Emerson, 2013; Quay et al., 2020). However, on long timescales, community production must be balanced by multiple export processes (Boyd et al., 2019; Siegel et al., 2022) that also include subduction of non-sinking organic carbon (Carlson et al., 1994; Dall'Olmo et al., 2016) and export via vertical migrations of zooplankton and fish (Longhurst et al., 1990; Steinberg et al., 2000; Bianchi et al., 2013). Using an euphotic viewpoint, and considering only gravitational settling, particle flux estimates have begun to converge on a value of 6 PgC y^{-1}, although with significant uncertainty (Boyd et al., 2019).

In the field, sediment traps and thorium-234 measurements have been used to quantify sinking particle fluxes. However, both types of observations lack detailed particle size information, vertical resolution, and have known biases, making extrapolations to global scales difficult (K. Buesseler et al., 2007; Le Gland et al., 2019). Recently, optical methods have gained traction to estimate particle export. These methods are based on *in situ* observations of particle size distribution (PSD), i.e., the number of particles, or abundance, as a function of size (Guidi et al., 2008; Bourne et al., 2019). Among optical instruments, the Underwater Vision Profiler 5 (UVP5) measures the abundance of particles in the 80 μm - 2.6 cm range (Picheral et al., 2010) and is routinely deployed on oceanographic expeditions (Kiko et al., 2022). The high vertical resolution of UVP5 observations, combined with empirical, size-dependent relationships for carbon content and sinking speed (Kriest, 2002; Stemmann et al., 2004a; Guidi et al., 2008), enables a uniquely detailed view into the three-dimensional ocean particle flux (Guidi et al., 2016). Observations from UVP5 have been used to quantify particulate fluxes from the surface ocean on a regional basis (Kiko et al., 2017; Cram et al., 2018; Forest et al., 2012), and to reconstruct carbon export across large-scale biomes based on limited sets of measurements (Guidi et al., 2015).

The growing number of UVP5 observations, their global distribution, high vertical resolution, and ability to resolve multiple particle size classes offer an unprecedented opportunity

to re-evaluate global carbon fluxes from the ocean's surface, testing the importance of the choice of depth horizon, the role of small vs. large particles, and the degree of autotrophy (i.e., net particle production) vs. heterotrophy (i.e., net particle consumption) across the euphotic zone. In this study, we use a global reconstruction of PSDs from UVP5 observations (Clements et al., 2022) to provide a new estimate of the magnitude and patterns of particulate carbon export from the ocean's surface. The approach relies on empirical relationships that relate particle size and abundance to sinking fluxes (Kriest, 2002; Guidi et al., 2008; Kiko et al., 2017), which we tune against a global data set of *in situ* sediment trap and thorium-derived particle flux observations (K. M. Bisson et al., 2018). we exploit the high vertical resolution of UVP5 measurements to estimate particle export at both the climatological euphotic zone depth and the maximum mixed layer depth, elucidating the importance of the export horizon for net carbon export and sequestration.

The rest of the paper is organized as follows. Section 2 describes the methods used to estimate particle fluxes from global PSD reconstructions and *in situ* observations. Section 3 presents the results of the new export estimates, comparing them to prior work, and discussing the implications, uncertainties, and caveats inherent to our approach. Section 4 summarizes the main findings and future directions.

3.2 Methods

The flux of particulate carbon ($\phi, \frac{mgC}{m^2 day}$) at any given depth can be expressed as a function of three size-dependent quantities: the number (#) of particles of a given size, i.e., the PSD ($n(s), \frac{\#}{m^3 cm}$), the sinking speed ($w(s), \frac{m}{s}$), and the carbon content of each particle ($c(s), \frac{mg}{\#}$), according to the following equation (Guidi et al., 2008; Stemmann & Boss, 2012):

$$\phi = \int_{s_{min}}^{s_{max}} n(s) \cdot w(s) \cdot c(s)\, ds, \tag{3.1}$$

Here, s (cm) indicates the particle ESD, or size, and s_{min} and s_{max} the minimum and maximum size of particles considered for export. Following previous work, we assume that

the quantities in Equation 3.1 can be approximated by power laws that depend on particle size, each characterized by an intercept (the size-independent coefficient) and a slope (the exponent for size-dependence) (Stemmann & Boss, 2012):

$$n(s) = n_0 \cdot s^{-\beta} \tag{3.2}$$

$$w(s) = w_0 \cdot s^{\eta} \tag{3.3}$$

$$c(s) = c_0 \cdot s^{\zeta}, \tag{3.4}$$

Thus, by using Equations 3.2-3.4, the total particle flux can be expressed as:

$$\phi = \int_{s_{min}}^{s_{max}} n_0 \cdot w_0 \cdot c_0 \cdot s^{-\beta+\eta+\zeta}\, ds = \int_{s_{min}}^{s_{max}} n_0 \cdot m_0 \cdot s^{-\beta+\mu}\, ds \tag{3.5}$$

where we combined the intercepts and exponents of the sinking speed and carbon content relationships by setting $m_0 = w_0 \cdot c_0$ and $\mu = \eta + \zeta$, following the approach by Guidi et al. (2008). we further approximate m_0 and μ with globally constant values, which we constrain with *in situ* observations. In practice, we calculate the continuous integral in Equation 3.5 as a discrete summation over the finite size bins that approximate the PSD observed by UVP5 instruments.

we use PSD properties (biovolume and slope) from a global UVP5-based reconstruction, shown in Figure 3.1 (Clements et al., 2022). Briefly, this reconstruction is based on a machine learning algorithm (a bagged Random Forest ensemble) applied to a global dataset of UVP5 observations (Kiko et al., 2022), and provides monthly varying climatological maps of PSD slope and biovolume in the upper ocean. we combine these PSD reconstructions with empirical relationships for sinking velocity and carbon content to estimate particle fluxes by solving Equation 3.5. Since the parameters that define the combined sinking speed and carbon content relationships, i.e., m_0 and μ, are poorly constrained (Kriest, 2002; Stemmann & Boss, 2012; Kiko et al., 2017), we optimized them by minimizing the mismatch between

predicted particle fluxes and *in situ* observations from sediment traps and thorium-uranium disequilibrium at the base of the euphotic zone (K. M. Bisson et al. (2018), see Section 3.2.1).

we exploit the three-dimensional nature of UVP5 observations to calculate particle fluxes at two different export horizons: the base of the euphotic zone (here defined by the 1% light level following Morel et al. (2007)) and the annual maximum mixed layer depth (Johnson et al., 2012). For the former, we take the PSD estimates from Clements et al. (2022). For the latter, we estimate the PSD at the base of the wintertime mixed layer, following the same procedure as Clements et al. (2022).

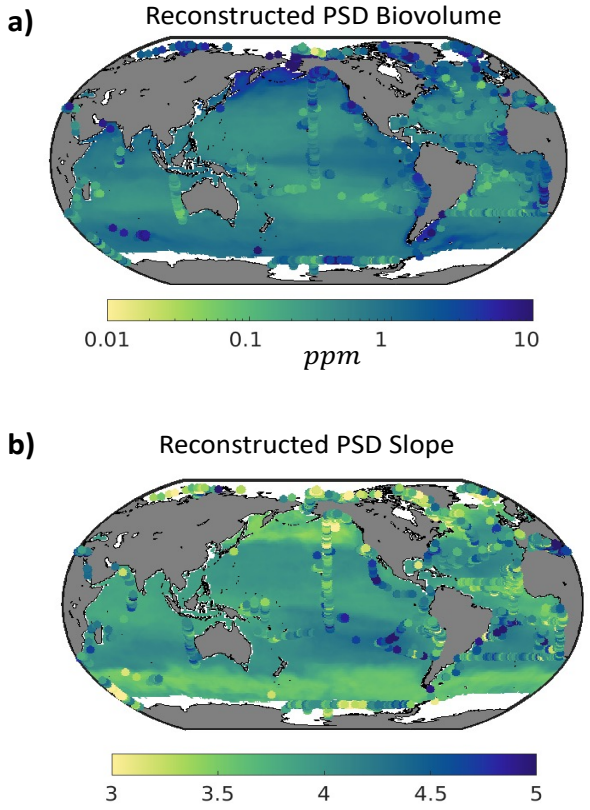

Figure 3.1: Global reconstructions of (a) PSD biovolume (ppm), and (b) PSD slope (non-dimensional), based on a machine-learning extrapolation of *in situ* UVP5 observations (Clements et al., 2022). Color contours show reconstructed variables as annual means. Dots show *in situ* quantities from UVP5 observations. Note that observations reflect specific months of the year, explaining some of the mismatches with annual mean quantities shown by the background colors.

3.2.1 Sinking Speed and Carbon Content

Particle sinking speed and carbon content have been empirically evaluated using power law relationships analogous to Equations 3.3 and 3.4, e.g., as compiled in Kriest (2002) and Stemmann et al. (2004a). Individually, most observational studies measure a range of particles that does not wholly encompass the sizes detected by the UVP5. Furthermore, these relationships are defined for specific particle types, which are not distinguished in the PSD reconstruction used here (Clements et al., 2022).

Since estimates of total flux are sensitive to the sinking speed and carbon content relationships encapsulated by the parameters m_0 and μ, we apply an optimization procedure to keep our results consistent with *in situ* particle flux measurements. Specifically, we find the values of m_0 and μ that minimize the sum of the square errors between the log of the particle flux reconstructions (Equation 3.1) and co-located *in situ* carbon flux measurements (K. M. Bisson et al., 2018). we use both trap and thorium flux data, corrected to be at the euphotic depth. we average together all *in situ* data onto the same grid of the PSD reconstructions (Clements et al., 2022), i.e., into 1 degree grids, by month, so that the optimizations are done on a climatological basis. Because *in situ* carbon flux measurements are uncertain (K. M. Bisson et al., 2018), we adopt a Monte Carlo approach for this optimization, repeating it 1000 times after perturbing each flux observation by applying a random observational error, assuming a log-normal distribution and an uncertainty of 1 standard deviation on the measurements. This Monte Carlo ensemble also allows us to estimate the error associated with the optimization of the sinking speed and carbon content parameters.

Because the size distribution of particles that contribute to the flux is poorly constrained, we perform this optimization for a range of plausible minimum and maximum sizes for Equation 3.5, selecting a physically reasonable combination for the final estimate. Ultimately, when optimizing the sinking carbon parameters, the total global export flux is not sensitive to the size range; however the resulting empirical relationships are. The insensitivity of the carbon flux to the size range indicates a compensatory effect between the sinking carbon parameters and the size range selected for the optimizations. Thus, choosing different size

combinations would result in a similar total flux, although it may slightly alter spatial or temporal patterns in a compensatory way (Supplementary Fig. B.1).

our final choice of size range is informed by average sinking speeds and carbon content previously reported (Kriest, 2002). Based on this optimization analysis, we set the minimum size class to be 35 μm, where the average sinking speed is near 1 m d^{-1} (Smayda, 1970; Kriest, 2002). Although this value is lower than the detection limit of the UVP5, the power law slope can likely be extended to this size range, as demonstrated for example by observations in the the Pacific Ocean (Stemmann, Eloire, et al., 2008). Most organic particles smaller than this size are likely rapidly remineralized, making their contribution to the sinking flux negligible (Riley et al., 2012). Even if some smaller particles could sink more rapidly (e.g., because of higher concentrations of mineral "ballast" and higher density) and could contribute more substantially to the total flux, neglecting them would not significantly affect our final export, because the optimized flux is nearly insensitive to the size range selected (Supplementary Fig. B.1). we choose 5 mm as the maximum size, i.e., the same maximum size used for the PSD reconstructions (Clements et al., 2022), roughly corresponding to the size where zooplankton become important contributors to the particle biovolume detected by UVP5 in a variety of regions (Forest et al., 2012; Stemmann, Youngbluth, et al., 2008; Stemmann & Boss, 2012).

Overall, this optimization approach results in a median value of 2.63 ± 0.06 for the exponent μ, and 18.0 ± 2.8 mgC m s^{-1} $cm^{-2.63}$ for the intercept m_0, both in the range suggested by *in situ* observations (Kriest, 2002), and comparable to values adopted by previous studies (Kriest, 2002; Stemmann et al., 2004a; Guidi et al., 2008; Kiko et al., 2017; Bianchi et al., 2018).

3.2.2 Flux reconstruction, error and evaluation

we first present results for fluxes estimated at the climatological euphotic zone depth, and then repeat the calculation at the maximum mixed layer depth. This requires an estimate of the PSD at the maximum mixed layer depth, which we obtain from UVP5 observations

following the same machine learning approach of Clements et al. (2022). we keep the same sinking speed and carbon content parameters (m_0 and μ), assuming that they do not change substantially between the two depths, which are often not too far apart from each other. Thus, the only methodological difference between the two estimates is the depth of the PSD reconstruction used to calculate the flux.

we take the ensemble mean of the Monte Carlo optimizations (Section 3.2.1) as the final carbon flux estimate. Error in this estimate could arise not only from the uncertainty in the particle sinking speed and carbon content parameters, but also from the uncertainty in the PSD reconstructions (Clements et al., 2022). we combine these two sources of error by summing the variances of two ensembles of carbon flux reconstructions. The first consists of the Monte Carlo optimization ensemble, based on the mean PSD from Clements et al. (2022). The second uses 100 different realizations of PSD from Clements et al. (2022), but sets m_0 and μ to the median values from the optimization. The final uncertainty is taken as the square root of the combined variances.

we evaluate reconstructed particle export fluxes by comparing them to *in situ* flux observations and previous global reconstructions. Specifically, we compare total fluxes, zonal averages, and seasonal cycles. For these comparisons, we divide the ocean into 14 biogeochemically consistent regions based on the boundaries identified by Weber et al. (2016), with an additional boundary along the equator to separate Northern and Southern Hemispheres.

3.3 Results and Discussion

3.3.1 Euphotic zone export fluxes

our resulting global carbon flux reconstruction at the base of the euphotic zone compares well with *in situ* sediment trap and thorium-based observations (Fig. 3.2), performing in a similar way as previous estimates (Henson et al., 2011; Dunne et al., 2007; Siegel et al., 2014). Compared to previous work, we reduce the uncertainty relative to observations, as expressed by the lower root mean square error and bias. However, our method also

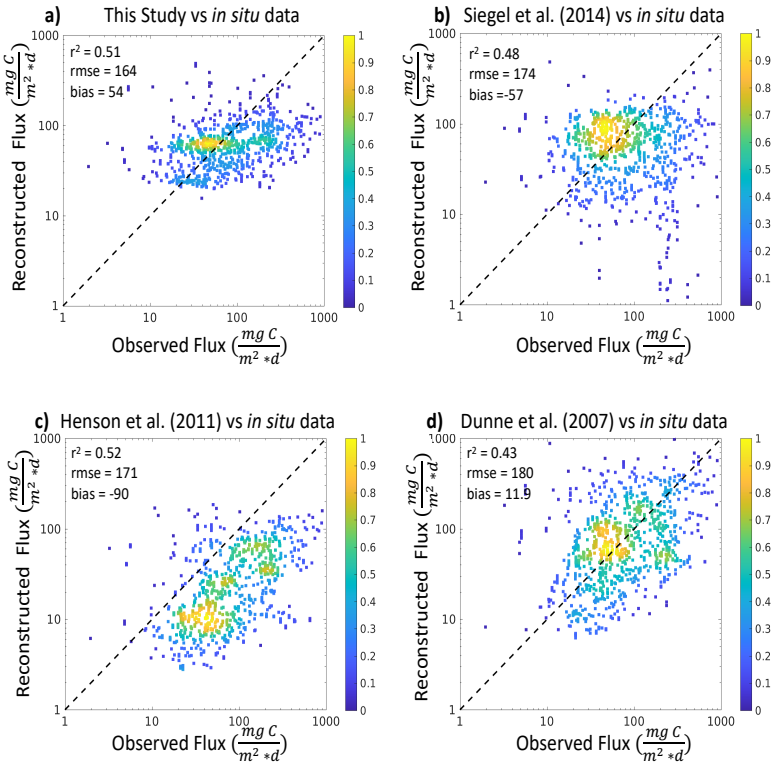

Figure 3.2: Density scatter plots showing the relationships between *in situ* flux observations and global flux reconstructions (mg C m^{-2} d^{-1}) at the base of the euphotic zone from (a) this study, (b) Siegel et al. (2014), (c) Henson et al. (2011), (d) (Dunne et al., 2007). Colored dots represent the relative density of grid points surrounding the data point, and the dashed line indicates a 1:1 ratio. Annotations show the coefficient of determination (r^2), root mean square error (RMSE), and average bias.

reduces the full range of reconstructed fluxes, i.e., it overestimates the flux at low values and underestimates it a high values compared to observations. This bias could be related to a similar underestimate of the range of PSD biovolume and slope that likely depends on the specific machine learning method used to extrapolate UVP5 observations (Clements et al., 2022). It is also possible that the optimization approach against an averaged global dataset of *in situ* fluxes fails to capture extremes in particle export at both the high and low range of observations.

Comparing sediment trap and thorium-based observations to the various estimates of

Fig. 3.2 highlights the relative strengths and weaknesses of each approach. The results from Dunne et al. (2007), based on combining satellite primary production with empirical estimates of particle export ratios, match the observed values well, but tend to overestimate the largest fluxes (not shown on these figure axes). The estimate by Henson et al. (2011), based on a similar approach as Dunne et al. (2007), follows a similar pattern as observations, as indicated by the high r^2, but systematically underestimates the flux magnitude, as shown by the negative bias. The satellite-driven, model-based estimate from Siegel et al. (2014) captures the overall magnitude of export, but misses some of the variability of observations, as indicated by the relatively low r^2. Overall, all estimates in Fig. 3.2 show combinations of strengths and weaknesses, and it would be difficult to highlight a specific model as unconditionally superior. This suggests that a combination of estimates should be used to asses export of carbon from the surface ocean, and that future efforts should strive to reduce the biases discussed above, potentially combining strengths from different approaches.

Extrapolated to the whole ocean, our method reveals spatial patterns of export fluxes in broad agreement with previous studies, with some notable differences (Fig. 3.3). Similar to other estimates, particle fluxes tend to decrease from high to low latitudes, and from coastal regions to the open ocean. A local maximum of export is reproduced along the equator, and is particularly evident in the Pacific Ocean. Compared to previous work, our method produces somewhat weaker gradients between coastal and offshore waters, with slightly higher fluxes near the centers of subtropical gyres, and suggests an asymmetry between the subpolar Atlantic and Pacific Oceans, with more intense particle export along the gulf of Alaska than in the North Atlantic (see also Section 3.3.1.1). we also reconstruct substantially stronger export than previously found in the Southern Ocean, in particular south of 50S (see discussion in Section 3.3.2).

Globally integrated, we estimate a particle export flux of 5.8 $\pm$ 0.1 PgC y^{-1}, in good agreement with the range of observational and model-based estimates of the biological gravitational pump (4-9 PgC y^{-1}, Boyd et al. (2019)). Compared to other spatially resolved reconstructions, our global flux sits between the low-value of Henson et al. (2011) (3.0 $\pm$ 0.3 PgC y^{-1}) and the high-value of Dunne et al. (2007) (8.5 $\pm$ 0.81 PgC y^{-1}). Seasonal maps

of the export and standard deviation are shown in Supplementary Figures B.2 and B.3.

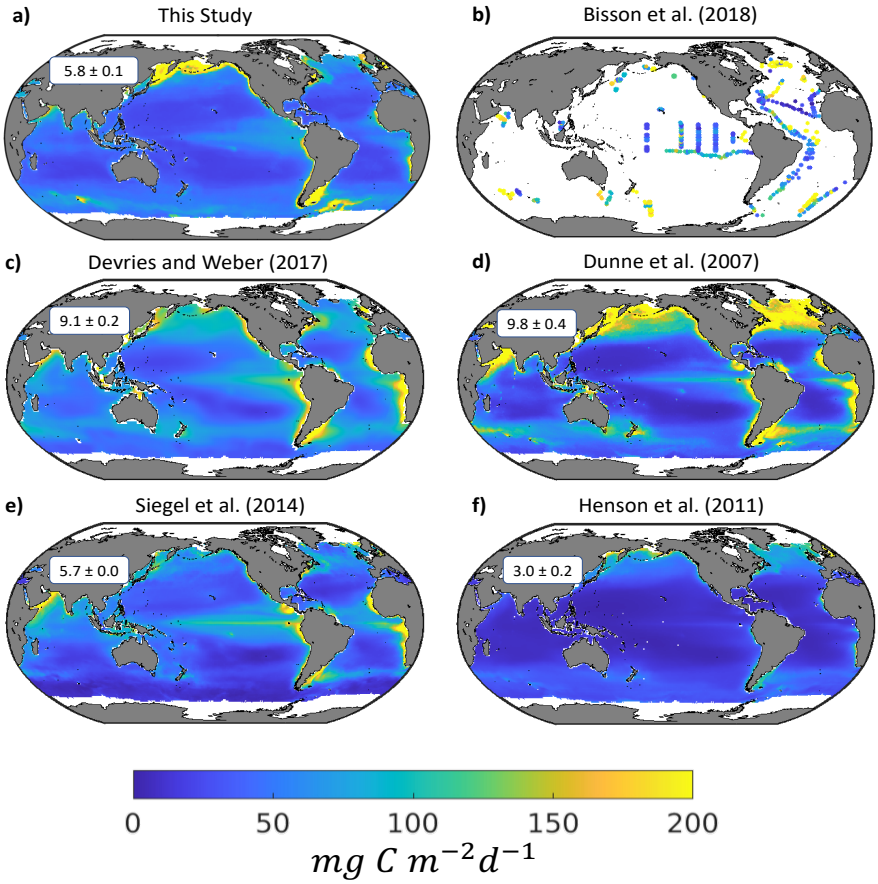

Figure 3.3: Annual average particle export flux (mg C m^{-2} d^{-1}) from the euphotic zone for (a) the global PSD-derived flux from this study, compared to (b) the *in situ* data of K. M. Bisson et al. (2018), (c) the steady state satellite-driven model SIMPLE-TRIM of DeVries et al. (2017), (d) the empirical model of Dunne et al. (2007), (e) the satellite-driven euphotic zone food web model Siegel et al. (2014), and (e) the empirical model of Henson et al. (2011). Annotations in each figure show the globally integrated export in Pg C y^{-1}, and the uncertainty reported by each study.

3.3.1.1 Spatial variability

Variations in export patterns derived with our approach (Equations 3.1 and 3.5) reflect a combination of spatially varying PSD biovolume and slope (Clements et al., 2022). we can

quantitatively describe the effect of the PSD as the relative contribution of small (35 μm - 418 μm) vs. large particles (418 μm - 5 mm) to the total flux (Figure 3.4), where 418 μm is the geometric mean of the size range considered here. Biovolume and PSD slope generally correlate in such a way that both factors contribute to increasing export fluxes in particle-rich productive waters, where large, rapidly sinking particles tend to be relatively more abundant than small particles, and to decreasing export fluxes in particle-poor oligotrophic waters where small particles dominate (Clements et al., 2022) (Figure 3.4).

High export in the eastern equatorial and tropical Pacific can be attributed to high biovolume, with a minor contribution from PSD slope, which appears to be more uniform across the region (Figure 3.1). The picture is somewhat different in the equatorial Atlantic Ocean, where a more substantial "flattening" of the PSD supports a higher contribution from large particles. A similar interaction of particle abundance and size-structure dramatically intensify fluxes at high latitudes, such as in the subpolar North Pacific and Southern Ocean, and to a lesser extent the subpolar Atlantic, where an increase in particle abundance is accompanied by a shift of the PSD toward large particles. In contrast, along many coastal regions, including eastern boundary upwelling systems and the Arabian Sea upwelling, increase in particle biovolume, rather than substantial changes in size structure, appears to drive enhanced export fluxes. we speculate that changes in community structure associated with more productive regions explain such a shift.

we illustrate the main spatial differences between our and other reconstructions by considering zonally averaged export fluxes (Fig. 3.5). The largest export rates are observed around the equator, in the subpolar Pacific Ocean, and in the mid- to high-latitudes of the South Atlantic Ocean, while more uniform export is observed in the Indian Ocean. In all basins, the minimum export rates are generally located at the latitude of the subtropical gyres. While export is nearly symmetrical around the equator in the Pacific Ocean (Fig. 3.5a), in the Atlantic Ocean it dramatically increases moving from the Northern to the Southern Hemisphere (Fig. 3.5b). These patterns reflect a combination of open-ocean and shelf enhanced particle fluxes. Specifically, high export in the Northern Pacific and Southern Atlantic Oceans are partly driven by large fluxes in the Bering Sea, the Sea of Okhotsk, and

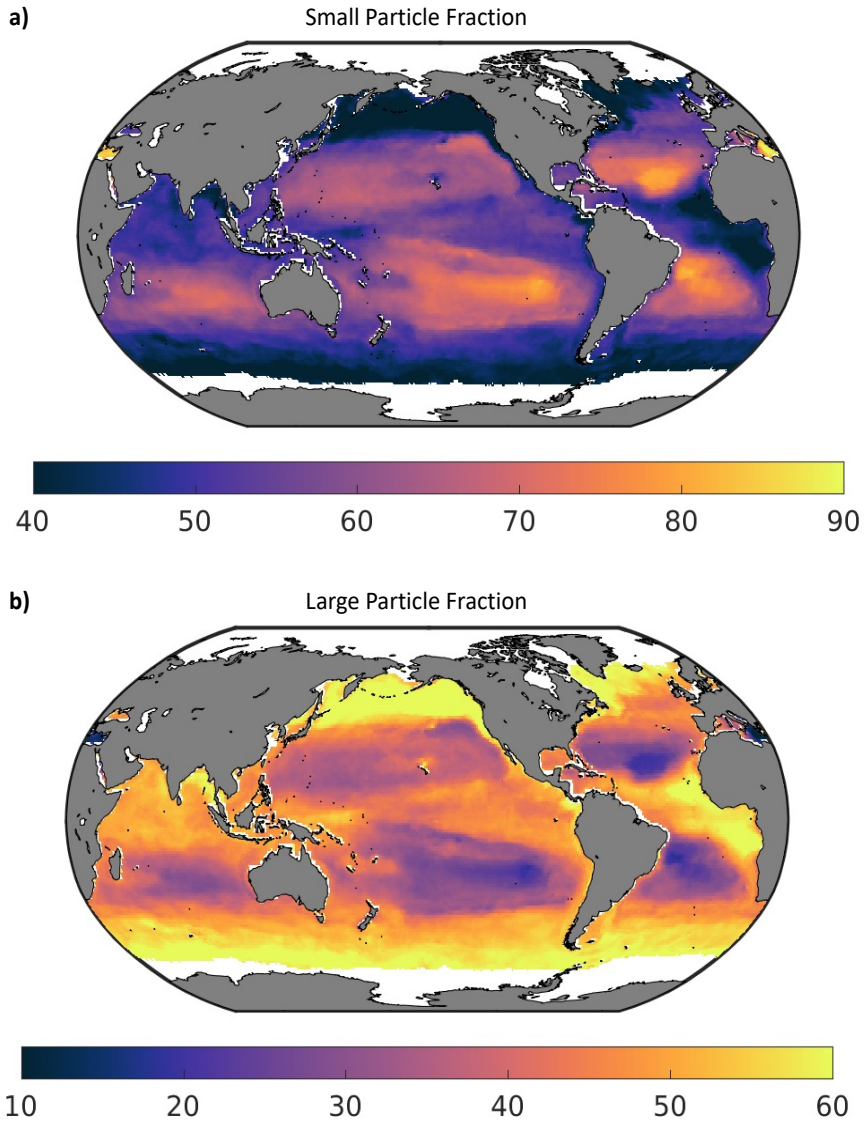

Figure 3.4: Role of small vs. large particles. The two panels show the fraction of carbon flux at the euphotic zone from (a) small particles ($35\mu m$ to $418\mu m$ ESD) and (b) large particles ($418\mu m$ to $5mm$ ESD).

the Patagonian shelf. At lower latitudes, coastal upwelling systems sustain particularly high export in the northern Indian Ocean and the tropical to subtropical Atlantic.

our reconstruction shows broad meridional patterns similar to previous estimates (Fig. 3.5); however, significant regional-level discrepancies remain. For example, in the low latitudes, we predict somewhat less intense equatorial export peaks and subtropical lows, com-

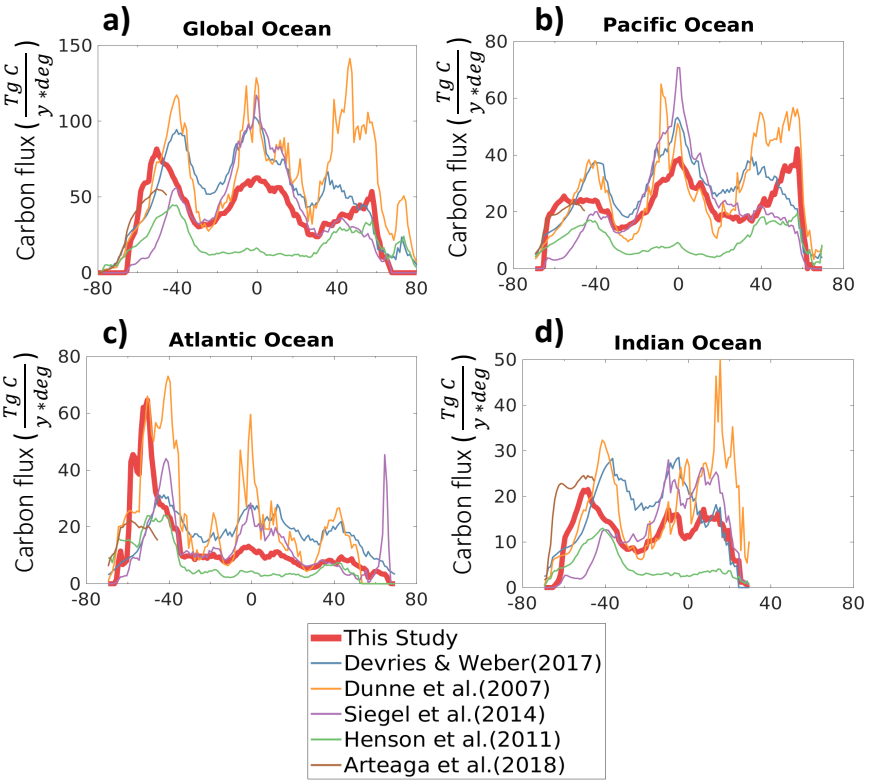

Figure 3.5: Zonally integrated annual mean export (Tg C y^{-1} $degree^{-1}$) from the base of the euphotic zone, for (a) the Global Ocean, (b) the Pacific Ocean, (c) the Atlantic Ocean, and (d) the Indian Ocean. Each color represents a different study, as shown in the legend (bottom).

pared to the estimates of Dunne et al. (2007) and Siegel et al. (2014). In this respect, our reconstruction is more in line with that of DeVries & Weber (2017). In the northern Pacific, we do broadly underestimate the transition zone as a persistent feature; however, seasonally, it is present (Supplementary figure B.2). Overall, in the subpolar North Pacific, our estimate shows a northward shift of maximum export towards the continental margins that is comparable to the results of Dunne et al. (2007). This is likely caused by intensification of particle fluxes in coastal waters and marginal seas, which may be related to regional processes such as more efficient nutrient recycling in shallow regions, or iron leakage from continental shelves (Nishioka et al., 2020) supporting large phytoplankton sizes. In the Atlantic Ocean, the gradual increase of export from northern to southern latitudes (mostly

driven by high export near the coast), and the rapid increase in the Southern Ocean (caused by high export near the Patagonian shelf), are similar to the reconstruction of Henson et al. (2011), although the magnitude is larger. In the Indian Ocean, our reconstruction matches other studies at low latitudes; however, it shows a more dramatic increase in export towards the Southern Ocean sector (discussed in more detail in Section 3.3.2).

3.3.1.2 Seasonal cycle

The seasonal cycle of particle export is comparable to previous studies, when averaged over large-scale coherent biomes (Fig. 3.6). However, significant discrepancies are also revealed. In general, our seasonal cycle is more muted than previous work, suggesting weaker month-to-month variability in some regions, while other regions match previous reconstructions more closely. The most significant discrepancy is observed in the Southern Ocean, in particular in the Antarctic zone, where our reconstruction is substantially higher than previous estimates, with sustained export throughout winter months. we discuss this deviation in Section 3.3.2.

The lower seasonality in our estimate is consistent with the reduced spatial gradients, and suggests overall weaker variations in net community production and export than previously assumed. The machine learning approach used to reconstruct the PSD relies on non-linear relationships with multiple ocean variables to reconstruct particle size distributions, which may accentuate compensatory relationships between different predictors. Surface chlorophyll, temperature, and net primary production have all been used in previous global reconstructions (Dunne et al., 2007; Henson et al., 2011; Siegel et al., 2014), but rarely together with additional variables that may be important in modulating spatial and seasonal patterns of export. It is also possible that our method somewhat underestimates variability compared to previous work. As previously noted, our PSD reconstructions reduces extremes in both biovolume and PSD slope (Clements et al., 2022), which may lead to underestimating variability in particle export fluxes derived from these quantities.

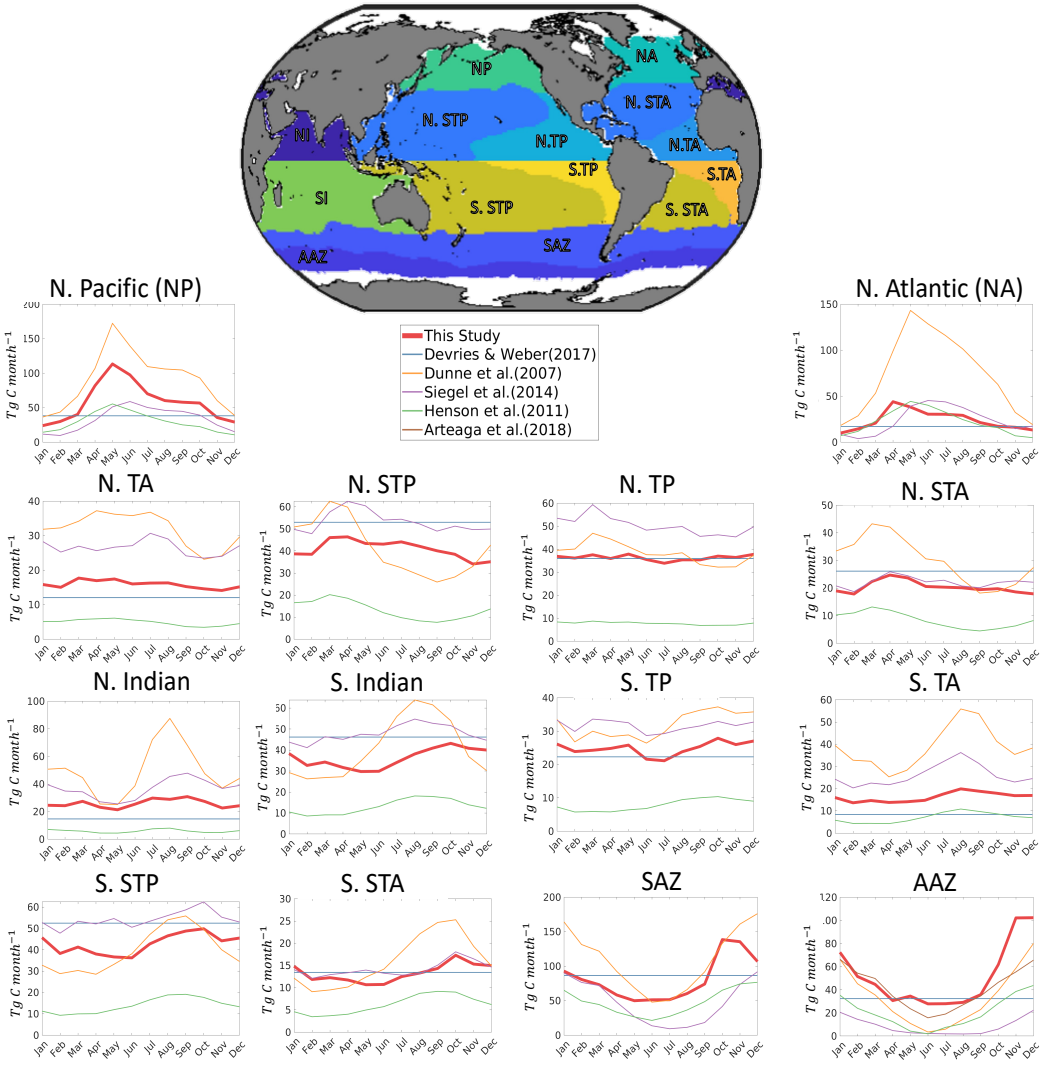

Figure 3.6: Annual seasonal cycle of particle flux from the euphotic zone (Tg C y^{-1}) for the regions specified in the map (top). Each line corresponds to a different estimate, as listed in the legend below the map. The same seasonal spatial mask was applied to each study. Note that the study by DeVries & Weber (2017) provides annual mean export fluxes, which are shown here as horizontal lines.

3.3.2 Southern Ocean Export

Export flux in the Antarctic zone of the Southern Ocean are larger in our estimate than other global reconstructions, especially during austral summer (Fig. 3.6). we also find a southward shift in export, with a peak around 50S, rather than around 40S as in other estimates. A

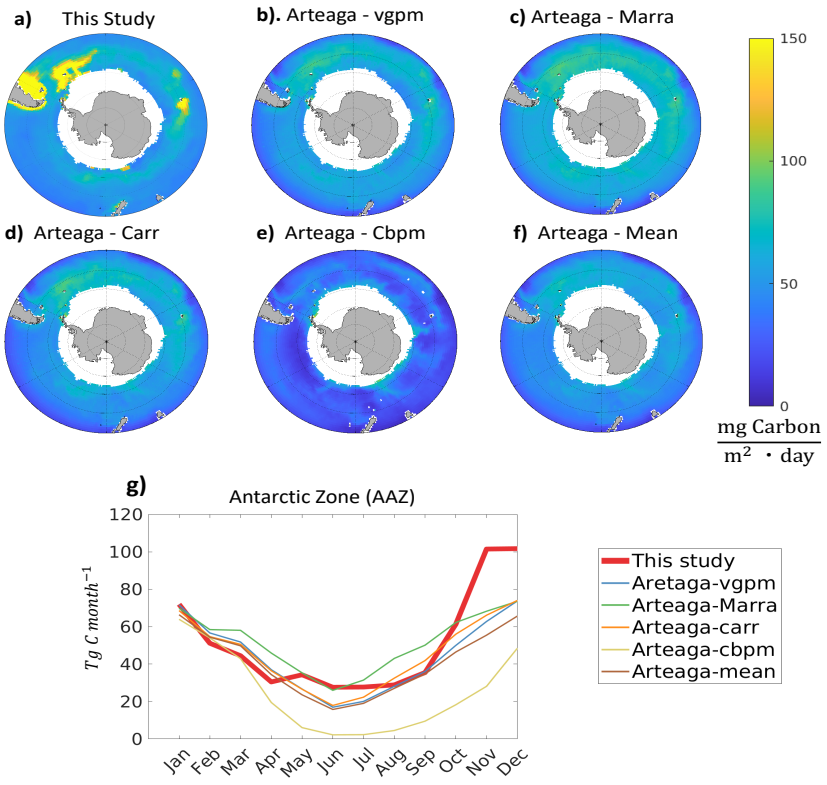

Figure 3.7: Southern Ocean particle export (mg C m^{-2} d^{-1}) for (a) this study, and (b-e) different data-based estimates from Arteaga et al. (2018), and (f) the mean from that study. Each data-based estimate from Arteaga et al. (2018) uses a different net primary production algorithm to derive export. (g) Seasonal cycle of export for each estimate in the Antarctic zone (shown in figure 3.6).

regional study based on 10 years of biogeochemical Argo measurements from 2006-2014, combined with satellite-based net primary production and export algorithms, similarly suggests higher than previously reported particle fluxes throughout the region (Arteaga et al., 2018), in better agreement with our results (Fig. 3.7). This similarity is mostly evident in the open ocean, and varies depending on the primary production algorithm chosen for the comparison. However, our estimate also suggests substantially higher export near landmasses, for example South Georgia and the South Sandwich Islands and the Kerguelen Plateau. Although estimates from Arteaga et al. (2018) do not show the same high flux in austal winter through the end of the year as our reconstruction, they do demonstrate that export fluxes from the

Antarctic zone of the Southern Ocean likely never decrease to the nearly negligible levels shown by other global estimates (Fig. 3.6).

The discrepancy in export from the Antarctic zone relative to prior global estimates could arise from a combination of factors. Observations in the Southern Ocean, particularly in austral winter, are scarce. This is true for both the UVP5 measurements and the climatological predictors used to reconstruct PSD (Clements et al., 2022). The UVP5 data compilation (Kiko et al., 2022) includes two major cruises in the Southern Ocean, which only cover the months of March to May. Satellite-based reconstructions of chlorophyll and primary production from ocean color are on the other hand poorly resolved in austral wintertime. Other climatological variables, such as nutrients and oxygen, are also the results of interpolation of fewer *in situ* observations relative to the rest of the ocean. The scarcity of observations to train the machine learning model used for the PSD reconstructions results in significant uncertainty in predicted PSD and export fluxes in this region (Clements et al., 2022).

our reconstruction reveals significant export primarily next to island masses. Proximity to Southern Ocean islands have been shown to increase productivity and carbon flux (Jouandet et al., 2014; Stemmann, Eloire, et al., 2008), presumably via enhanced vertical mixing and iron fertilization from sedimentary sources in otherwise high-nutrient low-chlorophyll waters (Gaiero et al., 2003). It is possible that other methods of flux reconstructions (Henson et al., 2011; Siegel et al., 2014; DeVries & Weber, 2017) underestimate this increased export, in particular during winter, when observations are scarce. Expanding the number of *in situ* particle flux and UVP5 observations from the Antarctic zone, downstream of major land masses and over the entire seasonal cycle, could help shed light on the patterns of export and their variability in this undersampled region.

3.3.3 Mixed layer versus euphotic zone export

Globally integrated, export from the maximum wintertime mixed layer depth is 6.1 $\pm$0.1 PgC y^{-1}, i.e., about 0.5 PgC y^{-1} larger than the global export from the euphotic zone. This estimate is lower than observational estimates of organic carbon export and annual net

community production from the same depth horizon (Emerson, 2013), which would include additional export mechanisms.

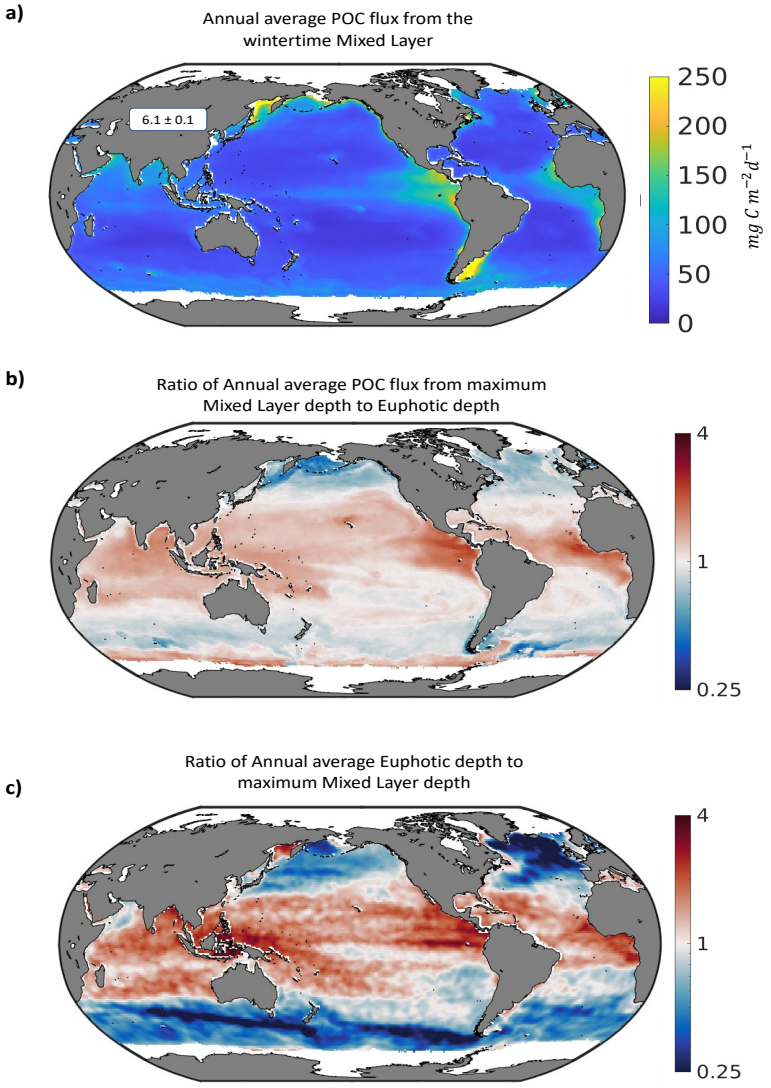

Figure 3.8: (a) Annual mean particle export (mg C m^{-2} d^{-1}) from the maximum mixed layer depth. Total export is 6.1 PgC y^{-1}. (b) Ratio of the export from the MLD to the export from the Euphotic zone. Red indicates a higher export from the MLD (c) Ratio of the annual mean Euphotic zone depth to the Maximum annual mixed layer depth. Red indicates where the euphotic zone is deeper.

Overall, export from the wintertime mixed layer follows broad spatial patterns similar to the export from the euphotic zone (Fig. 3.8a). The tropics and subtropics show larger

mixed layer export fluxes (locally, up to a few times), while high latitudes show overall weaker values (Fig. 3.8b). The low-latitude intensification of mixed layer fluxes is similar in all ocean basins, and more than compensates for the reduction at high latitudes (Supplementary Fig. B.6), thus producing an overall larger export from this horizon. Because of this low-latitude intensification, export from the mixed layer shows stronger gradients between the tropics and high latitudes. Gradients between the equatorial export peak and the subtropical export low are also intensified. Finally, export from the mixed layer in the Southern Ocean is substantially depressed compared to export from the euphotic zone. Similar to export from the euphotic zone, export from the mixed layer tends to be dominated by smaller particles (Supplementary Fig. B.7)

Differences between euphotic zone and mixed layer export can be best interpreted by considering the different depth of these horizons (Palevsky & Doney, 2018) (Fig. 3.8 C). The maximum mixed layer is shallower than the euphotic zone in the tropics and sub-tropics, and is deeper in high latitudes (Fig. 3.8c). This suggests that shallower export horizons are generally characterized by higher fluxes than deeper export horizons, which we attribute to remineralization of particles in surface layers. Specifically, we identify three main latitudinal bands with different horizon depths and export patterns, roughly corresponding to tropics and subtropics, mid-latitudes, and subpolar regions.

Over most of the tropics and the subtropics, the maximum wintertime mixed layer is shallower on average than the climatological euphotic zone (blue colors in Fig. 3.8 c). Here, particle remineralization between the wintertime mixed layer and the euphotic zone depth likely reduces export from the latter horizon, suggesting net heterotrophy in the deeper layers of the euphotic zone, consistent with observations of shallow particle regeneration in the tropics (Pavia et al., 2019).

Over subpolar regions, the wintertime mixed layer is deeper on average than the climatological euphotic zone. Thus, export fluxes reach maximum values within the euphotic zone, and decrease below it following typical flux attenuation profiles (Martin et al., 1987; Guidi et al., 2009). Finally, over most of mid-latitudes, the wintertime mixed layer is deeper on average than the climatological euphotic zone. However, export fluxes from the mixed layer

and euphotic zone are very similar in magnitude, suggesting a close seasonal compensation between enhanced euphotic zone fluxes when this horizon is found above the wintertime mixed layer, and reduced euphotic zone fluxes when it is found below it.

Ultimately, differences in export between the euphotic zone and the wintertime mixed layer are important when considering the role of the biological pump for carbon sequestration (Palevsky & Doney, 2018). Export below the wintertime mixed layer removes carbon from contact with the atmosphere for timescales longer than one year. our results suggest that more carbon is sequestered below the wintertime mixed layer than leaves the euphotic zone. we further suggest a role for particle consumption by heterotrophes (microbes and zooplankton) in reducing the abundance of organic particles in the lower euphotic zone, making it net heterotrophic rather than primarily autotrophic.

3.3.4 Caveats to our approach

our method relies on global PSD reconstructions from UVP5 observations, as well as *in situ* particle flux measurements, both of which are spatially and temporally limited. This in turns reduces the ability of our approach to obtain an accurate climatological picture of PSD and fluxes, and extrapolate local observations to larger regions and other times of the year. In particular, about forty-three percent of monthly particle flux observations contain only one data point, and entire ocean basins are represented by a handful of measurements (Fig. 3.2b). While more widely distributed than flux measurements, UVP5 observations are also characterized by large gaps in space and time (Kiko et al., 2022).

As previously discussed (Clements et al., 2022), regional correlations between environmental properties and PSD observations from UVP5 may not be well captured by extrapolation with a machine learning algorithm trained on data from different regions, especially when non-linear relationships between variables become important. Similarly, our reconstructions rely on the assumption that PSD can be well approximated by power law distributions. Analysis of UVP5 data suggest that this assumption is generally valid over the open ocean (Clements et al., 2022), although locally it may be flawed, and other distributions may be

more accurate (R. A. Reynolds & Stramski, 2021a). Expanding the coverage of *in situ* PSD and fluxes, in particular in under-sampled regions characterized by large variability, such as the Southern and northwest Pacific Oceans, would improve the robustness of our estimates, and shed additional light on regional export patterns not captured by previous work.

The conversion of PSD to export flux encapsulated by Equations 3.1 and 3.5 also suffers from limitations. Converting standing stocks of particles from UVP5 observations to sinking carbon flux using size-dependent relationships assumes that (1) all particles of a given size have the same carbon content, and (2) they all sink at a similar speed proportional to their size. Known biases exist with both assumptions. For example, densely packed fecal pellets often contain more carbon and sink faster than heterogeneous aggregates and marine snow of the same size (A. Alldredge, 1998). Biogenic and lithogenic minerals could alter these relationships on a regional basis, e.g., near continental margins, where lithogenic particles are generally more abundant (R. Reynolds et al., 2010; Trudnowska et al., 2020). Furthermore, we assume globally uniform relationships between particle size, sinking speed, and carbon content. However, these relationships remain highly uncertain (A. Alldredge, 1998; Stemmann, Eloire, et al., 2008; Stemmann & Boss, 2012; B. Cael et al., 2021), and are likely to depend on region and time of the year, reflecting variable particle characteristics and underlying oceanographic and ecological processes.

our approach, which optimizes carbon content and sinking velocity parameters against *in situ* particle fluxes, reduces to some extent the effect of these uncertainties. More work combining *in situ* and optical measurements should focus on constraining these quantities and their regional and temporal variability. Future studies could also improve our approach by distinguishing living and non-living particles, as well as particle type and composition, e.g., by analysis of UVP5 images or other optical methods in conjunction with *in situ* particle samples (Trudnowska et al., 2021).

3.4 Conclusions and future work

we provide a new, data-constrained estimate of particle export fluxes by combining global reconstructions of PSD from UVP5 observations and *in situ* export flux measurements. our estimate of particle export captures regional and seasonal variability in fluxes, and allows reconstruction of export fluxes from spatially variable euphotic zone and mixed layer depths, highlighting the importance of the choice of export horizon (Palevsky & Doney, 2018).

we obtain a global particle export flux of 5.8 $\pm$ 0.1 PgC y^{-1} from the euphotic zone, in line with previous work, although with regional and temporal differences. our results suggest weaker spatial and seasonal variability compared to previous studies, in particular in the open ocean, and highlight the importance of coastal waters and marginal seas for export at high latitudes. Results from the Southern Ocean suggest that processes that sustain elevated fluxes, in particular in wintertime and early austral summer, may not be completely captured by other global reconstructions, and that waters downstream of coasts and islands may harbor a significant source of carbon export to the deep ocean, which is only partially captured in one other reconstruction (Dunne et al., 2007).

we illustrate the ability of our method to obtain particulate organic carbon fluxes at multiple depth by reconstructing and comparing carbon export from the euphotic zone and the wintertime mixed layer depth. Export from the mixed layer is overall stronger than export from the euphotic zone in low and mid latitudes, and weaker in high latitudes, driving a marginally larger flux of 6.1 $\pm$ 0.1 PgC y^{-1}. Differences between euphotic zone and mixed layer export are only partially consistent with results from large scale models (Palevsky & Doney, 2018), and suggest important organic matter remineralization in the deeper parts of the euphotic zone. Three-dimensional reconstructions of particle fluxes would allow a closer investigation of the processes controlling export changes with depth and their implications for particle transfer efficiency and carbon sequestration.

our results highlight the relative importance of particle abundance and size structure in driving export. Total particle biovolume and the PSD slope are correlated in such a way to act synergistically on particle fluxes (Clements et al., 2022). Consistently, higher fluxes are

reconstructed in regions with larger particle biovolume and "flatter" slopes. we also suggest distinct deviations from these patterns, for example in the tropical and northern subtropical Pacific Ocean, where high abundance of all particles, rather than dominance of large relative to small particles, appears to drive elevated export. we further highlight the importance of the PSD by comparing export for small and large particles (here separated at a cutoff size of 418 μm), showing that, while small particles (35 μm - 418 μm range) overall dominate export, large particles (418 μm - 5 mm range) become proportionally more important in high latitudes and tropical regions, where they can account to up to 60 % of export fluxes.

we identify sources of uncertainty and limitations that should be addressed in future work. There remain areas of the ocean and times of the year with limited UVP5 observations and, critically, *in situ* flux measurements, driving uncertainty in both the PSD and flux reconstructions. As UVP5 observations increase in number, our analysis can be refined, for example by expanding comparison with *in situ* sediment trap and thorium-based particle flux observations (Mouw et al., 2016). New machine learning approaches should aim at better capturing fluxes at the high and low end of their range, reducing current biases (Clements et al., 2022). Furthermore, better constraints should be placed on particle carbon content and sinking speed parameters, reflecting regional variability and particle types.

The three-dimensional nature of UVP5 observations paves the way for fully three-dimensional reconstructions of particle export fluxes in the ocean interior. This will greatly benefit from particle flux compilations that span the full depth of the ocean (Mouw et al., 2016), and that harmonize discrepancies between different flux measurement methods (K. M. Bisson et al., 2018). Ongoing deployments of UVP instruments, including on Argo floats, will rapidly increase the number of PSD observations with high vertical resolution (Picheral et al., 2022). In turn, three-dimensional reconstructions of export will enable a better characterization of the ocean's ability to sequester carbon (Boyd et al., 2019), and, in combination with models (DeVries & Weber, 2017; Siegel et al., 2014), a better understanding of the processes behind the ocean's biological pump (Siegel et al., 2022).

CHAPTER 4

The distribution and fate of sinking organic matter in the global ocean interior

4.1 Introduction

Sinking organic particles shuttle organic matter from the productive euphotic ocean to the deep dark interior ocean. This shuttle provides nutrients to the deep ocean ecosystem, while removing and storing carbon from the atmosphere. The effective timescale of this storage is directly related to the depth at which the organic matter is remineralized to carbon dioxide. In the open ocean, the particle transfer efficiency (T_{eff}), i.e., the ratio between the sinking matter flux at a given depth in the ocean interior and the export leaving the surface ocean (generally the euphotic zone, or a shallow reference depth), encapsulates the proportion of organic matter that has escaped most remineralization, and can be stored in the deep ocean over timescales of centuries Boyd et al. (2019). Despite the importance for carbon sequestration, the controls and distribution of the transfer efficiency of sinking organic matter remain poorly constrained and extremely difficult to observe over global scales.

Sinking organic matter in the ocean is difficult to measure, especially below the surface. Historically, estimates of sinking organic matter fluxes were made through the direct capture of organic particles with sediment traps. Early compilations of sediment trap observations revealed a pattern of particle attenuation that is more intense in the upper ocean, and decreases progressively with depth, the so-called Martin curve (Martin et al., 1987). Martin et al. (1987) defined a simple power-law relationship to represent the sinking matter flux at

any depth, as defined by equation 4.1.

$$\phi_z = \phi_0(\frac{z_z}{z_0})^{-b}. \tag{4.1}$$

This pattern of decay has been supported by a variety of particle collection studies across different ocean basins (Martin et al., 1987; Berelson, 2001a). While the Martin curve is a useful, convenient approximation of sinking flux observations, it may not wholly capture the pattern of export at depth.

The exact shape and pattern of this attenuation has been called into question (B. B. Cael & Bisson, 2018). In fact, several studies (Henson et al., 2012; Marsay et al., 2015; Guidi et al., 2015; Weber et al., 2016; Cram et al., 2018) have attempted to resolve the strength of the flux attenuation, finding conflicting patterns, with spatially variable Martin curve coefficients (b in Equation 4.1) that remain poorly constrained. Advances in empirical, mechanistic, and geochemical inverse models of particle fluxes (Henson et al., 2012; Marsay et al., 2015; Weber et al., 2016; Cram et al., 2018) have allowed for a more robust characterization of the particle flux attenuation and the factors that control it. The particle transfer efficiency (T_{eff}) is a non-parametric diagnostic quantity that is often used to describe the POC attenuation. Here, we specifically select as reference depths the base of the mesopelagic ocean at 1000 m, and the the 100m depth horizon, defining the transfer efficiency as:

$$T_{eff} = \frac{\phi_{1000}}{\phi_{100}}. \tag{4.2}$$

However, similar to the Martin curve coefficient, the magnitude and variability of T_{eff} in space and time remain poorly constrained (Cram et al., 2018).

Another source of uncertainty is the role of small relative to large sinking particles for organic matter fluxes in the ocean interior. Previous work suggest that a large proportion of export can be attributed to relatively large, fast-sinking particles such as fecal pellets and phytoplankton aggregates, as opposed to smaller, slow-sinking particles. The importance of large sinking particles in driving the export flux has been well described by observational work (Honjo et al., 2008; A. L. Alldredge & Gotschalk, 1988; J. T. Turner, 2015), although several studies have also highlighted the importance of smaller aggregates, in particular

in the surface ocean (Alonso-González et al., 2010; Durkin et al., 2015; Kiko et al., 2017; Richardson, 2019; Clements et al., 2023).

As particles sink, small aggregates would be preferentially removed via remineralization, suggesting that large particles are primarily responsible for carbon sequestration in the ocean interior (Stemmann et al., 2004a; Burd & Jackson, 2009; Close et al., 2013; Richardson & Jackson, 2007; Devries et al., 2014; Kiko et al., 2017). Estimates of size-fractionated POC flux at depth have indicated an important role for small particle fluxes, at least in specific regions (Durkin et al., 2015), perhaps because of pervasive fragmentation processes (Briggs et al., 2020). However, these observations have not been yet extrapolated to the global scale.

Advances in *in situ* optical observations have allowed the quantification of particle size distribution (PSD), i.e., the number of particles of any given size, in the water column (Stemmann & Boss, 2012; Boss et al., 2015; Lombard et al., 2019). Among optical instruments, the Underwater Vision Profiler 5 (UVP5) is a particle counter that provides the *in situ* particle abundance and size for large particles (80 μm - 2.6 cm) in a given sampled volume (Picheral et al., 2010). The UVP5 consists of a camera that can be attached to a CTD rosette and lowered in the water column to capture images of particles at high frequency. UVP5 can capture PSD observations at a rate of up to 20 images per second, with downcast velocities of around 1 m s^{-1}, as deep as 6 km (Picheral et al., 2010). Since 2008, UVP5s have been routinely deployed on major oceanographic cruises, in all ocean basins (Kiko et al., 2022).

With knowledge of the sinking speed and carbon content of particles as a function of size, PSD observations can be translated into estimates of sinking POC fluxes in the water column (Guidi et al., 2008; Stemmann & Boss, 2012). Machine learning methods have recently been used to extrapolate sparse UVP5 observation of PSD (Kiko et al., 2021) to the global ocean, leading to in global surface POC export estimates (Clements et al., 2022, 2023). The approach relies on empirical relationships that relate particle size and abundance to sinking carbon fluxes (Kriest, 2002; Guidi et al., 2008; Kiko et al., 2017) tuned against a global data set of *in situ* sediment trap and thorium-derived particle flux observations (Mouw et al., 2016).

In this Chapter, we exploit the high vertical resolution of UVP5 measurements to extend the work of (Clements et al., 2022) and estimate global particle export fluxes in the ocean interior. This vertically resolved reconstruction of sinking particle fluxes in turn reveals new patterns of flux attenuation with depth, and thus sheds light on processes controlling the transfer efficiency and the deep ocean carbon sequestration.

The rest of this chapter is organized as follows. Section 2 describes the methods used to estimate vertical profiles of PSD from global UVP5 observations, and their translation to sinking particle fluxes. Section 3 presents the patterns of interior ocean PSD, export fluxes, and transfer efficiency, comparing them to prior estimates, and discussing the implications. Section 4 evaluates the uncertainties inherent to our approach, and remaining research needs. Section 5 summarizes the main findings and future directions for this work.

4.2 Methods

The flux of particulate carbon ($\phi, \frac{mgC}{m^2 day}$) at any given depth can be expressed as a function of three size-dependent quantities: the number (#) of particles of a given size, i.e., the PSD ($n(s), \frac{\#}{m^3 cm}$), the sinking speed ($w(s), \frac{m}{s}$), and the carbon content of paricles in each size class ($c(s), \frac{mg}{\#}$), according to the following equation (Guidi et al., 2008; Stemmann & Boss, 2012):

$$\phi = \int_{s_{min}}^{s_{max}} n(s) \cdot w(s) \cdot c(s)\, ds, \tag{4.3}$$

In the ocean, observed PSD can be well approximated by a power law (Bader, 1970; Sheldon et al., 1972) over a relatively broad size range (from micrometers to centimeters) which encompasses observations of organic particles from the UVP5 (Stemmann & Boss, 2012). Thus, following this power law assumption, we model observed PSD as:

$$n(s) = n_0 \cdot s^{-\beta} \tag{4.4}$$

Here, s (cm) indicates the particle size, and $n(s)$ is the abundance of particles in each

observed size class, between the maximum (s_{max}) and minimum (s_{min}) sizes observed. Following previous work, we assume that the sinking speed and carbon content can also be approximated by size-dependent powerlaws (eq. 4.5 and 4.6).

$$w(s) = w_0 \cdot s^{\eta} \tag{4.5}$$

$$c(s) = c_0 \cdot s^{\zeta}, \tag{4.6}$$

Similar to Guidi et al. (2008) and Clements et al. (2022), we further simplify equation 4.3 to be a function of just the PSD, and sinking carbon parameters m_0 and μ by substituting in equations 4.4 - 4.6 (E.q. 4.7).

$$\phi = \int_{s_{min}}^{s_{max}} n_0 \cdot w_0 \cdot c_0 \cdot s^{-\beta+\eta+\zeta}\, ds = \int_{s_{min}}^{s_{max}} n_0 \cdot m_0 \cdot s^{-\beta+\mu}\, ds \tag{4.7}$$

Using equation 4.7, we have a size dependent estimate for particle fluxes, dependent on four parameter; two for the PSD (n_0 and β) and two for the sinking carbon component(m_0 and μ), thus we can split our methods into two main parts, quantification of the interior ocean PSD, and translation into POC flux.

4.2.1 Calculating PSD parameters

Assuming a power law distribution of particles in the ocean, we estimate the slope β by fitting a linear least-squares regression through the log-transformed particle abundance and the log-transformed equivalent spherical diameter (ESD) of the particles. we then estimate the total volume of particles, here referred to as biovolume, assuming that particle are spherical, by multiplying the individual volume of particle of a given ESD by their abundance $n(s)$, and integrating over the size range sampled by the UVP5:

$$BV = \int_{s_{min}}^{s_{max}} n(s) \cdot \frac{\pi}{6} \cdot s^3\, ds. \tag{4.8}$$

Following the procedure outlined in Clements et al. (2022), we can invert equation 4.8 for the power law intercept n_0 as a function of the PSD slope and the observed biovolume:

$$n_0 = \frac{6 \cdot BV}{\pi} \cdot \left(\frac{s_{max}^{4-\beta}}{4-\beta} - \frac{s_{min}^{4-\beta}}{4-\beta} \right)^{-1}. \tag{4.9}$$

4.2.2 Extrapolation of PSD observations to the global ocean interior

Global climatological reconstructions of PSD require extrapolation of the power law slope and biovolume quantities to the whole ocean on monthly time scales. To do this, we use a bagged Random Forest (RF) algorithm (using the sklearn package in Python) to simultaneously reconstruct climatological PSD slope and biovolume globally, following an approach similar to Clements et al. (2022). A RF deploys an ensemble of decision trees to simultaneously solve a regression equation, and outputs the ensemble average reconstruction. Using a bagged RF, each individual decision tree is independently trained on a randomly selected subset of target data, with a random subset of predictors. The RF is able to learn statistical relationships between target variables (here, UVP5-derived slope and biovolume) and a series of predictors (here, environmental variables), to make reconstructions that minimize the error between predicted and observed data. Since machine learning algorithms are particularly sensitive to overfitting the data, we evaluate the method by quantifying the error between our RF model and the data that was not used for training, i.e., the so-called "out-of-bag" data.

our RF model relies on a set of predictors and target data gridded to match our desired output resolution. In our case, we use climatological monthly predictors at 1-degree spatial resolution, with 102 depth levels matching the grids used in the World Ocean Atlas (Locarnini et al., 2019). To construct our predictor set, we include variables that have been globally quantified and mechanistically linked to processes of particle production and remineralization. These range from physical variables (e.g., temperature and salinity) to ecosystem-level quantities (e.g., primary production, oxygen concentration and utilization). A list of all predictors is shown in Table 4.1.

Table 4.1: Variables used to predict PSD parameters, variations (i.e., vertical or temporal changes) and data sources.

Variable	**Variations**	**Source**
Temperature	Time Derivative Depth Derivative	Locarnini et al. (2019)
Salinity	Time Derivative Depth Derivative	Zweng et al. (2019)
Oxygen	Time Derivative Depth Derivative	H. E. Garcia et al. (2019)
AOU	Time Derivative Depth Derivative	H. E. Garcia et al. (2019)
Nitrate	Time Derivative Depth Derivative	H. Garcia et al. (2018)
Phosphate	Time Derivative Depth Derivative	H. Garcia et al. (2018)
Silicate	Time Derivative Depth Derivative	H. Garcia et al. (2018)
Chlorophyll	Time Derivative	NASA G.S.F.C (2014)
Mixed Layer	Time Derivative	Johnson et al. (2012) de Boyer Montégut et al. (2004)
Eppley VGPM VGPM CBPM CAFE	Time Derivative	Antoine & Morel (1996) Behrenfeld & Falkowski (1997) Westberry et al. (2008) Silsbe et al. (2016)
Iron	Time Derivative	Hamilton et al. (2019) Myriokefalitakis et al. (2018)
Shortwave Radiation	Time Derivative	Copernicus C.C.S (2017)

The predictor set is used to reconstruct PSD slope and biovolume on the three-dimensional ocean interior grid of the World Ocean Atlas (Locarnini et al., 2019). The reconstruction is based on the ensemble average of 250 RF realizations with variable hyper-parameters (the number of decision trees and their complexity), with the inter-model spread serving as an error estimate on the reconstructed PSD. Clements et al. (2022) demonstrated that using an ensemble of randomized RFs reduces biases associated with specific predictors and potential overfitting, making the results robust with respect to parameter tuning and the choice of different observational products (i.e., NPP algorithms). Thus, the PSD reconstructions are not the result of fine-tuning the model hyper-parameters, or choosing only the best predictors, but likely encapsulate real — albeit complex and highly non linear — relationships between water column conditions and PSD. To evaluate the robustness of our reconstructions, we report goodness-of-fit statistics including the coefficient of determination(r^2), the root mean square error (RMSE), and the average bias, calculated by comparing reconstructions to co-located *in situ* observations.

4.2.3 Calculating the sinking particle flux

Previous work shows that particle sinking speed and carbon content depend strongly on size, and can be approximated by power laws (Kriest, 2002), either separately or in combination. Combining estimates of PSD with observations of POC flux allows to estimate the parameters that control the size-dependent sinking carbon content of particles, i.e., m_0 and μ in Equation 4.7 (Guidi et al., 2008; Clements et al., 2023).

While previous work shows a general robustness of the approach to reconstructing these quantities near the surface, their changes with depth have not been previously explored. However, because of complex particle transformation processes in the ocean interior, we expect significant vertical variations in both the sinking speed and carbon content of particles of a given size (Berelson, 2001b). For example, the mineral fraction of particles is likely to increase with depth, because of slow dissolution rates relative to carbon remineralization. Thus, we employ a depth dependent optimization scheme to estimate the combined carbon

content and sinking speed parameters that allow translation of PSD into POC flux.

As a first approximation, we generate a step-wise linear profile of m_0 and μ that can be optimized against *in situ* observations of POC flux (Mouw et al., 2016). These profiles are constructed by defining a set of reference depths for which both m_0 and μ need to be specified, and letting them vary linearly in between. Given these profiles of sinking carbon parameters, particle flux at any depth can be calculated with knowledge of the PSD, by applying Equation 4.7.

we optimize for the values of m_0 and μ at the reference depths by minimizing the sum of the square errors between the log of *in situ* carbon flux measurements (Mouw et al., 2016) and the estimates from Equation 4.3, co-located at the same depth and month of observations, following the same approach of Clements et al. (2022). Here, for simplicity, we use three reference depths, respectively at 100 m, 458 m, and 2000 m, using a log-spacing between the top and bottom levels.

To first demonstrate the robustness of this method, we test it on a synthetic, three-dimensional flux data set generated by combining the global interior PSD reconstructions with prescribed step-wise carbon content and sinking parameter profiles with known parameters. To further validate our methodology we apply log-normal random errors to the synthetic data and re-optimize against a random, variable , more realistic, flux. Details of this validation are further described in Section 4.3.2.

4.3 Results and Discussion

4.3.1 Machine Learning reconstructions of interior ocean PSD

Fig 4.1 show the global reconstructions of PSD biovolume and slope. our reconstruction approach is able to capture most of the variability of the UVP5 observations, and robustly reproduces the gridded measurements of biovolume (r^2=0.99), and PSD slope (r^2=0.97) (Fig. C.1). Comparison with observations that are not used for training (out-of-bag) provide a more stringent test for the method's robustness. As shown in Fig. 4.1B and D, these out-

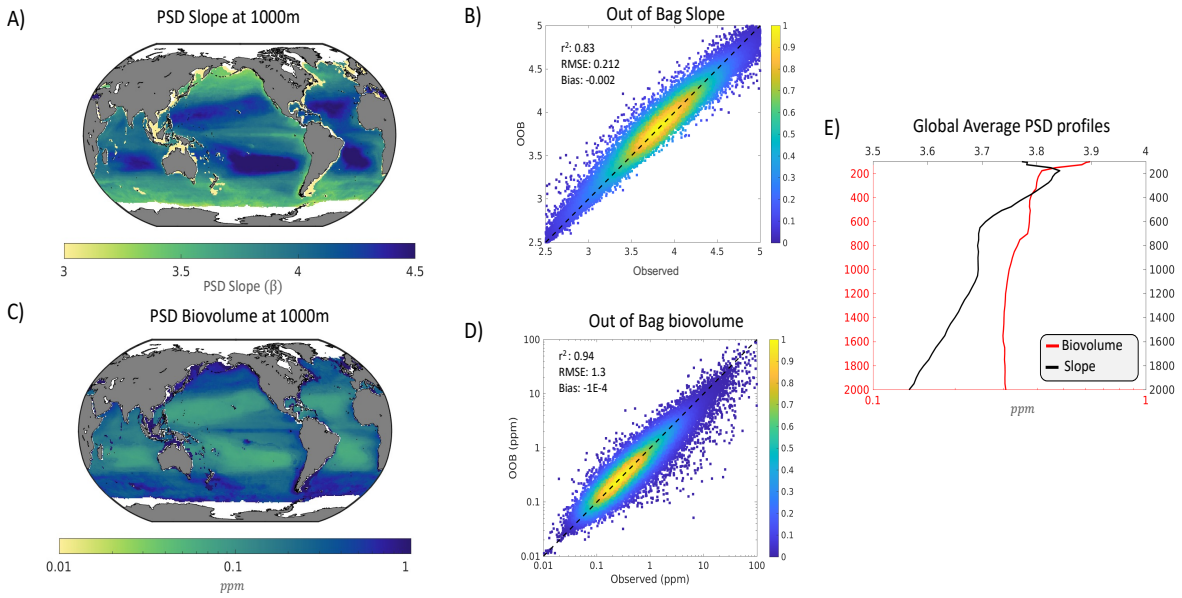

Figure 4.1: Reconstructed PSD biovolume (in parts per million, ppm) and slope at 1000m. (a) Annual mean PSD slope reconstructions at 1000m. (C) Annual mean PSD biovolume reconstructions at 1000m. (B,D) Performance of the RF reconstruction for PSD slope and biovolume, respectively, shown as density scatter plots of out-of bag prediction vs. observed quantities for all depths (colors indicate the normalized density of observations at each point). E) shows the global averaged depth profile of both PSD slope (black) and total biovolume (red). Annotations in (B,D) show the coefficient of determination (r^2), the rmse, and the global bias.

of-bag observations are also robustly predicted, with a RMSE of 1.3 ppm for biovolume (r^2=0.94) and 0.33 for slope (r^2=0.83). Relative to both the full data set and the out-of-bag observations, our reconstructions show negligible biases.

In comparison to results of surface export (Clements et al., 2022), our depth resolved reconstruction of PSD produce statistically stronger results. This can be attributed to the greatly increased data availability when considering observations in the water column rather than at a single surface depth horizon. With more data available, the ability of machine learning models to capture a larger proportion of the variability in observations increases. Additionally, the predictor data set is now more tightly coupled to the UVP5

observations in the vertical, because predictors are used at exactly the same depth of UVP5 observations, rather than by averaging them over a broader depth range as in (Clements et al., 2022). Operationally, we also use one single model to reconstruct both the PSD slope and biovolume simultaneously, as opposed to two independent models as in Clements et al. (2022). These factors likely work to reduce the uncertainty relative to previous work. However some discrepancies with the data, and caveats to this approach remain, as discussed in Section 4.4.

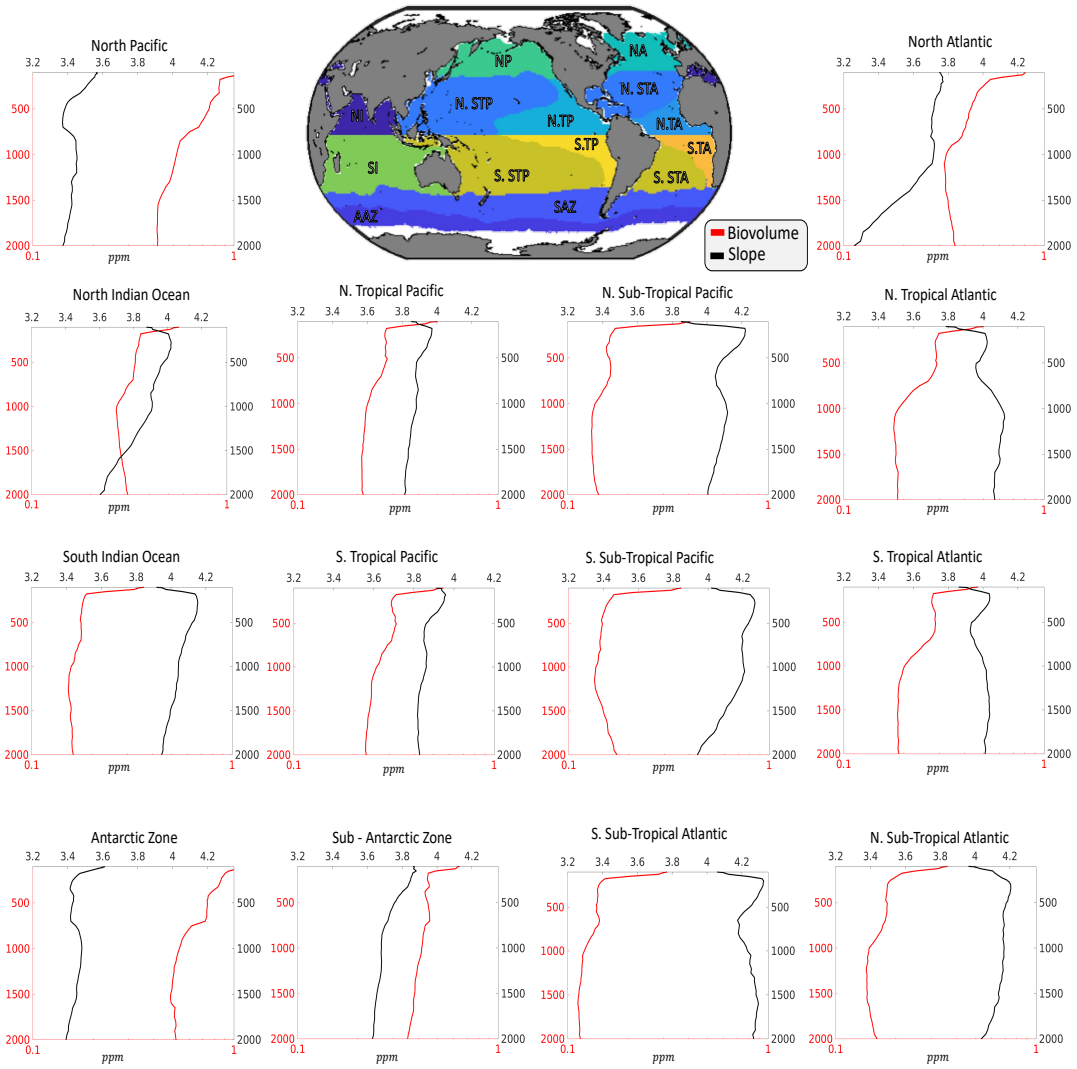

Figure 4.2: Depth varying PSD biovolume and slope for bio-regions as defined by Weber et al. (2016) and adapted by Clements et al. (2022). Red lines show the biovolume in PPM, black lines the slope.

The three-dimensional reconstructions of PSD reveal remarkable patterns with depth and region, with implications for export fluxes. we highlight a reduction in the global average biovolume from the surface (0.6 ppm (Clements et al., 2023)) to the base of the mesopelagic zone (0.27 ppm). Similarly, the PSD slope decreases from an average surface value of 3.9 (Clements et al., 2023) to 3.7 at 1000 m. The average PSD profile (Figure 4.1 E) shows an initial rapid attenuation of biovolume, followed by a slight increase at depths great than 1500 m. Similarly, the slope often initially increases with depth near the surface, before decreasing.

Across regions, the patterns and magnitudes of these changes also varies. Areas of high productivity, such as near the coast, in high latitudes, and along the equator, show higher biovolumes and lower slopes in the upper ocean. Meanwhile, low productivity regions, such as the subtropics, show opposite patterns, with low upper ocean biovolume and high slopes. In previous work (Clements et al., 2022), we described the coupling of these PSD properties near the euphotic zone depth, where slope and biovolume tend to be anti-correlated, and suggested that high productivity regions and periods tend to result in large numbers of particles (and hence high biovolume) and dominance of larger aggregates (and hence lower slopes). The progressive decrease of PSD slopes with depth suggests an increased importance of large particles, supporting the idea that, despite the presence of fragmentation processes (Briggs et al., 2020), small particles may be preferentially remineralized, leaving a surplus of large particles.

In several regions, in particular in the tropics and subtropics, a second-order pattern of variation with depth emerges, with a local minimum in slope between 400 m and 800 m approximately, often accompanied by a slightly shallower local maximum in biovolume. Previous work has suggested that the changes in particle PSD with depth, especially an increase in biovolume and in the importance of large particles with depth, could be caused by in situ particle production by vertically migrating animals (Bianchi et al., 2013), or by particle (dis-)aggregation and remineralization (Bianchi et al., 2018; Briggs et al., 2020). we suggest that this pattern is likely caused by the injection of relatively large new particles by diel vertical migrations of zooplankton and micronekton (Bianchi et al., 2013; Cram et al.,

2022).

4.3.2 Optimization of sinking carbon parameter: method validation

Given the poor knowledge of the sinking speed and carbon content parameters m_0 and μ, and uncertainties associated with optimization methods, we first demonstrate the ability of our method by testing it on a synthetic particle flux data set obtained by applying known profiles of these quantities. Figure 4.3 demonstrates the ability of our depth-dependent optimization scheme to recover the vertical variability of the synthetic flux profile (Figure 4.3 a-c,g). Although some uncertainty remains in the reconstruction of the known slope and intercept profiles (figure 4.3 a,b), the method robustly captures the resulting flux profile ($r^2 = 0.99$). The method also remains accurate after application of a random error to the synthetic flux data set (Figure 4.3 d-f,h), although with a higher degree of uncertainty (the coefficient of determination of the reconstruction dropping from 0.99 to 0.6).

While the method shows that the synthetic flux can be robustly recovered, capturing the exact sinking carbon parameters (intercept and slope) is more difficult. This is reflected in the fact that, as identified in Clements et al. (2023), there is a tight coupling of the two parameters, so that they often compensate to produce a nearly identical pattern of flux. Thus, interpretation of the optimized parameters should be done with caution, and the results should be interpreted as one of the many possible solutions that can produce profiles of flux consistent with observations.

4.3.3 Optimization of sinking carbon parameter: Results

To calculate the global particle export, we apply the step-wise linear profile for carbon content and sinking speed parameters, and optimize it against the global compilation of sediment trap and thorium-based sinking fluxes described by Mouw et al. (2016). we assume that the flux measurements in the compilation have been quality controlled, and do not apply any further checks. Given the observed compensation between optimized sinking carbon slope and intercept in the synthetic data, we perform this optimization 10 times,

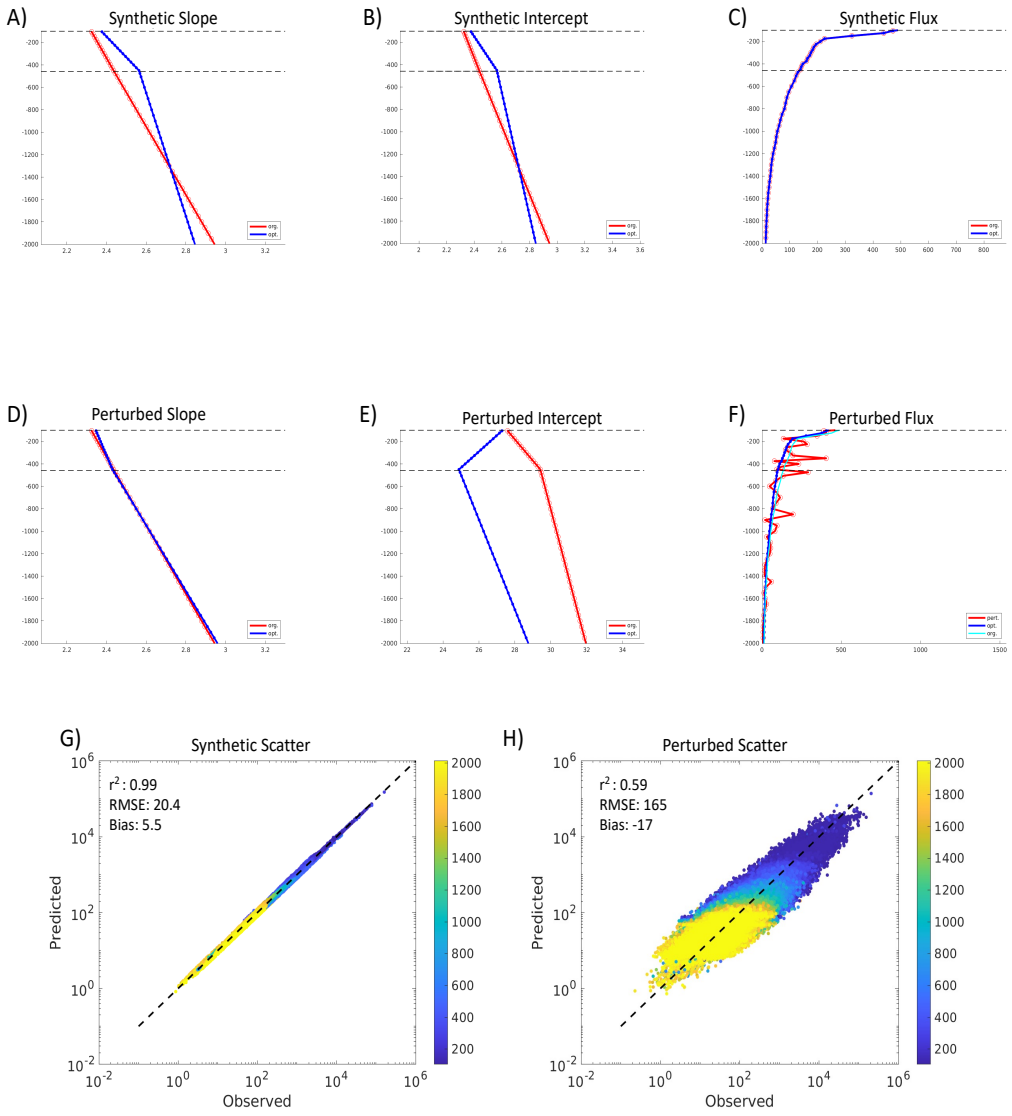

Figure 4.3: Synthetic data reconstructions of particle carbon content and sinking speed parameters and resulting fluxes. (A) Prescribed and optimized sinking carbon slope and (B) intercept, and (C) the resulting flux using a depth-dependent optimization. D-F) Show the same quantities, but with a log-normal random error added to the data. G-H) scatter plots over the whole ocean for (G) the unperturbed and (H) perturbed data. Colors show the depth of the observations, and annotations the summary statistics. A-F show the results for one specific profile, G-H) show the scatter of all data points.

with each optimization starting with randomly assigned initial values. The spread of the resulting fluxes provides a measure of the error associated to the optimization procedure.

The average optimized particle export fluxes are shown in Figure 4.5; the average slope and intercept profiles are shown in Figure 4.4.

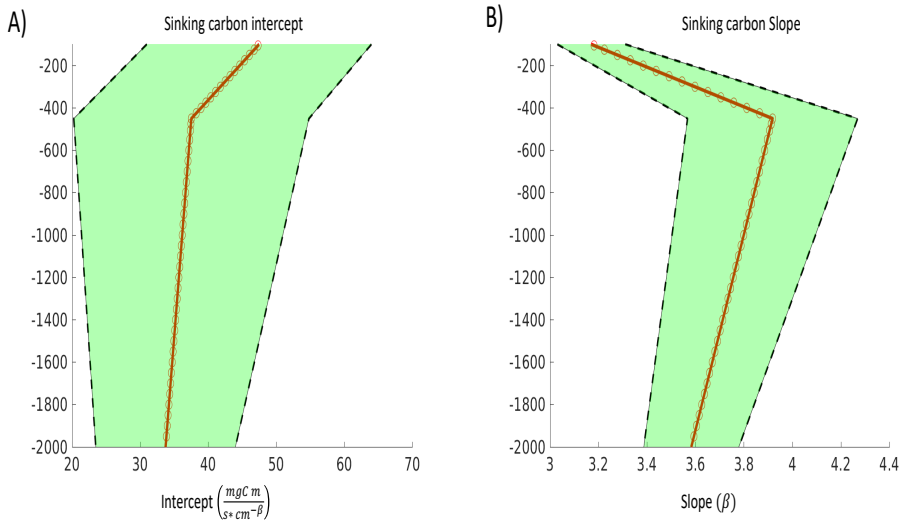

Figure 4.4: (A) Sensitivity of the sinking carbon intercept parameter, red line indicates the mean value used by this study. (B) Sensitivity of the sinking carbon slope parameter, red line indicates the mean value. Green shading on both, indicates one standard deviation error estimates.

Similar to the randomly perturbed synthetic data, we find a coefficient of determination of r^2 0.5 for the average flux reconstructions, optimized against the observations from the Mouw et al. (2016) database, indicating that our approach can capture a significant fraction of the variability of observed sinking fluxes. The resulting particle flux profiles match relatively well the observational compilation, showing a rapid attenuation near the surface, a stabilization at around 500 m, and a slight increase with depth below 1000 m (Figure 4.5D). While our annual average export lacks small scale variability seen in observation records, this is likely due to infrequent sampling by traps, especially in the deep ocean. Globally, we find a dramatic decrease of export from 6.0 PgC/yr leaving the surface (100 m) to 0.6 PgC/yr at a depth of 1000 m.

Seasonal and regional differences in the POC flux are also apparent. Figure 4.6 shows the regional average and seasonality of each ocean region, In general, the highest exports are

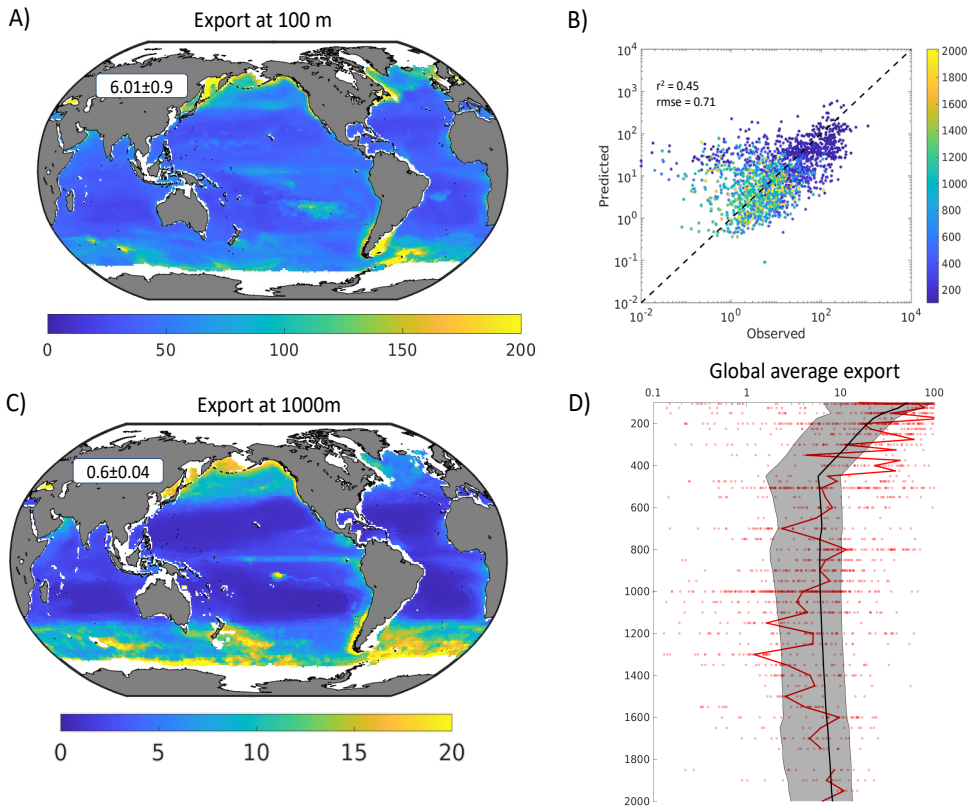

Figure 4.5: A,C) show the patterns and total annual particulate carbon export from the surface (100 m) and the deep mesopelagic zone (1000 m) respectively. B) Performance of our optimization results (y-axis) vs. observed particle export fluxes (x-axis) (Mouw et al., 2016). Annotations on the figure show the statistical summary of the fit. D) Shows the global average flux profile and seasonal flux range (Black line and shading) for this study, compared to the gridded observations from Mouw et al. (2016) (red dots) and the global average profile (red line).

observed in the high latitudes, and the lowest in the subtropics. we find the regions with the highest POC fluxes also have the highest seasonality, while the tropics and subtropics have the lowest. Seasonality generally peaks at the surface and decreases with depth. Similar to the pattern of particle biovolume, we find at many locations a slight increase in POC flux in the mid-mesopelagic ocean, which is likely caused by particle injection by vertically migrating animals.Remarkably, we also reconstruct slight increases of particle fluxes with

depth below 1000 m in most regions – a feature reminiscent of observations.

4.3.4 Global patterns of export and flux attenuation

our reconstructions produce a pattern of flux attenuation with depth comparable to observations (Fig. 4.5) (Martin et al., 1987; B. B. Cael & Bisson, 2018). In general, we find that the T_{eff} reflects patterns similar to the surface PSD and export (Clements et al., 2022, 2023). Similar to previous studies (Weber et al., 2016; Cram et al., 2018), the highest transfer efficiencies are observed in the high latitudes and upwelling regions – including the equator, while the lowest are observed in the subtropics. It is also apparent that the anoxic cores of the oceanic oxygen minimum zones, in particular in the eastern tropical North Pacific are characterized by a higher transfer efficiency than the surrounding oxygenated water, suggesting a reduction in flux attenuation within anoxic layers(Roullier et al., 2014; Weber & Bianchi, 2020; Cram et al., 2022).

our T_{eff} shows a finer degree of spatial variability than other studies. Specifically, our study highlights the importance of large-scale frontal structures, where we find both enhanced export at the surface, and enhanced T_{eff}. For example, we find high transfer efficiency in the mid-latitudes at the boundaries between subtropical and subpolar gyres, as well as along the equator. we suggest that along major frontal regions vigorous eddy mixing not only enhances the production of large particles at the surface, but also their subduction and vertical flux, and thus the transfer efficiency (Omand, D'Asaro, et al., 2015; Dall'Olmo et al., 2016; Stukel et al., 2018; Bach et al., 2020). With more dense sampling at fine spatial scales, it should be possible to quantify the role of sub-mesoscale fronts on enhancing export (Estapa et al., 2015).

4.3.5 Role of particle size for the sinking flux

While the transfer efficiency clearly show a decrease in the total POC flux with depth, our approach also allows us to quantify the importance of particle size for these fluxes. we follow the same definitions of small particle fraction (SPF, $35\mu m - 414\mu m$) and large

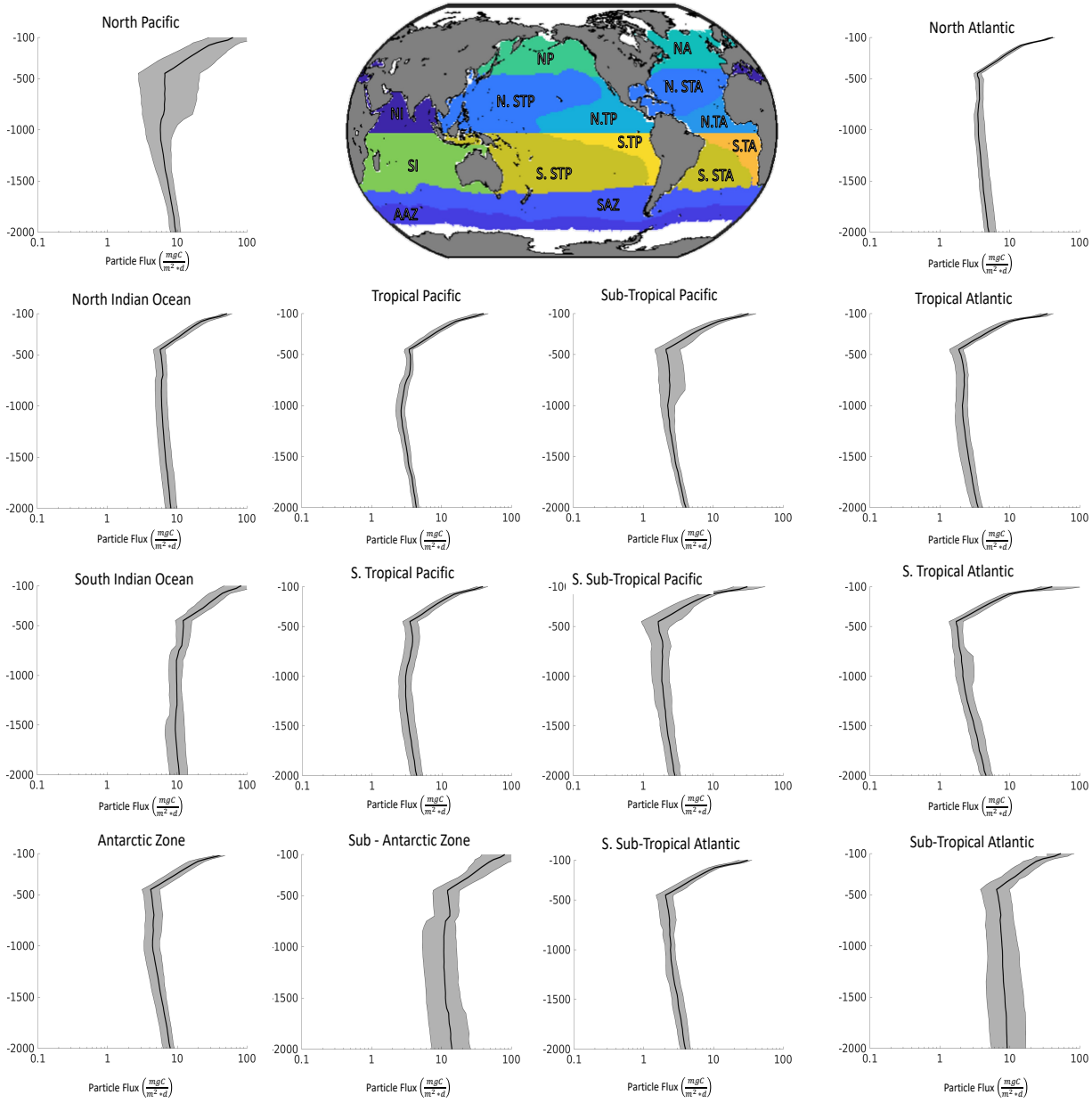

Figure 4.6: Particulate carbon flux profiles for each global region as defined by Weber et al. (2016). Black line indicates the annual mean export profile, while shading reflects the seasonal variability for reach region and depth.

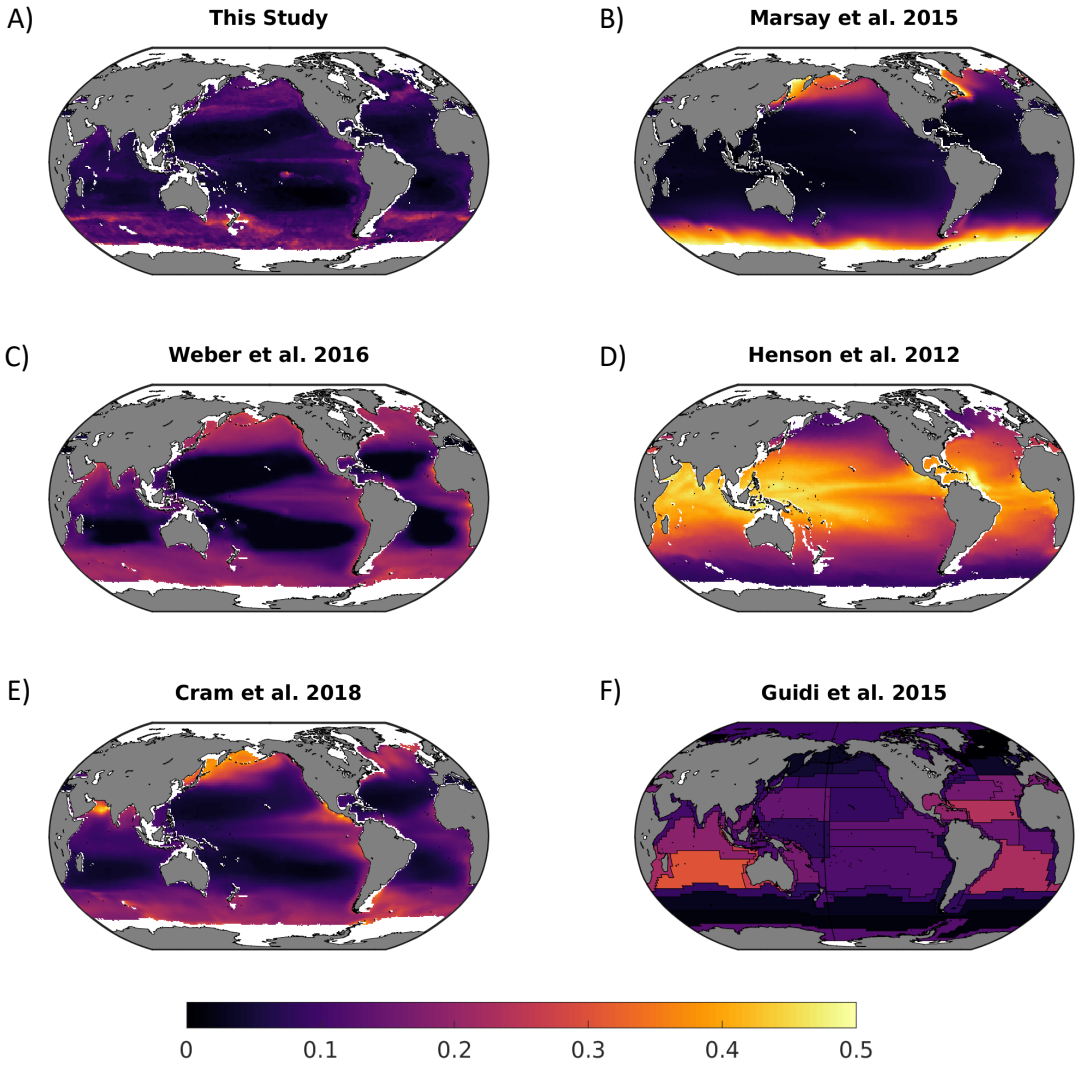

Figure 4.7: Particle flux transfer efficiency for A) This Study, B) Marsay et al. (2015), C) Weber et al. (2016), D) Henson et al. (2012), E) Cram et al. (2018), and F) Guidi et al. (2015). For this study, T_{eff} is defined as the ratio of export at 1000m to export at 100m, for all others it is the ratio between the export at 1000m to export at the base of the Euphotic zone. See Appendix C for a discussion on the calculation of T_{eff}

particle fraction (LPF, $414\mu m - 5mm$) as (Clements et al., 2023), and we quantify their relative contribution to the the sinking particle flux as a function of depth.

Figure 4.8 shows the relative contribution of the two size fractions near the base of the euphotic zone (100 m), near the base of the mesopelagic zone (1000 m) and entering the

abyssal ocean (2000 m). Near the surface, we confirm previous results (Clements et al., 2023), showing similar contributions of small and large particles, with small particles overall more important, especially in low-productivity waters. However, deeper in the water column, we find an increasingly stronger contribution from large particles.

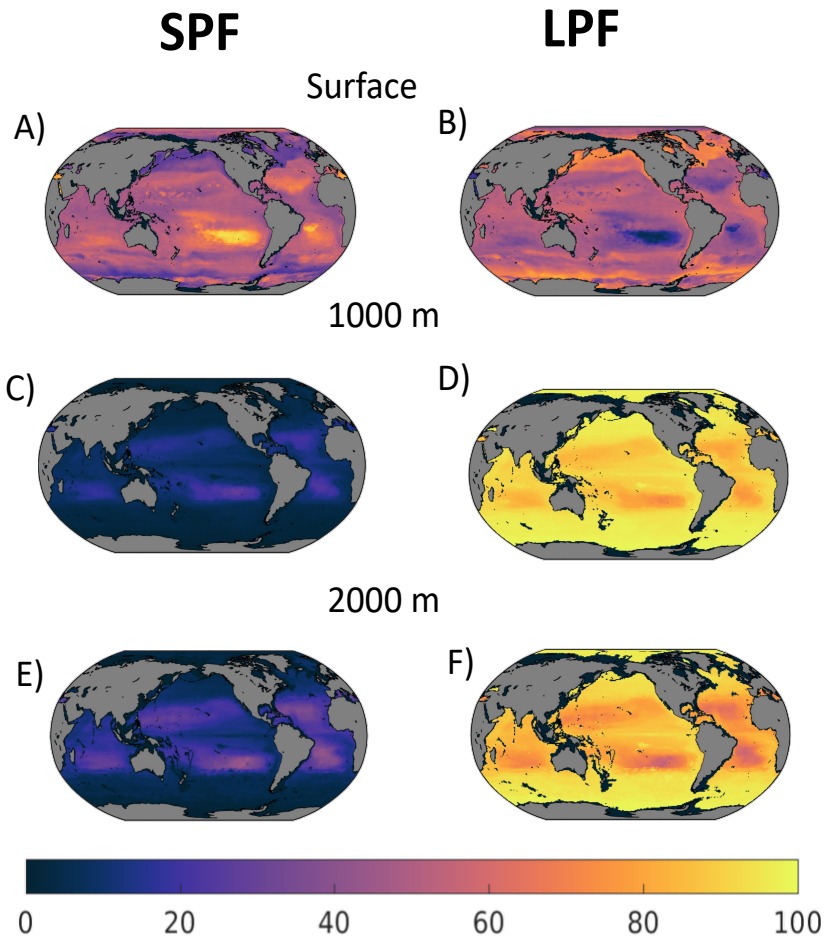

Figure 4.8: The percentage of the total carbon export for the small particle fraction (A,C,E) and the large particle fraction (B,D,F), at the surface (A,B), 1000 m (C,D), and 2000 m (E,F).

This pattern is the result of the combined effect of the decrease of the PSD slope with depth – i.e., the increase in the abundance of large vs. small particles, and the increase in the carbon content and sinking speed of a given particle (C_w), as encapsulated by the increase with depth of the slope of the sinking carbon power law relationship (4.7).

While we cannot directly quantify the mechanisms behind the shift in PSD, carbon

content and/or sinking speed of particles with depth, they likely reflect a combination of factors. For example, the decrease of the PSD slope with depth may reflect preferential remineralization of slowly sinking small particles Devries et al. (2014), aggregation of small particles by repackaging or differential settling, or production of large particles *in situ* by vertically migrating organisms Pinti et al. (2022). The increase with depth of the sinking speed and carbon content of individual particle may reflect a progressive enrichment by mineral ballast, or changes in particle properties and type as they reach the deep ocean.

4.4 Caveats to our approach

Previous work has described the limitations associated with translating UVP5 observations to global estimates of PSD and POC flux (Clements et al., 2022, 2023). Similar caveats remain for the present analysis, and relate mainly to three distinct error sources: the parametric description of PSD as power laws, continued limitations associated with the machine learning extrapolation approach, and difficulties in translating PSD to POC flux by using simple, globally uniform profiles of particle carbon content and sinking speed.

our method is based on global PSD reconstructions from UVP5 observations, which suffer from two main sources of error: first, the power law assumption used to approximate PSD globally, and second, the machine learning reconstruction. While power laws are a common and widely adopted first-order description of PSD in the ocean (Bader, 1970; Sheldon et al., 1972; Stemmann & Boss, 2012; Clements et al., 2022), significant deviations from it have been reported (Organelli et al., 2020; R. A. Reynolds & Stramski, 2021b). However, our tests suggest that a power law assumption is globally robust for the range of sizes sampled by UVP5 instruments (Clements et al., 2022). we also assume that all particles are organic and homogeneous in nature, while lithogenic particles may be present, especially in coastal regions or near the sediment-water column interface. Additionally, the random forest approach is able to robustly reconstruct observations, but provides limited mechanistic understanding on the relationship between environmental variables and PSD properties.

The conversion of PSD to export flux (Equation 4.3) also suffers from inherent limitations

(Clements et al., 2023). The approach assumes that particle carbon content and sinking speed are proportional to size, and that all particles of the same size have the same properties, while in reality a high degree of heterogeneity is observed (A. Alldredge, 1998; Stemmann, Eloire, et al., 2008; Stemmann & Boss, 2012; B. Cael et al., 2021). Although we account for simple changes in these particle properties depth, we assume that they are spatially uniform. Since particle composition is expected to reflect the community composition and oceanographic processes, regional and seasonal variability in particle properties should be expected. While we are able to capture the major patterns of particle fluxes globally, the shape of the profile chosen here should be more closely examined. Errors in observed particle fluxes, which are notoriously difficult to measure, are also likely to affect our estimates.

Further work should be considered to improve our results. First, expanding the coverage of observations with UVP5 and similar instruments, in particular in under-sampled and deep ocean regions such as coastal, high latitudes (in particular in the Southern Ocean) and to depths deeper than 1000 m, would improve the robustness of our estimates, and shed additional light on regional PSD patterns not captured here. Future studies could also improve our approach by distinguishing living and non-living particles, as well as particle type and composition. Analysis of UVP5 images or other optical methods in conjunction with *in situ* particle samples (Trudnowska et al., 2021), would allow a better characterization of the carbon content and sinking speed of particles. Finally, while we resolve first-order variation of these quantities with depth, more work should be done to quantify their regional and temporal variability.

4.5 Conclusions

In this Chapter, we provide a new, data-constrained estimate of particle size spectra and export fluxes in the ocean interior, based on global UVP5 observations from more than a decade of observations. our results show a rapid decrease of particle biovolume with depth just below the 100 m depth horizon, accompanied by a rapid increase in PSD slope. Further down in the water column, between a depth of few 100 m to 2000 m, both biovolume and

slope tend to decrease with depth, although changes are much less sharp than near the surface.

We also highlight fine scale variations in PSD with both depth and by region. Among the most remarkable patterns in our reconstruction, we reveal the consistent presence of a local maximum in biovolume accompanied by a slightly deeper local minimum in slope, found between depths of 400 and 800 m. We attribute this pattern to the injection of large particles by diel vertical migration of zooplankton and micronekton. Deeper in the water column, we observe a consistent increase in biovolume between depths of 1500 and 2000 m, which could reflect a combination of production of new particles at depth, increase in particle porosity (i.e., the water content), anf re-suspension from the sediment and horizontal transport in deep nepheloid layers. Future work should be focused on regional analysis of this deep increase to better parse its causes and implications for deep particle sinking fluxes.

Globally, we obtain a particle export flux of 6.0 PgC y^{-1} from the the surface ocean (100 m), which attenuates to 0.6 PgC y^{-1} by 1000 m depth, corresponding to a global transfer efficiency of 10%. We find that, as depth increases, large particle become relatively more important for sinking carbon fluxes than smaller particles, so that by 1000 m, most export can be attributed to the large particle fraction.

Regionally, we find significant variations in the transfer efficiency, with areas of high export characterized by high transfer efficiency. Our results agree with recent studies (Weber et al., 2016; Cram et al., 2018), although we find generally smaller values of transfer efficiency than previously proposed, with finer regional variations. Taken at face value, our results suggest an overall weaker carbon sequestration by sinking particles than previously assumed, with preferential, localized deep-sequestration pathways found at the boundaries between subtropical and subpolar regions, at the equator, and in oxygen deficient zones.

Fine scale patterns identified in the PSD reconstructions can also be recognized in the resulting particle fluxes. For example, we find a decreased flux attenuation, or even a flux increase, in the upper mesopelagic ocean (400-800 m), coincident with a local maximum in biovolume and a minimum in slope. This suggest that particle injection by diel vertical

migrations leaves a detectable imprint in particle fluxes at depth, likely contributing in a significant way to deep carbon storage (Pinti et al., 2022). Surprisingly, we also find a small increase of particle fluxes with depth below the base of the mesopelagic ocean, down to a depth of 2000 m. While this pattern is globally consistent with the sediment trap and Thorium-based flux compilation (Mouw et al., 2016), it implies a net production of organic material in abyssal layers, for which we can not provide a convincing mechanistic explanation. Potential causes, including deep chemosynthesis, animal migrations, or re-suspension from sediment and lateral transport appear unlikely. It is possible that, despite the introduction of depth-dependent variations, our optimization approach overestimates the sinking speed or carbon content of deep sinking particles, contributing to the abyssal flux increase with depth. Future efforts should focus on providing better observational constraints on the relationship between particle size, carbon content, and sinking speed in the deep ocean.

The large scale patterns identified in our study open the venue for new lines of research, but also highlight remaining limitations. Data coverage is still scarce, which is compounded by the high variability in observed PSD. Ongoing deployments of smaller, energy efficient UVP6 instruments, including on Argo floats, are likely to rapidly increase the number of PSD observations with high vertical resolution (Picheral et al., 2022), allowing to better constrain global patterns observed here. With a growing network of optical observations, remaining spatial and temporal gaps in PSD observations will likely be closed.

Mechanistic exploration of the processes controlling observed depth-dependent patterns is an important undertaking. Our analysis provides valuable constraints for this type of work. But perhaps, the most pressing need is the ability to characterize variations in particle carbon content and sinking speed with depth and region. Future work that combines optical observations with *in situ* sampling techniques, such as particle pumps, is likely to provide increasingly accurate *in situ* description of sinking organic matter for flux calculations, a critical undertaking in the context of continuing ocean variability and change.

CHAPTER 5

Summary and Conclusions

In this Chapter, I conclude the thesis with a summary of the results from Chapters 2–4. I also review the progress on the primary science questions presented in Section 1.3. Finally, I conclude with a discussion on future work that is inspired by the results presented in this thesis.

5.1 Summary

In this dissertation, I have demonstrated the ability of statistical machine learning models to improve estimates of sinking organic matter in the ocean. The three studies presented here have addressed specific, critically important facets of the ocean's biological carbon pump, and can be summarized as follows:

- Chapter 2 explored the particle size distribution of sinking organic particles at the base of the euphotic zone over the global Ocean. These new results capture the regional and seasonal variability of marine particle abundance and size distribution, and highlight the coupling (and decoupling) between particle biovolume and surface ocean chlorophyll, which we identify as a key driver of total organic matter. In contrast, I show that chlorophyll alone is not as strongly linked to particle size distribution, for which other variables of community structure, e.g. subsurface oxygen utilization, are also important. These results shed light on multiple ocean processes behind the ocean's biological pump, and lay the foundation for a better quantification of sinking particle fluxes (Guidi et al., 2008), and the rates of aggregation, disaggregation, and degradation of organic particles (Burd & Jackson, 2009; Briggs et al., 2020). My char-

acterization of particle size distributions is also critical to quantitative descriptions of the interaction between particles and marine microorganisms (DeLong et al., 1993; Church et al., 2021), migrating animals (Bianchi et al., 2013; Cram et al., 2022), and dissolved solutes, including trace elements that can be scavenged on particles (Ohnemus et al., 2019).

- Chapter 3 provided a new estimate of the global particle export flux from the base of the euphotic zone and the maximum mixed layer, building on the particle size distribution reconstructions discussed in Chapter 2. In this Chapter I provide a detailed characterization of the climatological export from these two depth horizons, highlighting spatial patterns, seasonal cycle, and relative contribution of large vs. small particles. In particular, I highlight the likely occurrence of net heterotrophy in the lower part of the euphotic zone, which slightly reduces the magnitude of the net export relative to the mixed layer. Ultimately, the results of this Chapter generated a novel, data-constrained estimate of particle export fluxes from two critical depth horizons, providing new context to interpret previous assessments of these quantities, and laying the foundation for a depth-resolved characterization of particle fluxes in the ocean interior.

- Chapter 4 took the two previous Chapters as ground-work, and provided a depth-resolved, data-based estimate of both PSD and particle export fluxes between 100m and 2000m. I quantified the amount of organic matter leaving the surface, and tracked where it was ultimately removed from the sinking pool. I find a particle export flux of 6.0 PgC y^{-1} from the surface ocean (100 m), which attenuates to 0.6 PgC y^{-1} by 1000 m depth, corresponding to a global transfer efficiency of 10%. My attenuation patterns follows the same broad features suggested by recent work (Cram et al., 2018; Weber et al., 2016), while they contrast with earlier estimates. Because of the finer resolution of my approach, I highlight the importance of small-scale, localized oceanographic features for particle export, including major oceanic frontal regions, boundary between water masses, the equatorial band, and oceanic island and coastal waters. My analysis suggests that these small-scale features are critically important in enhancing sinking

fluxes and, potentially, deep-ocean carbon sequestration. Lastly, I find a clear preference for the removal of small organic particles with depth, with particles larger than $418\mu m$ contributing disproportionately more to interior-ocean sinking carbon fluxes and sequestration than smaller particles.

5.2 Ongoing and future work

The work presented in this dissertation represents new advancements towards reducing the uncertainty in global carbon export from the ocean surface, and remineralization in the interior. It also outlines the path for future work, paving the way to a clearer understanding of the role of the ocean in sequestering carbon. In this section, I discuss new directions for future work, and highlight the importance of continued observational and modeling efforts to constrain the biological pump – an essential undertaking in a changing ocean.

Mechanistic understanding of particle flux attenuation is central for ocean biogeochemistry and carbon sequestration, and has received growing attention in recent years, with the development of increasingly complex modeling approaches (Weber et al., 2016; Bianchi et al., 2018; Cram et al., 2018, 2022; Amaral et al., 2022; Nowicki et al., 2022). My work highlights the magnitude and patterns of small particle removal, and the overall rapid attenuation of particle fluxes with depth. However, the causes of these changes, and their variability with depth and region remain uncertain, reflecting a poorly-constrained combination of remineralization, aggregation and disaggregation, and injection by vertically migrating animals. Currently, the best approaches to quantify the relative contribution of these processes are ocean circulation and biogeochemistry inverse models (Amaral et al., 2022), which remain limited in their ability to represent complex processes involving multiple particle sizes, and explicit mechanistic modelsDevries et al. (2014); Cram et al. (2022), which can provide a more detailed representation of particle transformations but remain harder to constrain observationally. To this end, I suggest a combination of the two modelling approaches to quantify magnitude and patterns in particle transformation rates. Critically, my reconstructions of PSD and sinking fluxes, which are resolved by particle size, depth, region, and time

of the year, can provide missing targets to constrain these models and apply them to the global ocean.

Secondly, multiple different observational databases of marine particles (both sinking and non-sinking) exist. However, these different types of observations are rarely used jointly. Recent studies have begun to use observations from particle pumps to estimate the sinking speed and carbon content of organic particles (Xiang et al., 2022). Other methods have similarly measured sinking carbon fluxes, although without detailed size information. However, most of these methods have never been combined to paint a consistent picture of sinking export fluxes. To address this gap, I suggest a two-part research direction. First, conduct a thorough inter-comparison of the different methodologies to observing sinking organic particle fluxes, encompassing multiple sites and oceanographic regimes whenever possible. Second, use the lessons learned from the meta-analysis to combine multiple observations of the biological pump, and provide tighter observational constraints for quantities discussed in this dissertation, in particular particle type, carbon content, sinking speeds. These additional observational constraint could then be used to provide an update to the particle size distribution and sinking flux estimates shown in this dissertation.

Together with efforts aimed at improving mechanistic models and reducing uncertainties in the biological pump, continued monitoring of *in situ* particles is critically important in a changing ocean. As the ocean warms up in response to carbon dioxide emissions to the atmosphere, its continuing ability to mitigate this change by sequestering carbon remains uncertain. For example, model projections indicate that the magnitude of primary productivity and particle export is expected to decrease over most of the low-latidude ocean Bopp et al. (2013), as the supply of nutrients to the surface is curtailed by increasing stratification. In parallel, a projected reduction in the concentration of oxygen in the water column Bopp et al. (2013), may work to reduce remineralization rates and increase the transfer efficiency Cavan et al. (2017), in turn favoring carbon sequestration. Additionally, in a more stratified ocean, water mass ventilation rates are likely to decrease Gnanadesikan et al. (2007); Bopp et al. (2013), allowing more remineralized carbon to accumulate in the interior Weber et al. (2016); Boyd et al. (2019). The combined effect of these and other competing changes

on ocean carbon sequestration remain uncertain. Ultimately, as optical particle observations increase in number, the work presented in this dissertation provides an unique ability to diagnose not only the current patterns of particle export and carbon sequestration, but also their changes over time, and the possible causes behind these changes, thus reducing uncertainties in future climate projections.

In this dissertation, I present a new statistical approach applied to a growing dataset of particle observations that revealed novel patterns of particle export and remineralization. As optical observations becomes more common, and the approach described here is refined, rapid quantification of patterns and trends in particle export becomes possible, allowing monitoring of sinking carbon fluxes, providing a baseline to evaluate changes in the biological pump, and serving as a constraint for Earth system models. These are all essential undertakings to better understand and mitigate ongoing changes in ocean biogeochemistry and ecosystem as a result of the climate emergency.

APPENDIX A

Supporting Information for Chapter 2

This sections provides additional figures not provided in chapter 2. Each figure is referenced within the text of the chapter and is meant to support the understanding of the analysis in the chapter.

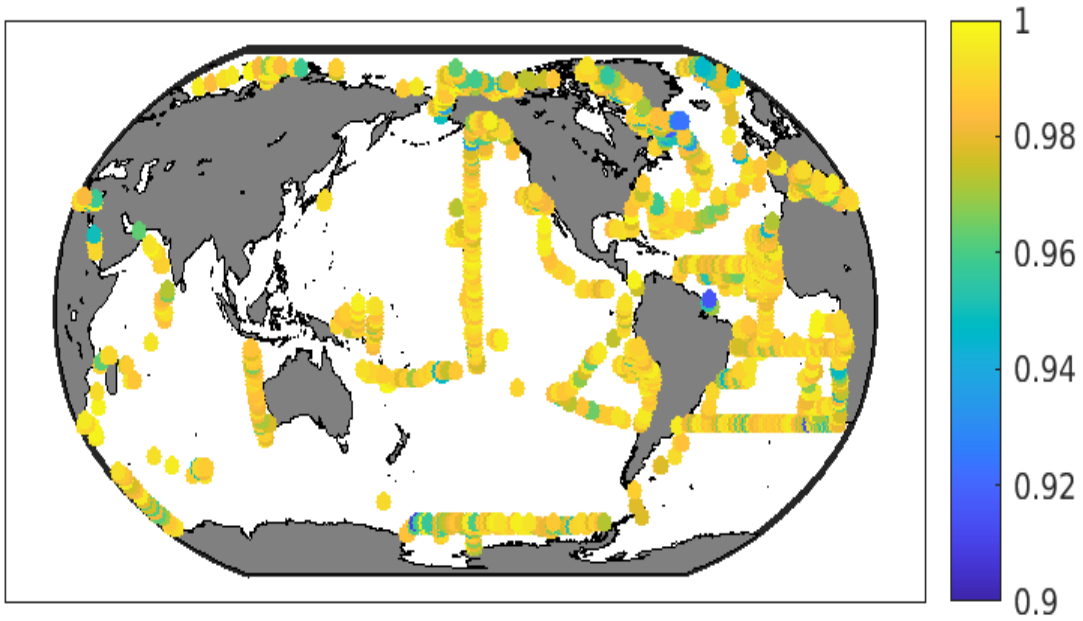

Figure A.1: The correlation coefficient (R^2) of the linear least squares fits between the log of the particle counts and the log of their size. The slope and intercept of these fits are the powerlaw parameters to describe the PSD. r-squared less than 0.9 are removed from the dataset, before extrapolating globally.

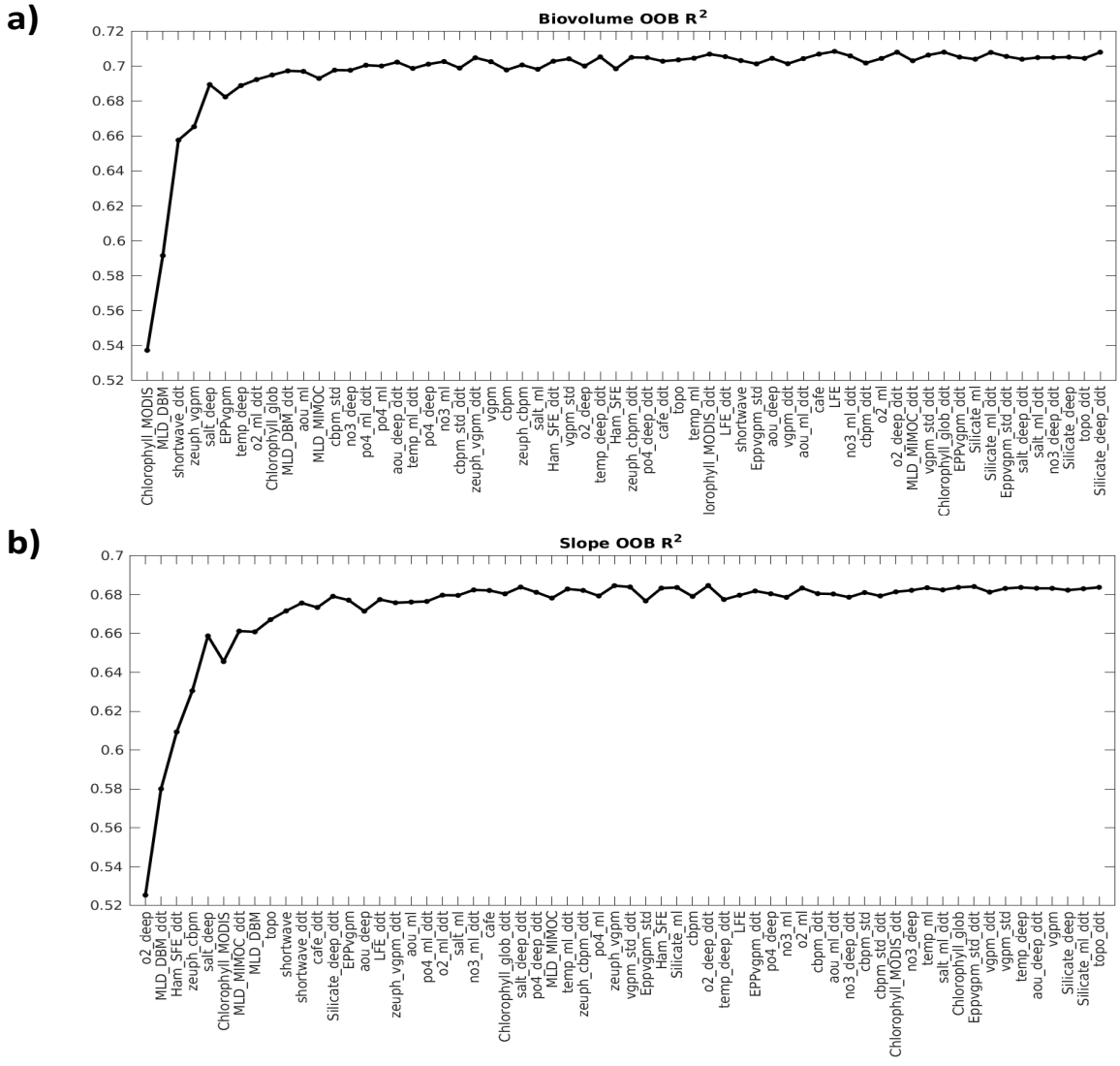

Figure A.2: The relative importance of each predictor for each predicted variable from the Euphotic zone. The y-axis shows the r-squared of a random forest, using all predictors at and to the point along the x-axis. It shows the strength of adding additional predictors. A) shows the ranking for the Biovolume, and B) for the Slope.

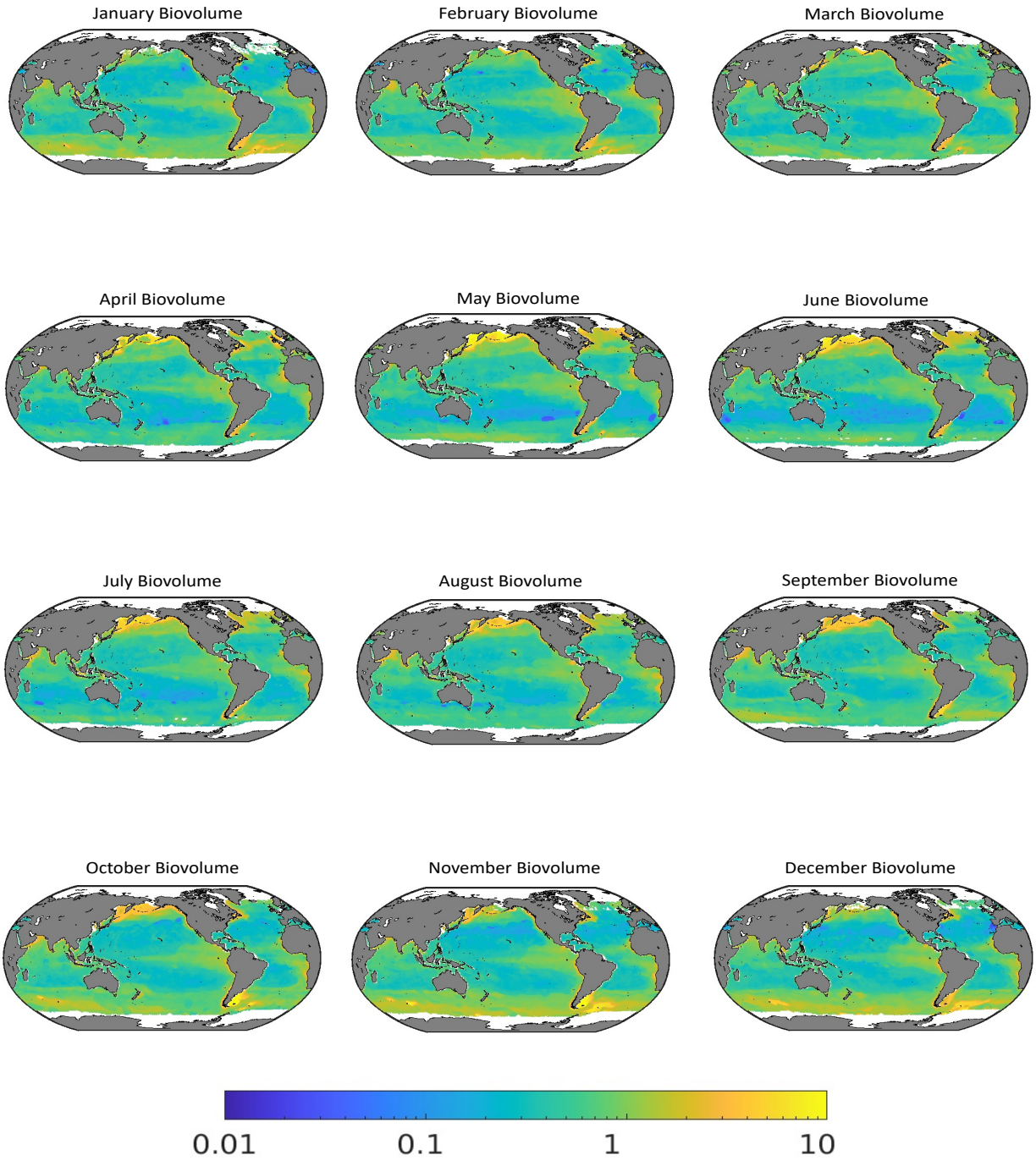

Figure A.3: Particulate carbon biovolume at the base of the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies.

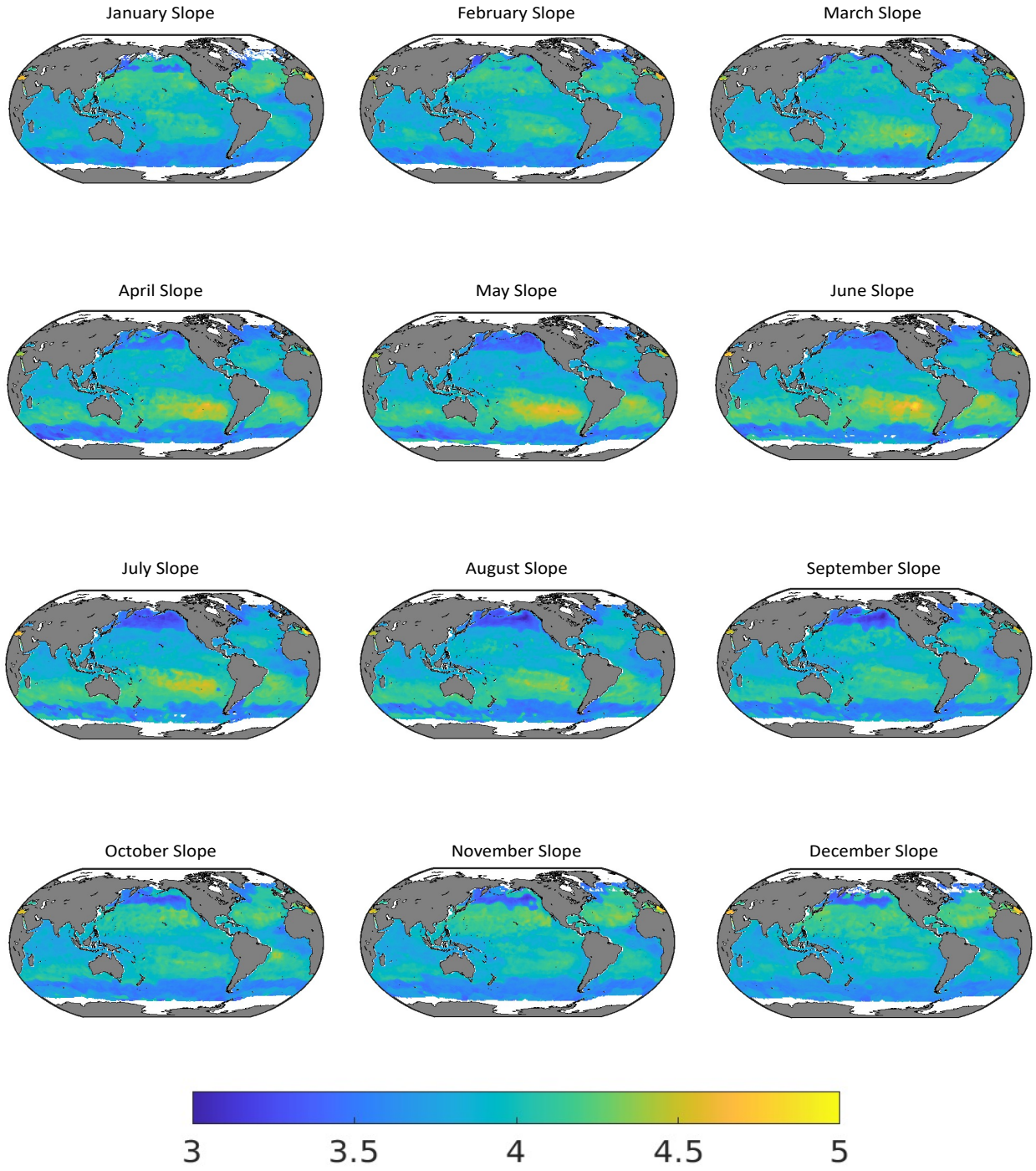

Figure A.4: Particulate carbon PSD slope at the base of the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies.

APPENDIX B

Supporting Information for Chapter 3

B.1 Supplemental figures

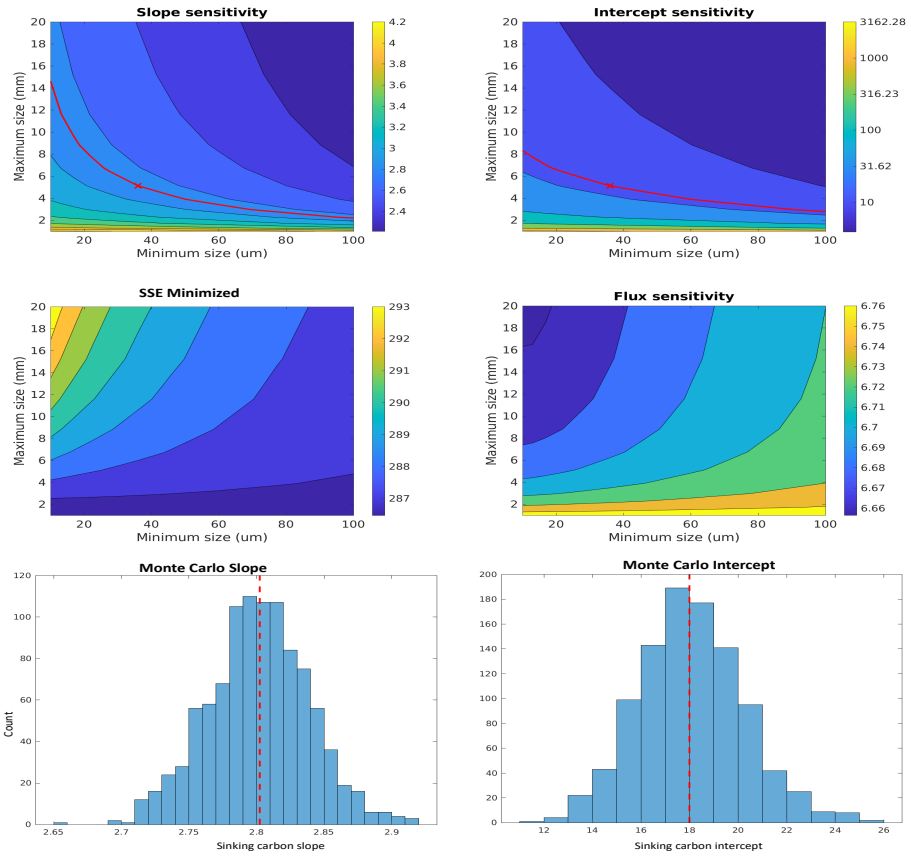

Figure B.1: (A-B) Sensitivty of the intercept(A) and slope (B) parameters, red line indicates the values used by this study. (C-D) Contour map showing the changes in the sum of the squared error (C) and export(D) which set the parameters chosen. E-F shows the range of slope and intercept values possible for this size range and histogram of result from the Monte Carlo optimization.

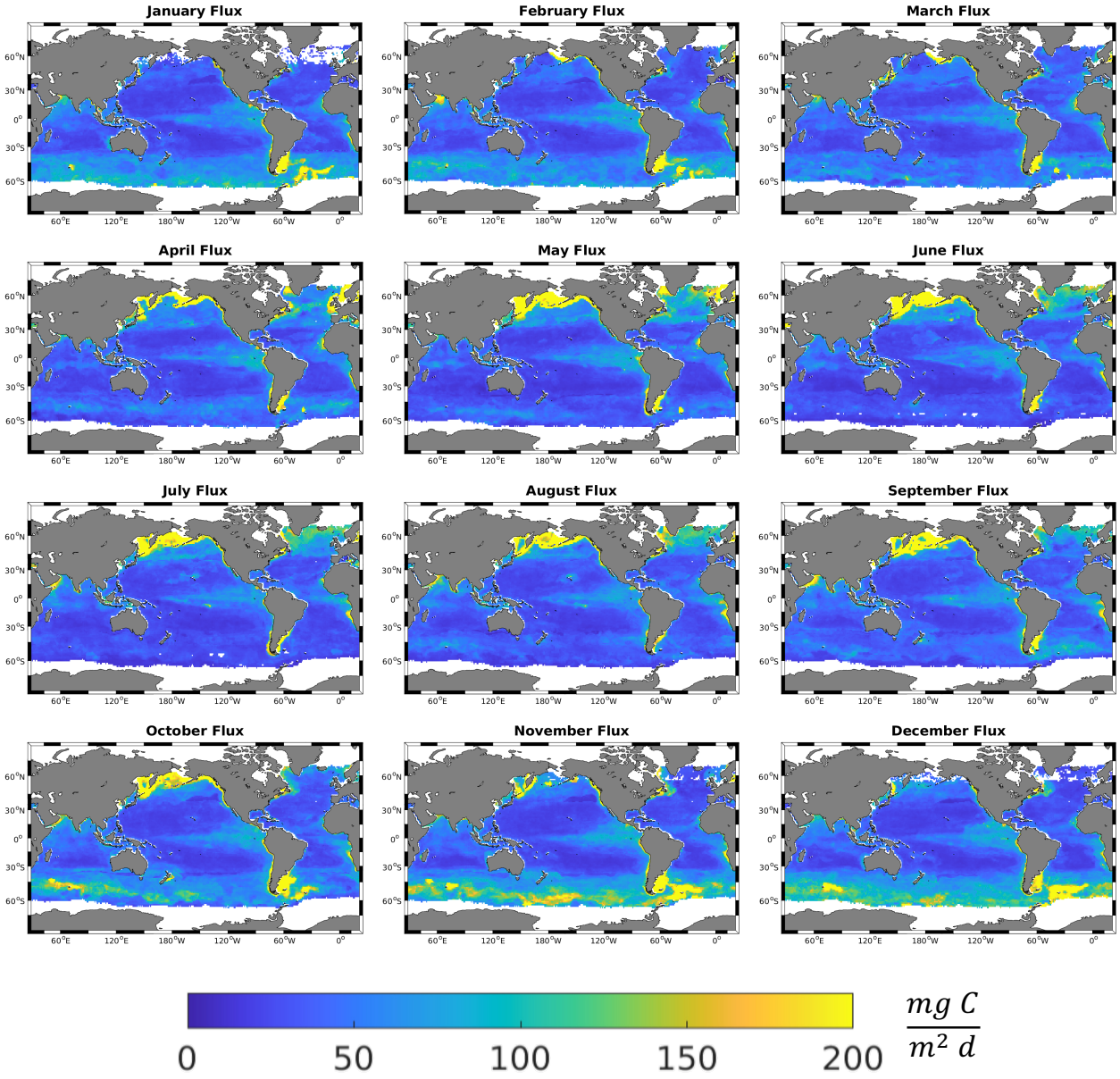

Figure B.2: Particulate carbon flux from the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies.

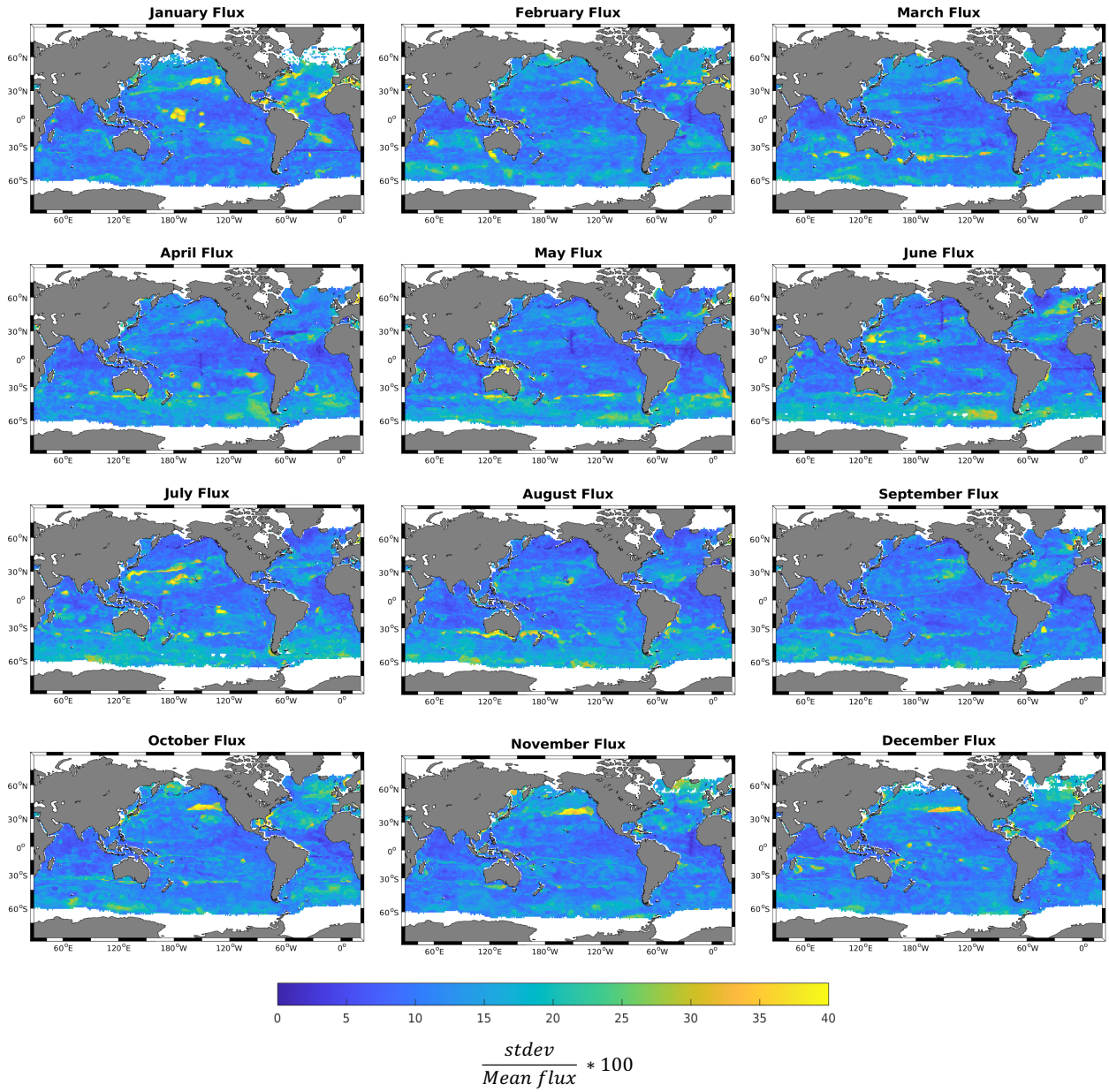

Figure B.3: Particulate carbon flux standard deviation from the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies of error. Color bar denotes percentage of error (error/flux).

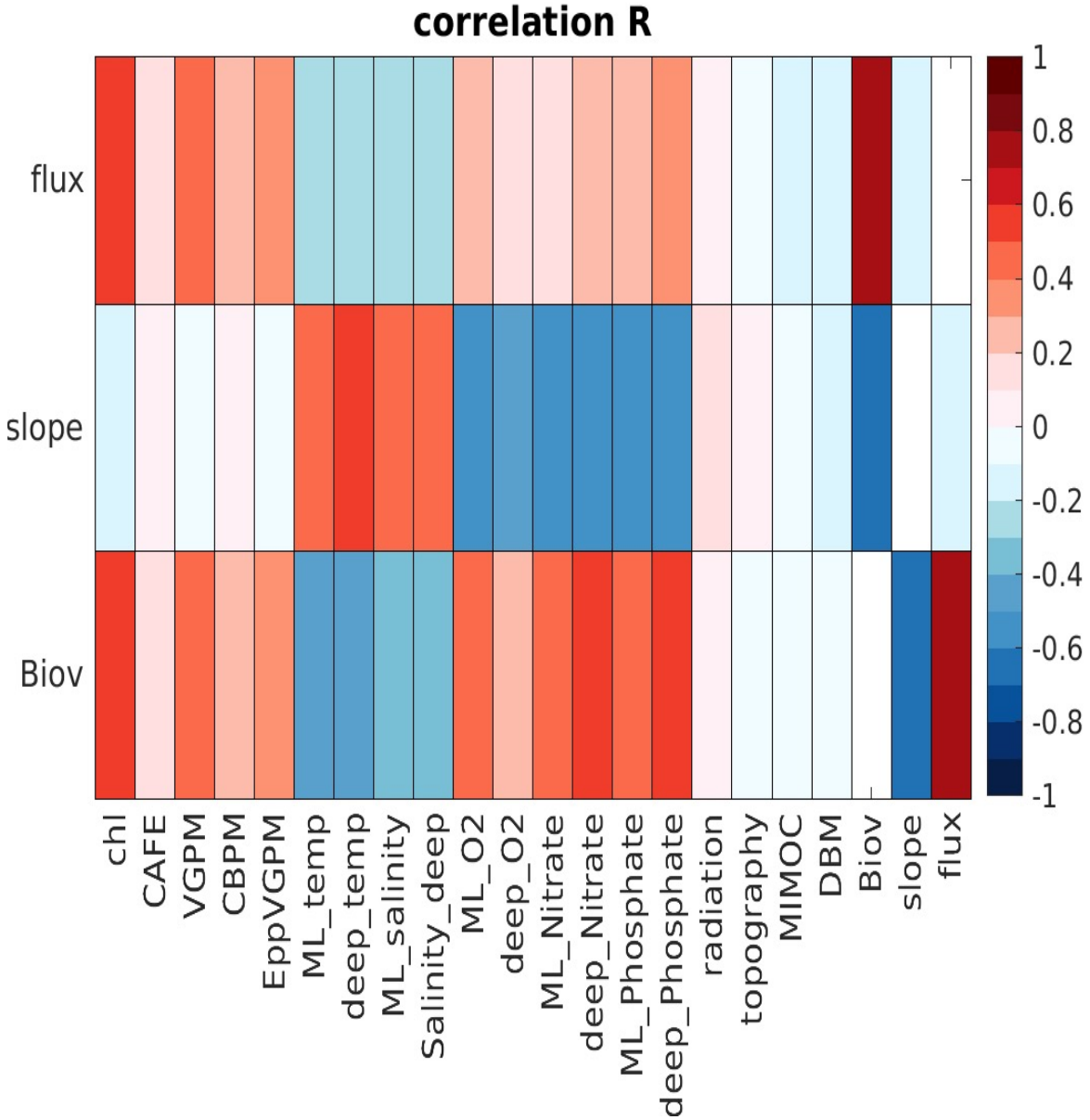

Figure B.4: A correlation matrix showing the correlation coefficients (R^2) between each predictor and our three reconstructed variable, the slope biovolume, and flux.

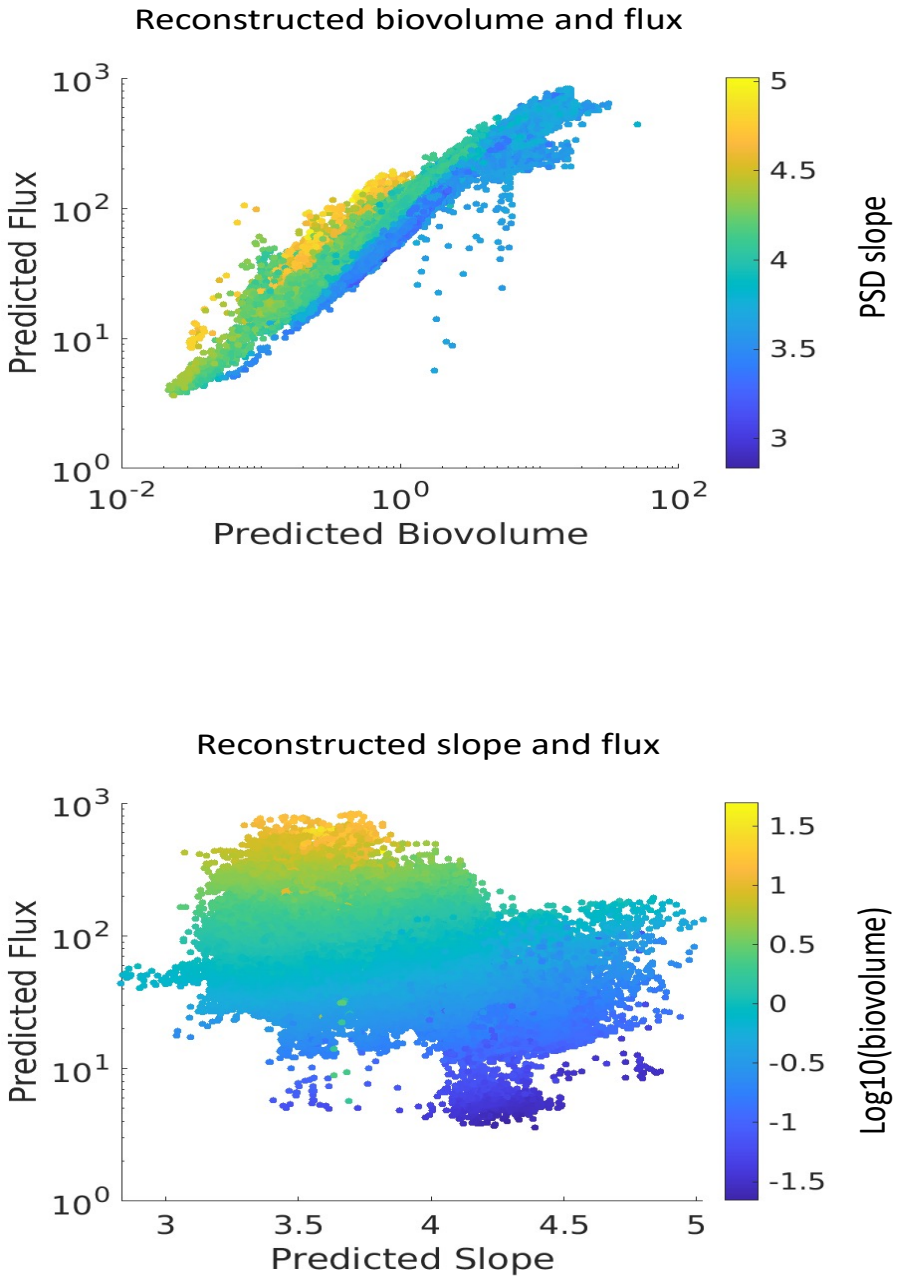

Figure B.5: Scatter plots showing the correlation between flux, and the PSD parameters. A) shows the log-log relationship between biovolume and flux, the colorbar shows the associated slope. B) shows the relationship between the log of flux and slope. The colorbar shows the log of biovolume.

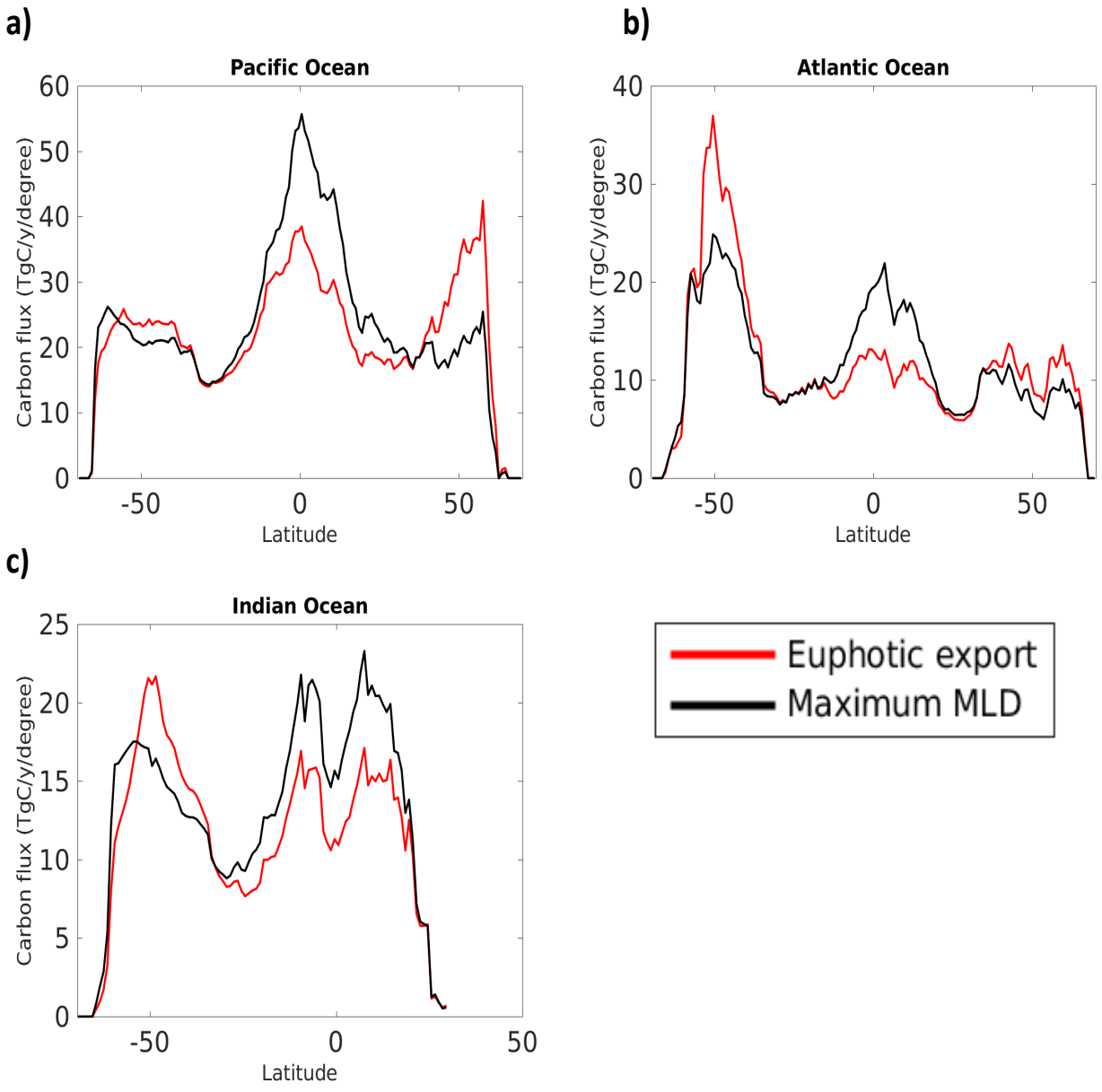

Figure B.6: Meridionally averaged export of carbon for each of the three main ocean basins. Black indicated the export from the maximum mixed layer. The red line indicates the flux from the euphotic zone.

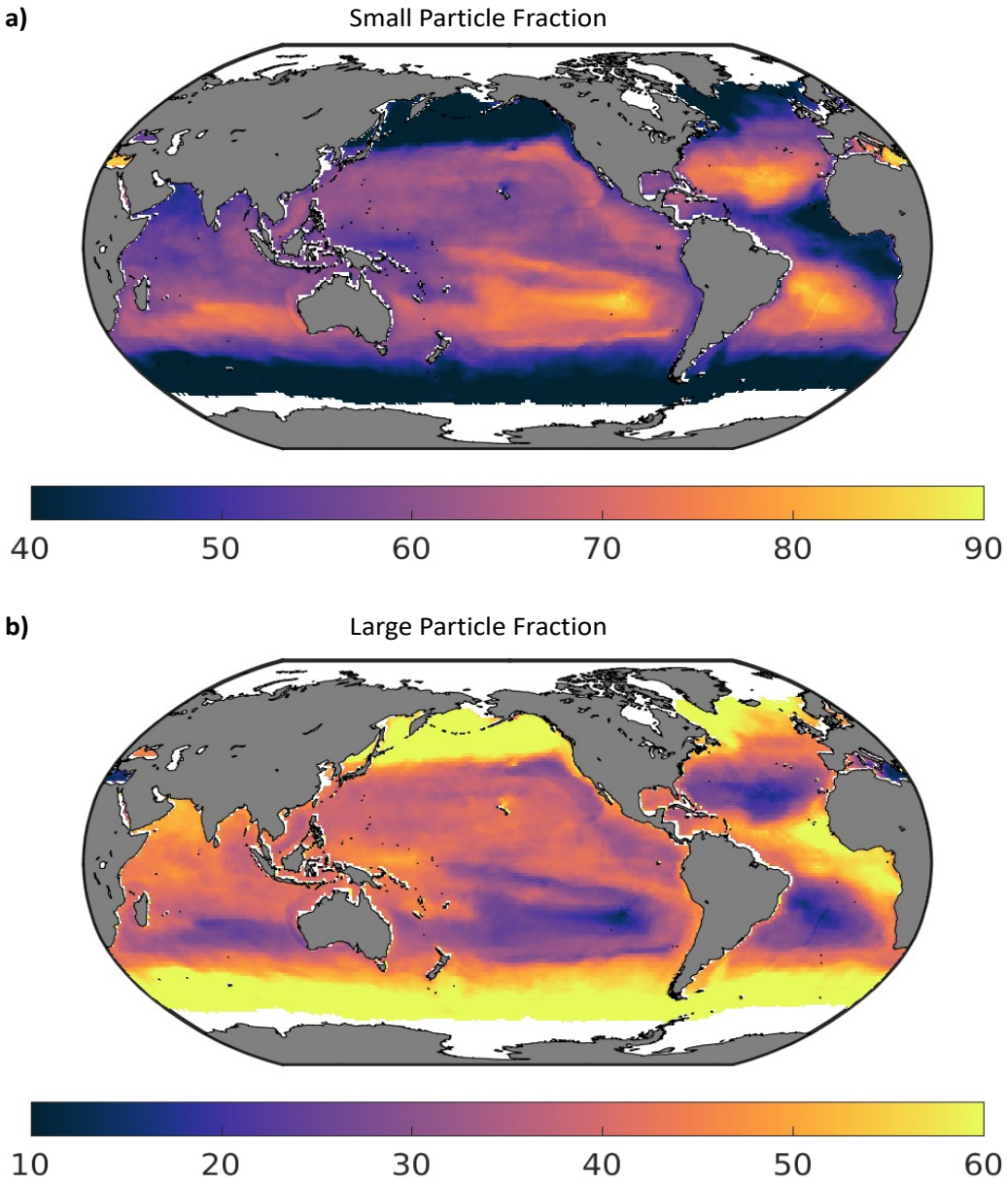

Figure B.7: Fraction of the particle flux contribution from particles 35μm to 1mm in ESD (A) and of particles 1mm to 5mm in ESD(B).

APPENDIX C

Supporting Information for Chapter 4

C.1 Quantification of sinking carbon parameters

As described in section 4.3.3 we are able to optimize our PSD results against *in situ* observations of particle flux. Here we show table C.1 which shows the mean and standard deviation of our optimization scheme at each defined hinge point. Values shows here can be used to estimate the particle flux, and sinking carbon profiles.

Table C.1: The mean sinking parameters ($\pm SD$) from the sinking carbon optimization. Intercept refers to the size independent coefficient, and slope the size-dependent component.

Variable	**100 m**	**458 m**	**1000 m**
Intercept	47.6 ±16.5	37.5±17.3	33.5±9.8
Slope	3.16 ±0.14	3.92±0.35	3.56±0.19

C.2 Calculation of Transfer efficiency

As discussed in Chapter 4, it possible to relate the Martin coefficient (b) to the particle transfer efficiency. For this study, we use this relationship to determine the particle transfer efficiency for each study listed in Fig. C.4. To do this, we use equation 4.1 and 4.2, and solve for the relationship as below:

$$\frac{\phi_z}{\phi_0} = (\frac{z_z}{z_0})^{-b}. \tag{C.1}$$

For this studies effective martin curve, we invert the relationship here, and solve for b. Here we define our effective depths of the surface (z_0) to be 100m, and the deep estimate (z_z) to be 1000m. Thus our effective martin curve is calculated using the form:

$$b = \frac{log_{10}(T_{eff})}{log_{10}(\frac{1000}{100})} \tag{C.2}$$

However, since all other studies described in fig. C.4 use their initial depth (z_0) to be at the base of the variable euphotic zone, it is necessary to "correct" all data onto the same depths used in this study (i.e. correct ϕ_0 and z_0 to 100 m).

We use the Martin equation (e.q. 4.1), and globally variable "b" values, to calculate the "transfer efficiency" of particle between the variable euphotic zone and both 100m ($T_{eff}100$) and 1000m ($T_{eff}1000$) using equation C.1. We then use equation C.3 to determine the t_{eff} as shown in figure 4.7.

$$\frac{\phi_{1000}}{\phi_{100}} = \frac{\phi_{1000}}{\phi_{zeu}} \cdot \frac{\phi_{zeu}}{\phi_{100}}. \tag{C.3}$$

C.3 Supplemental figures

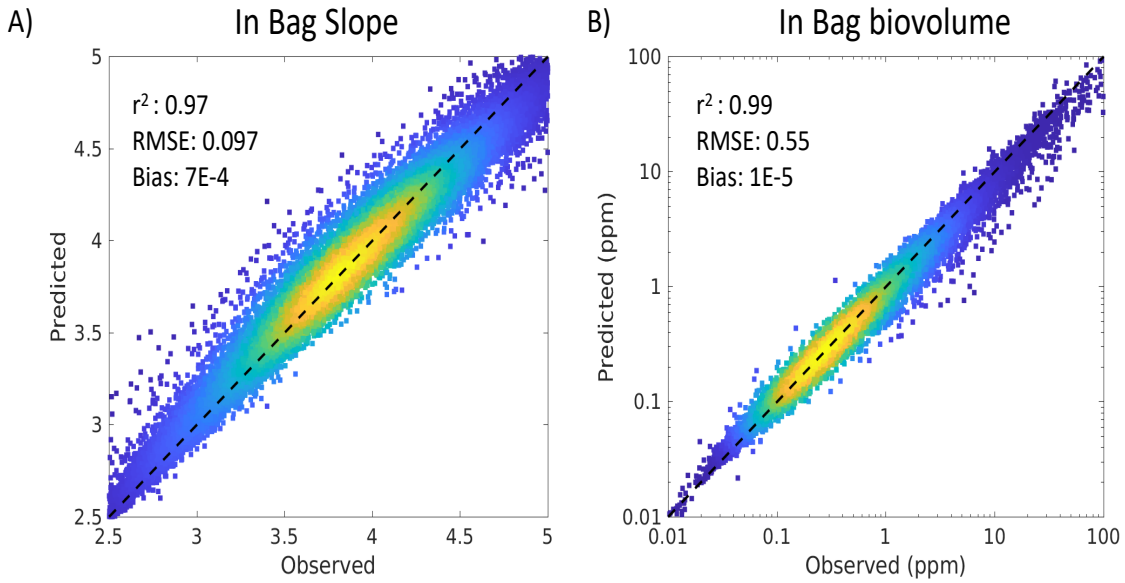

Figure C.1: In bag scatter plot for (A) the particle slope and (B) Biovolume. Annotations shoow the relevant statistics indicating robustness of the model. In Bag performance denotes the performance of our model on data used to train the model, thus may be susceptible to overfitting.

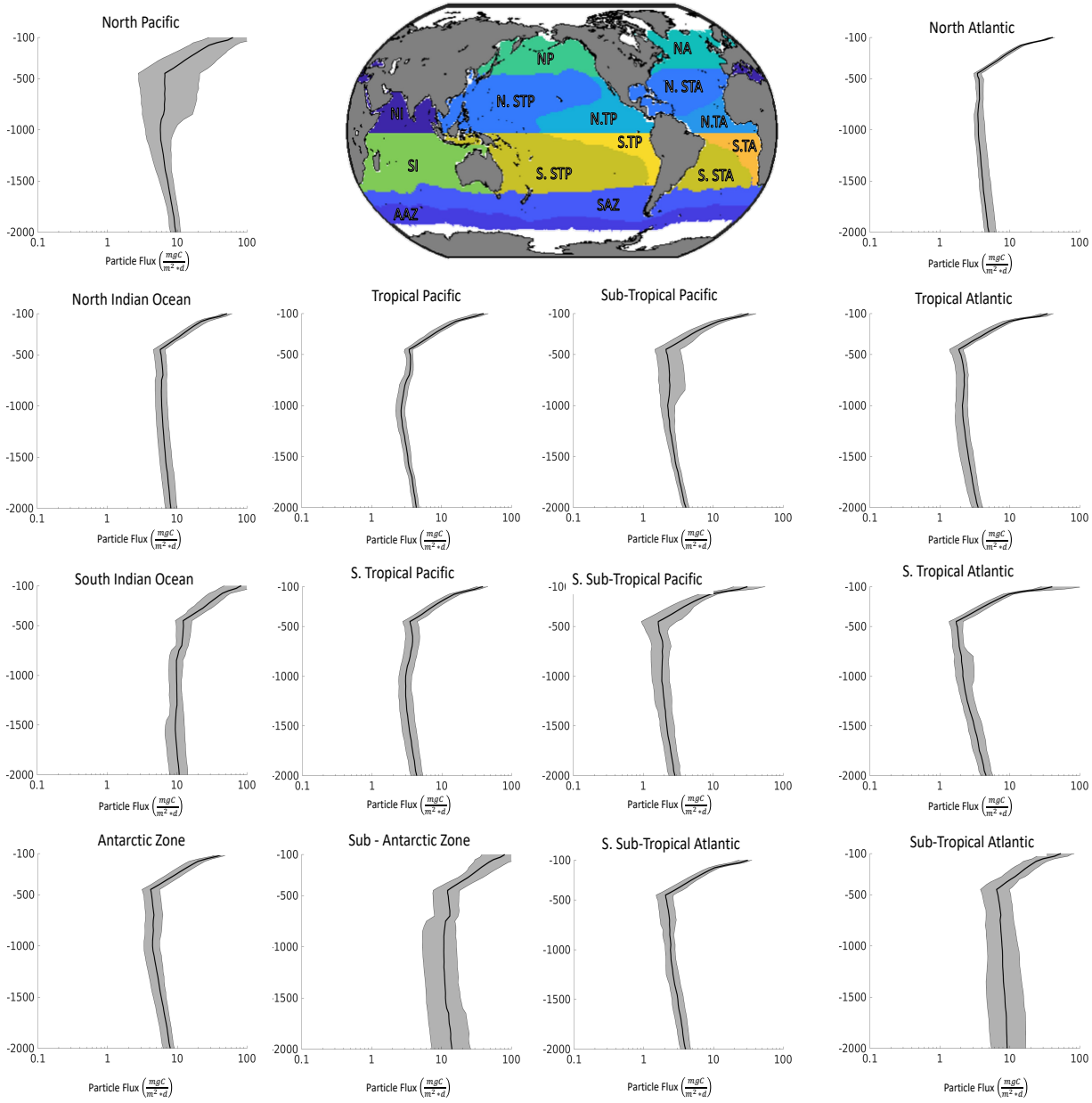

Figure C.2: Particulate carbon flux from the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies. The grey shading describes the seasonal range of the mean flux for each region.

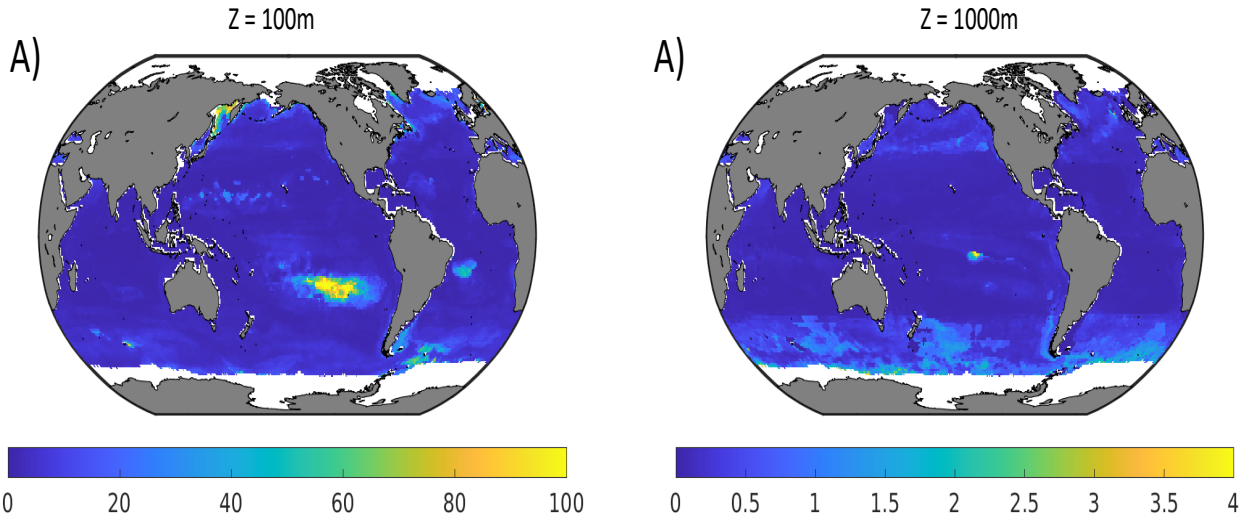

Figure C.3: Particulate carbon flux standard deviation from the euphotic zone reconstructed from the random forest calculations, showing monthly climatologies of error. Color bar denotes percentage of error (error/flux).

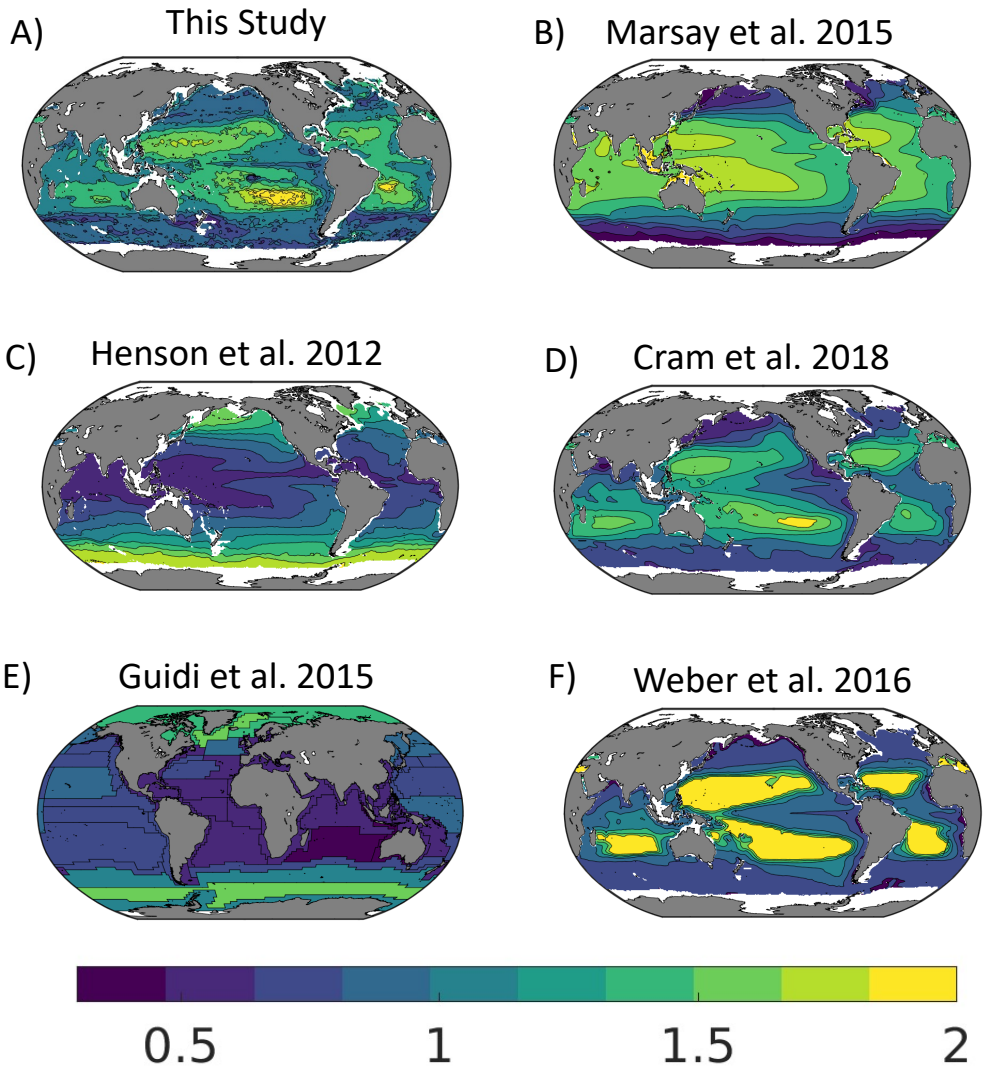

Figure C.4: The Effective Martin curve coefficient (b) for A) This Study, B) Marsay et al. (2015), C) Henson et al. (2012), D) Cram et al. (2018), E) Guidi et al. (2015), and F) Weber et al. (2016). For this study, b is defined as the Martin curve coefficient in equation 4.1.

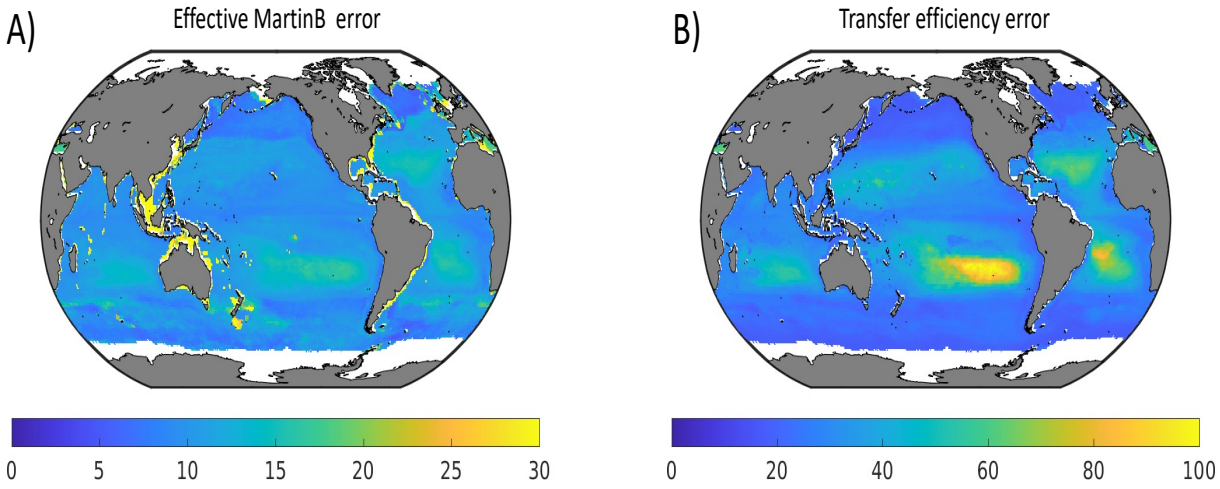

Figure C.5: (A)Error estimate for our effective martin curve depicted as the standard deviation from the our ensemble of sinking carbon estimates. (B) is the same as in (A) bute denotes the error of the calculated transfer efficiencies. Color bar denotes percentage of error (error/flux)

References

Alldredge, A. (1998). The carbon, nitrogen and mass content of marine snow as a function of aggregate size. *Deep-Sea Research Part I: Oceanographic Research Papers*, *45*(4-5), 529–541. doi: 10.1016/S0967-0637(97)00048-4

Alldredge, A. L., & Gotschalk, C. (1988). In situ settling behavior of marine snow. *Limnology and Oceanography*, *33*(3), 339–351. doi: 10.4319/lo.1988.33.3.0339

Alonso-González, I. J., Arístegui, J., Lee, C., Sanchez-Vidal, A., Calafat, A., Fabrés, J., ... Benítez-Barrios, V. (2010). Role of slowly settling particles in the ocean carbon cycle. *Geophysical Research Letters*, *37*(13). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010GL043827` doi: https://doi.org/10.1029/2010GL043827

Amaral, V. J., Lam, P. J., Marchal, O., Roca-Martí, M., Fox, J., & Nelson, N. B. (2022, aug). Particle cycling rates at Station P as estimated from the inversion of POC concentration data. *Elementa: Science of the Anthropocene*, *10*(1), 18. Retrieved from `https://doi.org/10.1525/elementa.2021.00018` doi: 10.1525/elementa.2021.00018

Antoine, D., & Morel, A. (1996, mar). Oceanic primary production: 1. Adaptation of a spectral light-photosynthesis model in view of application to satellite chlorophyll observations. *Global Biogeochemical Cycles*, *10*(1), 43–55. Retrieved from `http://doi.wiley.com/10.1029/95GB02831` doi: 10.1029/95GB02831

Arteaga, L., Haëntjens, N., Boss, E., Johnson, K. S., & Sarmiento, J. L. (2018). Assessment of Export Efficiency Equations in the Southern Ocean Applied to Satellite-Based Net Primary Production. *Journal of Geophysical Research: Oceans*, *123*(4), 2945–2964. doi: 10.1002/2018JC013787

Bach, L. T., Paul, A. J., Boxhammer, T., von der Esch, E., Graco, M., Schulz, K. G., ... Riebesell, U. (2020). Factors controlling plankton community production, export flux, and particulate matter stoichiometry in the coastal upwelling system off Peru. *Biogeosciences*,

17(19), 4831–4852. Retrieved from https://bg.copernicus.org/articles/17/4831/2020/ doi: 10.5194/bg-17-4831-2020

Bader, H. (1970). The hyperbolic distribution of particle sizes. *Journal of Geophysical Research (1896-1977)*, *75*(15), 2822-2830. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC075i015p02822 doi: https://doi.org/10.1029/JC075i015p02822

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., ... Ward, B. A. (2013). The biogeography of marine plankton traits. *Ecology Letters*, *16*(4), 522–534. doi: 10.1111/ele.12063

Behrenfeld, M. J., & Boss, E. S. (2014). Resurrecting the ecological underpinnings of ocean plankton blooms. *Annual review of marine science*, *6*, 167–194.

Behrenfeld, M. J., & Falkowski, P. G. (1997, jan). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*(1), 1–20. Retrieved from https://www.google.com/search?q=Engle+et+al.{\%}2C2000{\&}oq=Engle+et+al.{\%}2C2000{\&}aqs=chrome..69i57.11777j0j8{\&}sourceid=chrome{\&}ie=UTF-8http://doi.wiley.com/10.4319/lo.1997.42.1.0001 doi: 10.4319/lo.1997.42.1.0001

Berelson, W. M. (2001a, 10). The flux of particulate organic carbon into the ocean interior: A comparison of four u.s. jgofs regional studies. *Oceanography*, *14*(4), 59-67. Retrieved from https://doi.org/10.5670/oceanog.2001.07

Berelson, W. M. (2001b). Particle settling rates increase with depth in the ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, *49*(1), 237–251. Retrieved from https://www.sciencedirect.com/science/article/pii/S0967064501001023 doi: https://doi.org/10.1016/S0967-0645(01)00102-3

Bianchi, D., Stock, C., Galbraith, E. D., & Sarmiento, J. L. (2013). Diel vertical migration: Ecological controls and impacts on the biological pump in a one-dimensional

ocean model. *Global Biogeochemical Cycles*, *27*(2), 478–491. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/gbc.20031 doi: https://doi.org/10.1002/gbc.20031

Bianchi, D., Weber, T. S., Kiko, R., & Deutsch, C. (2018). Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nature Geoscience*, 1–6. Retrieved from http://dx.doi.org/10.1038/s41561-018-0081-0 doi: 10.1038/s41561-018-0081-0

Bishop, J. K., Fong, M. B., & Wood, T. J. (2016). Robotic observations of high wintertime carbon export in California coastal waters. *Biogeosciences*, *13*(10), 3109–3129. doi: 10.5194/bg-13-3109-2016

Bisson, K., Boss, E., Werdell, P. J., Ibrahim, A., Frouin, R., & Behrenfeld, M. (2021). Seasonal bias in global ocean color observations. *Applied optics*, *60*(23), 6978–6988.

Bisson, K., & Cael, B. (2021). How are under ice phytoplankton related to sea ice in the southern ocean? *Geophysical Research Letters*, *48*(21), e2021GL095051.

Bisson, K. M., Boss, E., Werdell, P. J., Ibrahim, A., Frouin, R., & Behrenfeld, M. J. (2021, Aug). Seasonal bias in global ocean color observations. *Appl. Opt.*, *60*(23), 6978–6988. Retrieved from http://opg.optica.org/ao/abstract.cfm?URI=ao-60-23-6978 doi: 10.1364/AO.426137

Bisson, K. M., Siegel, D. A., DeVries, T., Cael, B. B., & Buesseler, K. O. (2018). How Data Set Characteristics Influence Ocean Carbon Export Models. *Global Biogeochemical Cycles*, *32*(9), 1312–1328. doi: 10.1029/2018GB005934

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., ... others (2013). Multiple stressors of ocean ecosystems in the 21st century: projections with cmip5 models. *Biogeosciences*, *10*(10), 6225–6245.

Boss, E., Guidi, L., Richardson, M. J., Stemmann, L., Gardner, W., Bishop, J. K., ... Sherrell, R. M. (2015). Optical techniques for remote and in-situ characterization of particles pertinent to geotraces. *Progress in Oceanography*, *133*, 43–54.

Bourne, H. L., Bishop, J. K., Wood, T. J., Loew, T. J., & Liu, Y. (2019). Carbon Flux Explorer optical assessment of C, N and P fluxes. *Biogeosciences*, *16*(6), 1249–1264. doi: 10.5194/bg-16-1249-2019

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., & Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature*, *568*(7752), 327–335. Retrieved from `https://doi.org/10.1038/s41586-019-1098-2` doi: 10.1038/s41586-019-1098-2

Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of co2 by the oceans. *Science*, *367*(6479), 791–793.

Broecker, W. S., Peng, T.-H., et al. (1982). *Tracers in the sea* (Vol. 690). Lamont-Doherty Geological Observatory, Columbia University Palisades, New York.

Buesseler, K., Antia, A., Chen, M., Fowler, S., Gardner, W., Gustafsson, O., ... Trull, T. (2007, 05). An assessment of the use of sediment traps for estimating upper ocean particle fuxes. *Journal of Marine Research*, *65*. doi: 10.1357/002224007781567621

Buesseler, K. O., & Boyd, P. W. (2009). Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnology and Oceanography*, *54*(4), 1210–1232. doi: 10.4319/lo.2009.54.4.1210

Buesseler, K. O., Boyd, P. W., Black, E. E., & Siegel, D. A. (2020). Metrics that matter for assessing the ocean biological carbon pump. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9679–9687. doi: 10.1073/pnas.1918114117

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., ... others (2007). Revisiting carbon flux through the ocean's twilight zone. *science*, *316*(5824), 567–570.

Burd, A. B., Hansell, D. A., Steinberg, D. K., Anderson, T. R., Arístegui, J., Baltar, F., ... others (2010). Assessing the apparent imbalance between geochemical and biochemical indicators of meso-and bathypelagic biological activity: What the $#! is wrong with present calculations of carbon budgets? *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(16), 1557–1571.

Burd, A. B., & Jackson, G. A. (2009). Particle aggregation. *Annual Review of Marine Science*, *1*(1), 65–90. Retrieved from `http://www.annualreviews.org/doi/10.1146/annurev.marine.010908.163904` doi: 10.1146/annurev.marine.010908.163904

Cael, B., Cavan, E. L., & Britten, G. L. (2021). Reconciling the size-dependence of marine particle sinking speed. *Geophysical Research Letters*, *48*(5), e2020GL091771.

Cael, B. B., & Bisson, K. (2018). Particle Flux Parameterizations: Quantitative and Mechanistic Similarities and Differences. *Frontiers in Marine Science*, *5*. Retrieved from `https://www.frontiersin.org/articles/10.3389/fmars.2018.00395` doi: 10.3389/fmars.2018.00395

Carlson, C. A., Ducklow, H. W., & Michaels, A. F. (1994). Annual flux of dissolved organic carbon from the euphotic zone in the northwestern sargasso sea. *Nature*, *371*(6496), 405–408.

Cavan, E. L., Trimmer, M., Shelley, F., & Sanders, R. (2017). Remineralization of particulate organic carbon in an ocean oxygen minimum zone. *Nature Communications*, *8*(1), 1–9.

Cavender-Bares, K. K., Rinaldo, A., & Chisholm, S. W. (2001). Microbial size spectra from natural and nutrient enriched ecosystems. *Limnology and Oceanography*, *46*(4), 778–789.

Church, M. J., Kyi, E., Hall Jr, R. O., Karl, D. M., Lindh, M., Nelson, A., & Wear, E. K. (2021). Production and diversity of microorganisms associated with sinking particles in the subtropical north pacific ocean. *Limnology and Oceanography*, *66*(9), 3255–3270.

Clements, D. J., Yang, S., Weber, T., McDonnell, A. M. P., Kiko, R., Stemmann, L., & Bianchi, D. (2022). Constraining the Particle Size Distribution of Large Marine Par-

ticles in the Global Ocean With In Situ Optical Observations and Supervised Learning. *Global Biogeochemical Cycles*, *36*(5), e2021GB007276. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GB007276 doi: https://doi.org/10.1029/2021GB007276

Clements, D. J., Yang, S., Weber, T., McDonnell, A. M. P., Kiko, R., Stemmann, L., & Bianchi, D. (2023). New estimate of organic carbon export from optical measurements reveals the role of particle size distribution and export depth. *In Revision for Global Biogeochemical Cycles*.

Close, H. G., Shah, S. R., Ingalls, A. E., Diefendorf, A. F., Brodie, E. L., Hansman, R. L., ... Pearson, A. (2013). Export of submicron particulate organic matter to mesopelagic depth in an oligotrophic gyre. *Proceedings of the National Academy of Sciences*, *110*(31), 12565-12570. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.1217514110 doi: 10.1073/pnas.1217514110

Copernicus C.C.S. (2017). *Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate.* Copernicus Climate Change Service Climate Data Store (CDS). (accessed: 11-13-2019)

Cram, J. A., Fuchsman, C. A., Duffy, M. E., Pretty, J. L., Lekanoff, R. M., Neibauer, J. A., ... McDonnell, A. M. P. (2022). Slow particle remineralization, rather than suppressed disaggregation, drives efficient flux transfer through the eastern tropical north pacific oxygen deficient zone. *Global Biogeochemical Cycles*, *36*(1), e2021GB007080. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GB007080 (e2021GB007080 2021GB007080) doi: https://doi.org/10.1029/2021GB007080

Cram, J. A., Weber, T., Leung, S. W., McDonnell, A. M., Liang, J. H., & Deutsch, C. (2018). The Role of Particle Size, Ballast, Temperature, and Oxygen in the Sinking Flux to the Deep Sea. *Global Biogeochemical Cycles*, *32*(5), 858–876. doi: 10.1029/2017GB005710

Dall'Olmo, G., Dingle, J., Polimene, L., Brewin, R. J., & Claustre, H. (2016). Substantial energy input to the mesopelagic ecosystem from the seasonal mixed-layer pump. *Nature Geoscience*, *9*(11), 820–823. doi: 10.1038/ngeo2818

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 C: Oceans*, *109*(12), 1–20. doi: 10.1029/2004JC002378

DeLong, E. F., Franks, D. G., & Alldredge, A. L. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnology and oceanography*, *38*(5), 924–934.

DeVries, T., Holzer, M., & Primeau, F. (2017). Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature*, *542*(7640), 215–218. Retrieved from `http://dx.doi.org/10.1038/nature21068` doi: 10.1038/nature21068

Devries, T., Liang, J. H., & Deutsch, C. (2014). A mechanistic particle flux model applied to the oceanic phosphorus cycle. *Biogeosciences*, *11*(19), 5381–5398. doi: 10.5194/bg-11-5381-2014

DeVries, T., Primeau, F., & Deutsch, C. (2012). The sequestration efficiency of the biological pump. *Geophysical Research Letters*, *39*(13).

DeVries, T., & Weber, T. (2017). The export and fate of organic matter in the ocean: New constraints from combining satellite and oceanographic tracer observations. *Global Biogeochemical Cycles*, *31*(3), 535–555. doi: 10.1002/2016GB005551

Dilling, L., & Alldredge, A. L. (2000). Fragmentation of marine snow by swimming macrozooplankton: A new process impacting carbon cycling in the sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(7), 1227–1245.

Dunne, J. P., Sarmiento, J. L., & Gnanadesikan, A. (2007). A synthesis of global particle export from the surface ocean and cycling through the ocean interior and on the seafloor. *Global Biogeochemical Cycles*, *21*(4), 1–16. doi: 10.1029/2006GB002907

Durkin, C. A., Estapa, M. L., & Buesseler, K. O. (2015). Observations of carbon export by small sinking particles in the upper mesopelagic. *Marine Chemistry*, *175*, 72–81. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0304420315000390` doi: https://doi.org/10.1016/j.marchem.2015.02.011

D'Errico, J. (2016). *Inpaint nans (matlab central file exchange, 2012).*

Emerson. (2013). Global Biogeochemical Cycles the biological carbon flux in the ocean. *Global Biogeochemical Cycles*, 14–28. doi: 10.1002/2013GB004680

Estapa, M. L., Siegel, D. A., Buesseler, K. O., Stanley, R. H. R., Lomas, M. W., & Nelson, N. B. (2015). Decoupling of net community and export production on submesoscales in the sargasso sea. *Global Biogeochemical Cycles*, *29*(8), 1266-1282. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GB004913` doi: https://doi.org/10.1002/2014GB004913

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *science*, *281*(5374), 200–206.

Fontanez, K. M., Eppley, J. M., Samo, T. J., Karl, D. M., & DeLong, E. F. (2015). Microbial community structure and function on sinking particles in the north pacific subtropical gyre. *Frontiers in microbiology*, *6*, 469.

Forest, A., Stemmann, L., Picheral, M., Burdorf, L., Robert, D., Fortier, L., & Babin, M. (2012). Size distribution of particles and zooplankton across the shelf-basin system in southeast Beaufort Sea: Combined results from an Underwater Vision Profiler and vertical net tows. *Biogeosciences*, *9*(4), 1301–1320. doi: 10.5194/bg-9-1301-2012

Gaiero, D., Probst, J.-L., Depetris, P., Bidart, S., & Leleyter, L. (2003). Iron and other transition metals in patagonian riverborne and windborne materials: geochemical control and

transport to the southern south atlantic ocean. *Geochimica et Cosmochimica Acta*, *67*(19), 3603-3623. Retrieved from https://www.sciencedirect.com/science/article/pii/S0016703703002114 doi: https://doi.org/10.1016/S0016-7037(03)00211-4

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, R., ... Reagan, J. (2018). World Ocean Atlas 2018. Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate and nitrate+nitrite, silicate). *NOAA Atlas NESDIS 84*, *84*(July), 35.

Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, R. A., ... Reagan, J. R. (2019). World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Oxygen Saturation. *NOAA Atlas NESDIS*, *3*(83), 38 pp.

Gehlen, M., Bopp, L., Emprin, N., Aumont, O., Heinze, C., & Ragueneau, O. (2006). Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model. *Biogeosciences*, *3*(4), 521–537.

Giering, S. L., Sanders, R., Lampitt, R. S., Anderson, T. R., Tamburini, C., Boutrif, M., ... others (2014). Reconciliation of the carbon budget in the ocean's twilight zone. *Nature*, *507*(7493), 480–483.

Giering, S. L. C., Cavan, E. L., Basedow, S. L., Briggs, N., Burd, A. B., Darroch, L. J., ... Waite, A. M. (2020). Sinking organic particles in the ocean—flux estimates from in situ optical devices. *Frontiers in Marine Science*, *6*. Retrieved from https://www.frontiersin.org/article/10.3389/fmars.2019.00834 doi: 10.3389/fmars.2019.00834

Gnanadesikan, A., Russell, J. L., & Zeng, F. (2007). How does ocean ventilation change under global warming? *Ocean Science*, *3*(1), 43–53. Retrieved from https://os.copernicus.org/articles/3/43/2007/ doi: 10.5194/os-3-43-2007

Goldthwait, S., Carlson, C., Henderson, G., & Alldredge, A. (2005). Effects of physical fragmentation on remineralization of marine snow. *Marine Ecology Progress Series*, *305*, 59–65.

Gregor, L., & Gruber, N. (2021, 3). Oceansoda-ethz: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, *13*, 777-808. doi: 10.5194/essd-13-777-2021

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., ... Gorsky, G. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470. Retrieved from `http://dx.doi.org/10.1038/nature16942` doi: 10.1038/nature16942

Guidi, L., Jackson, G. A., Stemmann, L., Miquel, J. C., Picheral, M., & Gorsky, G. (2008). Relationship between particle size distribution and flux in the mesopelagic zone. *Deep-Sea Research Part I: Oceanographic Research Papers*, *55*(10), 1364–1374. doi: 10.1016/j.dsr.2008.05.014

Guidi, L., Legendre, L., Reygondeau, G., Uitz, J., Stemmann, L., & Henson, S. A. (2015, jul). A new look at ocean carbon remineralization for estimating deepwater sequestration. *Global Biogeochemical Cycles*, *29*(7), 1044–1059. Retrieved from `http://doi.wiley.com/10.1002/2014GB005063` doi: 10.1002/2014GB005063

Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., ... Gorsky, G. (2009). Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis. *Limnology and Oceanography*, *54*(6), 1951–1963. doi: 10.4319/lo.2009.54.6.1951

Hague, M., & Vichi, M. (2021). Southern ocean biogeochemical argo detect under-ice phytoplankton growth before sea ice retreat. *Biogeosciences*, *18*(1), 25–38.

Hamilton, D. S., Scanza, R. A., Feng, Y., Guinness, J., Kok, J. F., Li, L., ... Mahowald, N. M. (2019). Improved methodologies for Earth system modelling of atmospheric soluble iron and observation comparisons using the Mechanism of Intermediate complexity for Modelling Iron (MIMI v1.0). *Geoscientific Model Development*, *12*(9), 3835–3862. doi: 10.5194/gmd-12-3835-2019

Henson, S. A., Sanders, R., & Madsen, E. (2012). Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochemical Cycles*, *26*(1). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GB004099 doi: https://doi.org/10.1029/2011GB004099

Henson, S. A., Sanders, R., Madsen, E., Morris, P. J., Le Moigne, F., & Quartly, G. D. (2011). A reduced estimate of the strength of the ocean's biological carbon pump. *Geophysical Research Letters*, *38*(4), 10–14. doi: 10.1029/2011GL046735

Honjo, S., Manganini, S. J., Krishfield, R. A., & Francois, R. (2008). Particulate organic carbon fluxes to the ocean interior and factors controlling the biological pump: A synthesis of global sediment trap programs since 1983. *Progress in Oceanography*, *76*(3), 217-285. Retrieved from https://www.sciencedirect.com/science/article/pii/S0079661108000025 doi: https://doi.org/10.1016/j.pocean.2007.11.003

Huete-Ortega, M., Rodríguez-Ramos, T., López-Sandoval, D., Cermeño, P., Blanco, J. M., Palomino, R. L., ... Marañón, E. (2014). Distinct patterns in the size-scaling of abundance and metabolism in coastal and open-ocean phytoplankton communities. *Marine Ecology Progress Series*, *515*, 61–71.

Jackson, G. A. (1989). Simulation of bacterial attraction and adhesion to falling particles in an aquatic environment. *Limnology and Oceanography*, *34*(3), 514–530.

Jackson, G. A. (1990). A model of the formation of marine algal flocs by physical coagulation processes. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(8), 1197–1211.

Johnson, G. C., Schmidtko, S., & Lyman, J. M. (2012). Relative contributions of temperature and salinity to seasonal mixed layer density changes and horizontal density gradients. *Journal of Geophysical Research: Oceans*, *117*(4). doi: 10.1029/2011JC007651

Jonasz, M., & Fournier, G. (1996). Approximation of the size distribution of marine particles by a sum of log-normal functions. *Limnology and Oceanography*, *41*(4), 744–754.

Jouandet, M. P., Jackson, G. A., Carlotti, F., Picheral, M., Stemmann, L., & Blain, S. (2014). Rapid formation of large aggregates during the spring bloom of Kerguelen Island: Observations and model comparisons. *Biogeosciences*, *11*(16), 4393–4406. doi: 10.5194/bg-11-4393-2014

Karl, D., Knauer, G., Martin, J., & Ward, B. (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature*, *309*(5963), 54–56.

Karl, D., Knauer, G. A., & Martin, J. H. (1988). Downward flux of particulate organic matter in the ocean: a particle decomposition paradox. *Nature*, *332*(6163), 438–441.

Karp-Boss, L., Azevedo, L., & Boss, E. (2007). Lisst-100 measurements of phytoplankton size distribution: Evaluation of the effects of cell shape. *Limnology and Oceanography: Methods*, *5*(11), 396–406.

Kiko, R., Biastoch, A., Brandt, P., Cravatte, S., Hauss, H., Hummels, R., ... Stemmann, L. (2017). Biological and physical influences on marine snowfall at the equator. *Nature Geoscience*, *10*(11), 852–858. doi: 10.1038/NGEO3042

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., ... Stemmann, L. (2021). The global marine particle size distribution dataset obtained with the underwater vision profiler 5 - version 1. *PANGAEA*.

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., ... Stemmann, L. (2022). A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5. *Earth System Science Data*, *14*(9), 4315–4337. Retrieved from `https://essd.copernicus.org/articles/14/4315/2022/` doi: 10.5194/essd-14-4315-2022

Kiørboe, T., Andersen, K., & Dam, H. (1990). Coagulation efficiency and aggregate formation in marine phytoplankton. *Marine Biology*, *107*(2), 235–245.

Kiørboe, T., Balxter, J., & Southward, A. (1993). Turbulence, phytoplankton cell size, and the structure of pelagic food webs. *Advances in Marine Biology*, *29*, 1–72.

Kiørboe, T., Grossart, H.-P., Ploug, H., & Tang, K. (2002). Mechanisms and rates of bacterial colonization of sinking aggregates. *Applied and Environmental Microbiology*, *68*(8), 3996–4006.

Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J., ... Uitz, J. (2017). Inter-comparison of phytoplankton functional type phenology metrics derived from ocean color algorithms and earth system models. *Remote Sensing of Environment*, *190*, 162-177. Retrieved from `https://www.sciencedirect.com/science/article/pii/S003442571630459X` doi: https://doi.org/10.1016/j.rse.2016.11.014

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research: Oceans*, *114*(9), 1–22. doi: 10.1029/2009JC005303

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2010a). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeosciences*, *7*(10), 3239–3257. Retrieved from `https://bg.copernicus.org/articles/7/3239/2010/` doi: 10.5194/bg-7-3239-2010

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2010b). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. *Biogeosciences*, *7*(10), 3239–3257. Retrieved from `https://bg.copernicus.org/articles/7/3239/2010/` doi: 10.5194/bg-7-3239-2010

Kriest, I. (2002). Different parameterizations of marine snow in a 1D-model and their influence on representation of marine snow, nitrogen budget and sedimentation. *Deep-Sea Research Part I: Oceanographic Research Papers*, *49*(12), 2133–2162. doi: 10.1016/S0967-0637(02)00127-9

Kriest, I., & Evans, G. T. (1999). Representing phytoplankton aggregates in biogeochemical models. *Deep Sea Research Part I: Oceanographic Research Papers*, *46*(11), 1841–1859.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., ... others (2020). Twenty-first century ocean warming, acidification, deoxygenation,

and upper-ocean nutrient and primary production decline from cmip6 model projections. *Biogeosciences*, *17*(13), 3439–3470.

Kwon, E. Y., Primeau, F., & Sarmiento, J. L. (2009). The impact of remineralization depth on the air–sea carbon balance. *Nature Geoscience*, *2*(9), 630–635.

Laws, E. A., D'Sa, E., & Naik, P. (2011). Simple equations to estimate ratios of new or export production to total production from satellite-derived estimates of sea surface temperature and primary production. *Limnology and Oceanography: Methods*, *9*(DECEMBER), 593–601. doi: 10.4319/lom.2011.9.593

Le Gland, G., Aumont, O., & Mémery, L. (2019). An estimate of thorium 234 partition coefficients through global inverse modeling. *Journal of Geophysical Research: Oceans*, *124*(6), 3575-3606. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JC014668` doi: https://doi.org/10.1029/2018JC014668

Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., ... Wu, G. (2021). Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. *Remote Sensing of Environment*, *256*(January), 112316. Retrieved from `https://doi.org/10.1016/j.rse.2021.112316` doi: 10.1016/j.rse.2021.112316

Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., ... Smolyar, I. V. (2019). World Ocean Atlas 2018, Volume 1: Temperature. A. Mishonov, Technical Editor. *NOAA Atlas NESDIS*, *1*(81), 52pp.

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., ... others (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, *6*, 196.

Longhurst, A., Bedo, A., Harrison, W., Head, E., & Sameoto, D. (1990). Vertical flux of respiratory carbon by oceanic diel migrant biota. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(4), 685–694.

Lowry, K. E., Pickart, R. S., Selz, V., Mills, M. M., Pacini, A., Lewis, K. M., ... others (2018). Under-ice phytoplankton blooms inhibited by spring convective mixing in refreezing leads. *Journal of Geophysical Research: Oceans*, *123*(1), 90–109.

Marsay, C. M., Sanders, R. J., Henson, S. A., Pabortsava, K., Achterberg, E. P., & Lampitt, R. S. (2015). Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1089-1094. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1415311112` doi: 10.1073/pnas.1415311112

Martin, J. H., Knauer, G. A., Karl, D. M., & Broenkow, W. W. (1987). VERTEX: carbon cycling in the northeast Pacific. *Deep Sea Research Part A, Oceanographic Research Papers*, *34*(2), 267–285. doi: 10.1016/0198-0149(87)90086-0

McDonnell, A. M., & Buesseler, K. O. (2010). Variability in the average sinking velocity of marine particles. *Limnology and Oceanography*, *55*(5), 2085–2096.

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 Environment*, *111*(1), 69–88. doi: 10.1016/j.rse.2007.03.012

Moscoso, J. E., Bianchi, D., & Stewart, A. L. (2022). Controls and characteristics of biomass quantization in size-structured planktonic ecosystem models. *Ecological Modelling*, *468*, 109907. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0304380022000321` doi: https://doi.org/10.1016/j.ecolmodel.2022.109907

Mouw, C. B., Barnett, A., McKinley, G. A., Gloege, L., & Pilcher, D. (2016). Global ocean particulate organic carbon flux merged with satellite parameters. *Earth System Science Data*, *8*(2), 531–541.

Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J., Bricaud, A., ... others (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Frontiers in Marine Science*, *4*, 41.

Myriokefalitakis, S., Ito, A., Kanakidou, M., Nenes, A., Krol, M. C., Mahowald, N. M., ... Duce, R. A. (2018). Reviews and syntheses: The GESAMP atmospheric iron deposition model intercomparison study. *Biogeosciences*, *15*(21), 6659–6684. doi: 10.5194/bg-15-6659-2018

NASA G.S.F.C . (2014). *Modis-aqua ocean color data.* NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. doi: dx.doi.org/10.5067/AQUA/MODIS_OC.2014.0

N.G.D.C. (2006). *2-minute gridded global relief data (etopo2) v2.* National Geophysical Data Center, NOAA. (accessed: 11-13-2019) doi: 10.7289/V5J1012Q

Nguyen, T. T., Zakem, E. J., Ebrahimi, A., Schwartzman, J., Caglar, T., Amarnath, K., ... others (2022). Microbes contribute to setting the ocean carbon flux by altering the fate of sinking particulates. *Nature Communications*, *13*(1), 1–9.

Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC bioinformatics*, *11*(1), 1–13.

Nishioka, J., Obata, H., Ogawa, H., Ono, K., Yamashita, Y., Lee, K., ... Yasuda, I. (2020). Subpolar marginal seas fuel the North Pacific through the intermediate water at the termination of the global ocean circulation. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(23), 12665–12673. doi: 10.1073/pnas.2000658117

Nowicki, M., DeVries, T., & Siegel, D. A. (2022). Quantifying the carbon export and sequestration pathways of the ocean's biological carbon pump. *Global Biogeochemical Cycles*, *36*(3), e2021GB007083. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GB007083` (e2021GB007083 2021GB007083) doi: https://doi.org/10.1029/2021GB007083

Ohnemus, D. C., Torrie, R., & Twining, B. S. (2019). Exposing the distributions and elemental associations of scavenged particulate phases in the ocean using basin-scale multi-element data sets. *Global Biogeochemical Cycles*, *33*(6), 725–748.

Omand, M. M., D'Asaro, E. A., Lee, C. M., Perry, M. J., Briggs, N., Cetinić, I., & Mahadevan, A. (2015). Eddy-driven subduction exports particulate organic carbon from the spring bloom. *Science*, *348*(6231), 222–225. doi: 10.1126/science.1260062

Omand, M. M., D'Asaro, E. A., Lee, C. M., Perry, M. J., Briggs, N., Cetinić, I., & Mahadevan, A. (2015). Eddy-driven subduction exports particulate organic carbon from the spring bloom. *Science*, *348*(6231), 222-225. Retrieved from `https://www.science.org/doi/abs/10.1126/science.1260062` doi: 10.1126/science.1260062

Omand, M. M., Govindarajan, R., He, J., & Mahadevan, A. (2020). Sinking flux of particulate organic matter in the oceans: Sensitivity to particle characteristics. *Scientific reports*, *10*(1), 1–16.

Organelli, E., Dall'Olmo, G., Brewin, R. J., Nencioli, F., & Tarran, G. A. (2020). Drivers of spectral optical scattering by particles in the upper 500 m of the atlantic ocean. *Optics Express*, *28*(23), 34147–34166.

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*(9), 4171–4179. doi: 10.1029/2017GL076498

Pavia, F. J., Anderson, R. F., Lam, P. J., Cael, B. B., Vivancos, S. M., Fleisher, M. Q., ... Edwards, R. L. (2019). Shallow particulate organic carbon regeneration in the South Pacific Ocean. *Proceedings of the National Academy of Sciences*, *116*(20), 201901863. doi: 10.1073/pnas.1901863116

Picheral, M., Catalano, C., Brousseau, D., Claustre, H., Coppola, L., Leymarie, E., ... others (2022). The underwater vision profiler 6: an imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms. *Limnology and Oceanography: Methods*, *20*(2), 115–129.

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of

particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(SEPT), 462–473. doi: 10.4319/lom.2010.8.462

Pinti, J., DeVries, T., Norin, T., Serra-Pompei, C., Proud, R., Siegel, D. A., ... Visser, A. W. (2022). Model estimates of metazoans' contributions to the biological carbon pump. *EGUsphere*, *2022*, 1–20. Retrieved from `https://egusphere.copernicus.org/preprints/egusphere-2022-1227/` doi: 10.5194/egusphere-2022-1227

Quay, P., Emerson, S., & Palevsky, H. (2020). Regional Pattern of the Ocean's Biological Pump Based on Geochemical Observations. *Geophysical Research Letters*, *47*(14), 1–10. doi: 10.1029/2020GL088098

Reynolds, R., Stramski, D., Wright, V., & Woźniak, S. (2010). Measurements and characterization of particle size distributions in coastal waters. *Journal of Geophysical Research: Oceans*, *115*(C8).

Reynolds, R. A., & Stramski, D. (2021a). Variability in oceanic particle size distributions and estimation of size class contributions using a non-parametric approach. *Journal of Geophysical Research: Oceans*, *126*(12), e2021JC017946.

Reynolds, R. A., & Stramski, D. (2021b). Variability in oceanic particle size distributions and estimation of size class contributions using a non-parametric approach. *Journal of Geophysical Research: Oceans*, *126*(12), e2021JC017946. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JC017946` (e2021JC017946 2021JC017946) doi: https://doi.org/10.1029/2021JC017946

Reynolds, R. A., Stramski, D., & Neukermans, G. (2016). Optical backscattering by particles in arctic seawater and relationships to particle mass concentration, size distribution, and bulk composition. *Limnology and Oceanography*, *61*(5), 1869–1890.

Reynolds, R. A., Stramski, D., Wright, V. M., & Woźniak, S. B. (2010). Measurements and characterization of particle size distributions in coastal waters. *Journal of Geophysical*

Research: Oceans, *115*(C8). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JC005930 doi: https://doi.org/10.1029/2009JC005930

Richardson, T. L. (2019). Mechanisms and pathways of small-phytoplankton export from the surface ocean. *Ann. Rev. Mar. Sci*, *11*(1), 57–74.

Richardson, T. L., & Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science*, *315*(5813), 838-840. Retrieved from https://www.science.org/doi/abs/10.1126/science.1133471 doi: 10.1126/science.1133471

Riley, J. S., Sanders, R., Marsay, C., Le Moigne, F. A., Achterberg, E. P., & Poulton, A. J. (2012). The relative contribution of fast and slow sinking particles to ocean carbon export. *Global Biogeochemical Cycles*, *26*(1), 1–10. doi: 10.1029/2011GB004085

Robinson, C., Steinberg, D. K., Anderson, T. R., Aristegui, J., Carlson, C. A., Frost, J. R., ... others (2010). Mesopelagic zone ecology and biogeochemistry–a synthesis. *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(16), 1504–1518.

Roullier, F., Berline, L., Guidi, L., Durrieu De Madron, X., Picheral, M., Sciandra, A., ... Stemmann, L. (2014). Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. *Biogeosciences*, *11*(16), 4541–4557. Retrieved from https://bg.copernicus.org/articles/11/4541/2014/ doi: 10.5194/bg-11-4541-2014

Roy, S., Sathyendranath, S., Bouman, H., & Platt, T. (2013). The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data. *Remote Sensing of Environment*, *139*, 185–197. Retrieved from http://dx.doi.org/10.1016/j.rse.2013.08.004 doi: 10.1016/j.rse.2013.08.004

Runyan, H., Reynolds, R. A., & Stramski, D. (2020). Evaluation of particle size distribution metrics to estimate the relative contributions of different size fractions based on measurements in arctic waters. *Journal of Geophysical Research: Oceans*, *125*(6), e2020JC016218.

Sarmiento, J. L., & Gruber, N. (2006a). *Ocean biogeochemical dynamics*. Princeton University Press. Retrieved from http://www.jstor.org/stable/j.ctt3fgxqx

Sarmiento, J. L., & Gruber, N. (2006b). Ocean biogeochemical dynamics. In *Ocean biogeochemical dynamics.* Princeton University Press.

Schartau, M., Landry, M. R., & Armstrong, R. A. (2010). Density estimation of plankton size spectra: a reanalysis of ironex ii data. *Journal of plankton research*, *32*(8), 1167–1184.

Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., ... others (2020). Tracking improvement in simulated marine biogeochemistry between cmip5 and cmip6. *Current Climate Change Reports*, 1–25.

Sheldon, R., Prakash, A., & Sutcliffe Jr., W. H. (1972). The size distribution of particles in the ocean1. *Limnology and Oceanography*, *17*(3), 327-340. Retrieved from https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.1972.17.3.0327 doi: https://doi.org/10.4319/lo.1972.17.3.0327

Sheldon, R., Sutcliffe Jr, W., & Paranjape, M. (1977). Structure of pelagic food chain and relationship between plankton and fish production. *Journal of the Fisheries Board of Canada*, *34*(12), 2344–2353.

Siegel, D. A., Buesseler, K. O., Doney, S. C., Sailley, S. F., Behrenfeld, M. J., & Boyd, P. W. (2014). Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochemical Cycles*, *28*(3), 181–196. doi: 10.1002/2013GB004743

Siegel, D. A., Cetinić, I., Graff, J. R., Lee, C. M., Nelson, N., Perry, M. J., ... others (2021). An operational overview of the export processes in the ocean from remote sensing (exports) northeast pacific field deployment. *Elem Sci Anth*, *9*(1), 00107.

Siegel, D. A., DeVries, T., Cetinić, I., & Bisson, K. M. (2022). Quantifying the ocean's biological pump and its carbon cycle impacts on global scales. *Annual Review of Marine Science*, *15*.

Siegel, D. A., Doney, S. C., & Yoder, J. A. (2002). The North Atlantic spring phytoplankton bloom and Sverdrup's critical depth hypothesis. *Science*, *296*(5568), 730–733. doi: 10 .1126/science.1069174

Silsbe, G. M., Behrenfeld, M. J., Halsey, K. H., Milligan, A. J., & Westberry, T. K. (2016). The CAFE model: A net production model for global ocean phytoplankton. *Global Biogeochemical Cycles*, *30*(12), 1756–1777. doi: 10.1002/2016GB005521

Smayda, T. J. (1970). The Suspension and sinking of phytoplankton in the sea. *Oceanography Marine Biology, Annual Review*, *8*, 353–414.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., & Michaels, A. F. (2000). Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the sargasso sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(1), 137–158.

Steinberg, D. K., Van Mooy, B. A., Buesseler, K. O., Boyd, P. W., Kobari, T., & Karl, D. M. (2008). Bacterial vs. zooplankton control of sinking particle flux in the ocean's twilight zone. *Limnology and Oceanography*, *53*(4), 1327–1338.

Stemmann, L., & Boss, E. (2012). Plankton and Particle Size and Packaging: From Determining Optical Properties to Driving the Biological Pump. *Annual Review of Marine Science*, *4*(1), 263–290. Retrieved from `http://www.annualreviews.org/doi/10.1146/annurev-marine-120710-100853` doi: 10.1146/annurev-marine-120710-100853

Stemmann, L., Eloire, D., Sciandra, A., Jackson, G., Guidi, L., Picheral, M., & Gorsky, G. (2008). Volume distribution for particles between 3.5 to 2000 μm in the upper 200 m region of the South Pacific Gyre. *Biogeosciences Discussions*, *5*, 299–310. doi: 10.5194/bgd-4-3377-2007

Stemmann, L., Gorsky, G., Marty, J.-C., Picheral, M., & Miquel, J.-C. (2002). Four-year study of large-particle vertical distribution (0–1000m) in the nw mediterranean in relation to hydrology, phytoplankton, and vertical flux. *Deep Sea Research Part*

II: Topical Studies in Oceanography, *49*(11), 2143-2162. Retrieved from https://www.sciencedirect.com/science/article/pii/S0967064502000322 (Studies at the DYFAMED (France JGOFS) Time-Series Station, N.W. M editerranean Sea) doi: https://doi.org/10.1016/S0967-0645(02)00032-2

Stemmann, L., Jackson, G. A., & Ianson, D. (2004a). A vertical model of particle size distributions and fluxes in the midwater column that includes biological and physical processes - Part I: Model formulation. *Deep-Sea Research Part I: Oceanographic Research Papers*, *51*(7), 865–884. doi: 10.1016/j.dsr.2004.03.001

Stemmann, L., Jackson, G. A., & Ianson, D. (2004b). A vertical model of particle size distributions and fluxes in the midwater column that includes biological and physical processes—part i: model formulation. *Deep Sea Research Part I: Oceanographic Research Papers*, *51*(7), 865–884.

Stemmann, L., Youngbluth, M., Robert, K., Hosia, A., Picheral, M., Paterson, H., ... Gorsky, G. (2008). Global zoogeography of fragile macrozooplankton in the upper 100-1000 m inferred from the underwater video profiler. *ICES Journal of Marine Science*, *65*(3), 433–442. doi: 10.1093/icesjms/fsn010

Stukel, M. R., Aluwihare, L. I., Barbeau, K. A., Chekalyuk, A. M., Goericke, R., Miller, A. J., ... Landry, M. R. (2017). Mesoscale ocean fronts enhance carbon export due to gravitational sinking and subduction. *Proceedings of the National Academy of Sciences*, *114*(6), 1252–1257. Retrieved from http://www.pnas.org/lookup/doi/10.1073/pnas.1609435114 doi: 10.1073/pnas.1609435114

Stukel, M. R., Song, H., Goericke, R., & Miller, A. J. (2018). The role of subduction and gravitational sinking in particle export, carbon sequestration, and the remineralization length scale in the california current ecosystem. *Limnology and Oceanography*, *63*(1), 363-383. Retrieved from https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.1002/lno.10636 doi: https://doi.org/10.1002/lno.10636

Sverdrup, H. (1953). On conditions for the vernal blooming of phytoplankton. *J. Cons. Int. Explor. Mer*, *18*(3), 287–295.

Trudnowska, E., Lacour, L., Ardyna, M., Rogge, A., Irisson, J. O., Waite, A. M., ... Stemmann, L. (2021). Marine snow morphology illuminates the evolution of phytoplankton blooms and determines their subsequent vertical export. *Nature communications*, *12*(1), 1–13.

Trudnowska, E., Stemmann, L., Błachowiak-Samołyk, K., & Kwasniewski, S. (2020). Taxonomic and size structures of zooplankton communities in the fjords along the atlantic water passage to the arctic. *Journal of Marine Systems*, *204*, 103306. Retrieved from https://www.sciencedirect.com/science/article/pii/S0924796320300026 doi: https://doi.org/10.1016/j.jmarsys.2020.103306

Turekian, K. K. (1977). The fate of metals in the oceans. *Geochimica et Cosmochimica Acta*, *41*(8), 1139–1144.

Turner, J. S., Pretty, J. L., & McDonnell, A. M. (2017). Marine particles in the Gulf of Alaska shelf system: Spatial patterns and size distributions from in situ optics. *Continental Shelf Research*, *145*(July), 13–20. Retrieved from http://dx.doi.org/10.1016/j.csr.2017.07.002 doi: 10.1016/j.csr.2017.07.002

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205–248.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric co2 changes. In *The carbon cycle and atmospheric co2: Natural variations archean to present* (p. 99-110). American Geophysical Union (AGU). Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/GM032p0099 doi: https://doi.org/10.1029/GM032p0099

Ward, B. A., Dutkiewicz, S., & Follows, M. J. (2014). Modelling spatial and temporal patterns in size-structured marine plankton communities: Top-down and bottom-up controls. *Journal of Plankton Research*, *36*(1), 31–47. doi: 10.1093/plankt/fbt097

Weber, T., & Bianchi, D. (2020). Efficient particle transfer to depth in oxygen minimum zones of the pacific and indian oceans. *Frontiers in Earth Science*, *8*, 376.

Weber, T., Cram, J. A., Leung, S. W., DeVries, T., & Deutsch, C. (2016). Deep ocean nutrients imply large latitudinal variation in particle transfer efficiency. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(31), 8606–8611. doi: 10.1073/pnas.1604414113

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2), 1–18. doi: 10.1029/2007GB003078

Xiang, Y., Lam, P. J., Burd, A. B., & Hayes, C. T. (2022). Estimating mass flux from size-fractionated filtered particles: Insights into controls on sinking velocities and mass fluxes in recent u.s. geotraces cruises. *Global Biogeochemical Cycles*, *36*(4), e2021GB007292. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GB007292` (e2021GB007292 2021GB007292) doi: https://doi.org/10.1029/2021GB007292

Yang, S., Chang, B. X., Warner, M. J., Weber, T. S., Bourbonnais, A. M., Santoro, A. E., ... Bianchi, D. (2020). Global reconstruction reduces the uncertainty of oceanic nitrous oxide emissions and reveals a vigorous seasonal cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(22). doi: 10.1073/pnas.1921914117

Zhang, G., & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, *39*(1), 151-160. Retrieved from `https://doi.org/10.1080/02664763.2011.578621` doi: 10.1080/02664763.2011.578621

Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Antonov, J. I., Locarnini, R. A., ... Smolyar, I. V. (2019). World Ocean Atlas 2018, Volume 2: Salinity. *NOAA Atlas NESDIS*, *2*(82), 50.