A new estimate of Global Ocean Carbon Flux from *In Situ* Optical Observations and Supervised Learning.

D.J. Clements[1], S. Yang[1], T. Weber[2], A.M.P. McDonnell[3], R.Kiko[4], L.Stemmann[4], D.Bianchi[1]

[1]Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA.
[2]Department of Earth and Environmental Sciences, University of Rochester,Rochester, New York, USA
[3]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, Alaska 99775-7220, USA.
[4]Sorbonne Université, CNRS, UMR 7093, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, Villefranche-sur-Mer, France.

Key Points:

- We use optical observations of particle size distribution to reconstruct global particulate carbon fluxes.
- We quantify the importance of particle size distribution slope and biovolume on the reconstructed fluxes.
- Fluxes estimated at two depths show that more carbon is sequestered from the wintertime mixed layer than the euphotic zone.

Corresponding author: D.J Clements, dclements@atmos.ucla.edu

Corresponding author: D. Bianchi, dbianchi@atmos.ucla.edu

Abstract

Export of sinking particles from the surface ocean is critical for carbon sequestration and for providing energy to the deep-ocean biosphere. The magnitude and spatial patterns of this flux have been estimated in the past by satellite-based algorithms and ocean biogeochemical models; however, these estimates remain uncertain. Here, we present a novel analysis of a global compilation of *in situ* ocean particle size spectra from Underwater Vision Profiler 5 (UVP5) measurements, from which we determine particulate carbon fluxes. Using a machine learning algorithm, we extrapolate sparse observations of particle abundance by size to the global ocean from oceanographic variables that are more commonly observed. We reconstruct global maps of particle size distribution parameters for large sinking particles (80 µm to 2.6 cm), and combine them with empirical relationships to calculate the sinking carbon flux from the euphotic zone and the wintertime mixed layer depth. Our flux reconstructions are comparable to other estimates, but suggest a less variable seasonal cycle in the tropical ocean, and a more continuous export in the Southern Ocean than previously thought. Because our estimates are not bounded by a specific depth horizon, we reconstruct export at multiple depths, and find that export from the wintertime mixed layer globally exceeds that from the euphotic zone. Our estimates provide a baseline for more accurate understanding of particle cycles in the ocean, and open the way to fully three-dimensional global reconstructions of particle size spectra and fluxes in the ocean, supported by the growing database of UVP5 observations.

1 Introduction

In the ocean, primary production and other complex biogeochemical processes interact to form the ocean's biological pump. Aggregation of particulate organic matter into particles denser than seawater leads to gravitational settling (Alldredge & Gotschalk, 1988), eventually storing inorganic carbon and nutrients in the deep ocean for timescales of decades to centuries (Boyd et al., 1999, 2019). The export of particulate organic carbon provides energy to the deep ocean ecosystem (Siegel et al., 2014), influences atmospheric carbon dioxide and climate (Kwon et al., 2009; Palevsky & Doney, 2018), and indirectly affects on the ocean's microbiome (Karl et al., 1984; Bianchi et al., 2018). Several studies have estimated this global particle flux from the euphotic zone, resulting in substantially variable estimates ranging from 3 to 10 PgC/y (Henson et al., 2011; Siegel et al., 2014; DeVries & Weber, 2017; Dunne et al., 2007).

Current reconstructions of the sinking particle flux tend to vary in the total global export, depending on the methods used (Quay et al., 2020). Biogeochemical models yield a global export of 4-6 PgC/y when tuned to match particle observations, but could reach up to 10 PgC/y when tuned to match in situ profiles of nutrients and other biogeochemical tracers (Siegel et al., 2014; DeVries et al., 2017). Models that use satellite inputs and empirically derived export ratios tend to result in a larger flux (Dunne et al., 2007; Laws et al., 2011; Guidi et al., 2015), near 10 PgC/y. This is similar to annual net community production estimates at the base of the mixed layer, which include additional export of dissolved organic carbon (Emerson, 2013; Quay et al., 2020).

Although the globally integrated particle flux is similar when comparing model, geochemical, and satellite-based estimates, the regional patterns predicted by these methods differ substantially. Differences in regional flux estimates 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. Based on *in situ* geochemical observations, Quay et al. (2020) suggest a weaker meridional variability in export flux than other estimates, stressing the need for expanding and combining observational approaches and models to fully constrain particle export.

Recent studies have highlighted the importance of standardized methods and metrics used to quantify particle fluxes (Buesseler et al., 2020). In particular, the depth horizon of export has been identified as a leading cause of diverging estimates (Palevsky & Doney, 2018). Two choices of export horizon have been commonly adopted: the base of the euphotic zone, either as a variable depth or global average (Buesseler & Boyd, 2009; Siegel et al., 2016, 2014; Bisson et al., 2018; Dunne et al., 2007; DeVries et al., 2017; Henson et al., 2011), and the mixed layer depth, both as seasonally varying and maximum depth (Emerson, 2013; Quay et al., 2020). These choices underlie different interpretations of export fluxes: euphotic export takes an ecosystem view, while mixed layer export takes a carbon storage view.

Gravitational settling is thought to be the primary export mechanism globally, contributing to about 60% of the total carbon export, and about half of the carbon storage in the deep ocean (Boyd et al., 2019). Using both a euphotic viewpoint, and considering only gravitational settling, particle flux estimates have begun to converge on a value of 5-6 PgC/y (Palevsky & Doney, 2018; Boyd et al., 2019).

Advances in ocean optical observations have begun to provide a three-dimensional view into the life of particulate matter in the ocean (Stemmann & Boss, 2012; Kiko et al., 2017; Guidi et al., 2009). The Underwater Vision Profiler 5 (UVP5) is an optical particle counter which provides the *in situ* particle abundance for large 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, and quantifies the particle abundances at high frequency as it is lowered in the water column. Vertical profiles of particle size distribution (PSD) from the UVP5 are commonly taken at up to 20 images per second with downward speeds of $1m/s$, with observations as deep at 6 km (Picheral et al., 2010). Since 2008, UVP5s have been routinely deployed on ocean expeditions, resulting in over 9,000 profiles to date, with observations from all ocean basins.

Although the UVP5 cannot directly determine carbon flux, because particle sinking speed and carbon content are not measured, empirical relationships have been used to define these as a function of size, making flux estimates possible (e.g., as compiled in Kriest (2002) and Stemmann et al. (2004)). The vertical resolution of the UVP5, coupled with these empirical relationships, enables a unique high-resolution view into the three-dimensional ocean particle flux. Observations from UVP5 have been used to quantify particulate flux to the mesopelagic ocean on a regional basis (Guidi et al., 2008, 2009; Kiko et al., 2017), and a smaller dataset has been used to reconstruct global fluxes by large-scale biomes (Guidi et al., 2015). However, the expanded dataset has not yet been used to quantify fluxes from the surface ocean yet.

In this study, we reconstruct global particle carbon export by training a supervised machine learning algorithm to extrapolate PSD from a rapidly growing dataset of UVP5 observations and well-sampled oceanographic variables. We combine these estimates with *in situ* sediment trap and thorium-derived particle flux observations to better constrain empirical relationships between particle size, sinking speed and carbon content, producing robust estimates of regional and seasonal flux variability. By comparing patterns in PSD and flux with observations of environmental drivers, we further gain insight into the mechanisms responsible for particle export and its variability. Finally, we exploit the high vertical resolution of UVP5 measurements to estimate particle fluxes at both the climatological euphotic zone and the maximum mixed layer depth, revealing the importance of the export horizon for this quantity.

The rest of the paper is organized as follows. Section 2 describes the machine-learning approach used to globally extrapolate PSD and reconstruct particle fluxes. Section 3 presents the reconstructions of particle distributions and export fluxes, and compares our results to previous studies at global and regional scales. Section 4 summarizes the main find-

120 ings, discussing the uncertainties and caveats inherent to our approach, and future di-
121 rections.

122 2 Methods

123 The flux of particulate carbon (ϕ) at any given depth can be expressed as a func-
124 tion of three size-dependent quantities: PSD ($n(s)$, $\frac{\#}{L \cdot cm}$), sinking speed ($w(s)$, $\frac{m}{s}$), and
125 particle carbon content ($c(s)$, mgC), according to the following equation (Stemmann &
126 Boss, 2012):

$$\phi = \int_{s_{min}}^{s_{max}} n(s) \cdot w(s) \cdot c(s)\, ds, \tag{1}$$

127

128 Here, s indicates the particle equivalent spherical diameter, or size (cm), s_{min} and
129 s_{max} are respectively the minimum and maximum size of particles considered for export.
130 We further assume that the quantities in Equation 1 can be approximated by power laws
131 that depend on particle size, each characterized by an intercept (the size-independent
132 coefficient) and a slope (the exponent for size-dependence) (Stemmann & Boss, 2012):

$$n(s) = n_0 \cdot s^{-\beta} \tag{2}$$

133

$$w(s) = w_0 \cdot s^{\eta} \tag{3}$$

134

$$c(s) = c_0 \cdot s^{\zeta}, \tag{4}$$

135

136 Thus, by using Equations 2-4, the total particle flux can be expressed as:

$$\phi = \int_{s_{min}}^{s_{max}} n_0 \cdot w_0 \cdot c_0 \cdot s^{-\beta+\eta+\zeta}\, ds = \int_{s_{min}}^{s_{max}} n_0 \cdot m_0 \cdot s^{-\beta+\mu}\, ds \tag{5}$$

137

138 where we combined the intercepts and exponents of the sinking speed and carbon
139 content relationships by setting $m_0 = w_0 \cdot c_0$ and $\mu = \eta + \zeta$. We further approximate
140 m_0 and μ with globally constant values.

141 We use UVP5 observations to reconstruct PSDs (i.e., n_0 and β) at the chosen ex-
142 port horizon, by fitting Equation 2 to the observed particle abundance. We use a 20 me-
143 ter depth bin around the export horizon to average the observations and smooth out small-
144 scale noise and variability. We then extrapolate sparse UVP5 observations to a global
145 grid, by training a supervised learning algorithm to predict spatially-varying PSD from
146 well-sampled environmental predictors. To completely reconstruct fluxes based on Equa-
147 tion (1), we constrain the parameters of the combined sinking speed and carbon content
148 relationships, i.e., m_0 and μ, by optimizing predicted particle fluxes against *in situ* flux
149 estimates from sediment traps and thorium-uranium disequilibrium measurements.

150 We exploit the three-dimensional nature of UVP5 observations to perform these
151 calculation at two different export horizons: the base of the euphotic zone (here defined
152 by the 1% light level according to Morel et al. (2007)) and the annual maximum mixed
153 layer depth. The steps used to solve Equation (1) are illustrated in the workflow schematic
154 in Fig. 1, and are detailed in the following sections.

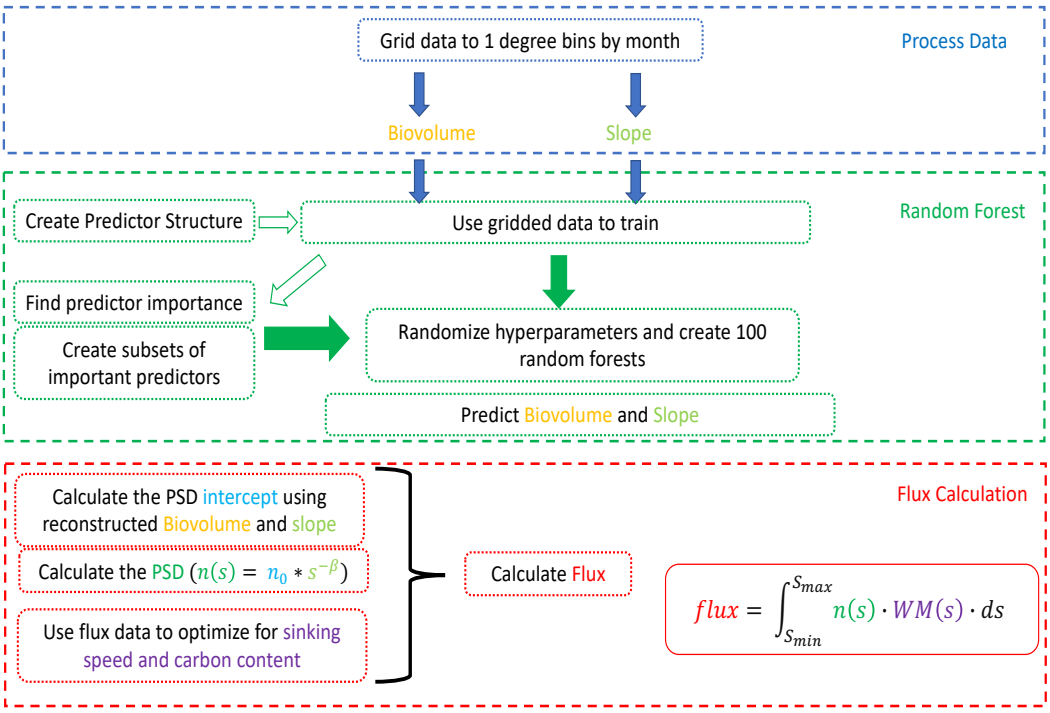

Figure 1. Schematic diagram illustrating the general workflow of processing UVP5 observations into a flux reconstruction. Observations are ensembled onto a normal 1 degree grid, with observation representing an average of a 20 meter vertical bin about the export horizon. All observations are the calculated values for the 105 μm to 5 mm size bins. Specific parameters are globally extrapolated using the Random Forest algorithm. These new global data are used to construct the flux, using equation 1.

2.1 Reconstructions of particle size spectra from UVP5 data

Observations from UVP5 provide particle counts for the 80 μm - 2.6 cm size range at any specific location and depth. Under the power law assumption, the two parameters n_0 and β are needed to fully capture the PSD (Equation (2)) (Brun-Cottan, 1971; Stemmann et al., 2004; Stemmann & Boss, 2012; Devries et al., 2014).

The slope β quantifies the relative abundance of large vs. small particles, while the intercept n_0 is a measure of the total abundance of particles for an arbitrary reference size. We determine the slope for each UVP5 measurement by fitting a linear least-squares regression through the log-transformed particle abundance and size. We determine the intercept by using the particle biovolume (BV), which can be directly derived from UVP5 observations, and applying the following definition:

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

By fixing the size range, we solve Equation 6 for the intercept as a function of slope and biovolume:

$$n_0 = \frac{6 \cdot BV}{\pi} \cdot \left(\frac{s_{max}^{4-\beta}}{4-\beta} - \frac{s_{min}^{4-\beta}}{4-\beta}\right)^{-1}. \tag{7}$$

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 of 105 μm to avoid a potential slight instrument bias in the lowest size classes. We constrain the maximum size to 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 et al., 2008; Stemmann & Boss, 2012). The sensitivity of our results to these choices is discussed in Section 2.2.

We coarsen the temporal and spatial resolution of the 6856 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). To this end, we combine profiles in a given grid cell and month together, thus reducing the noisy and episodic nature of particle observations. To reconstruct global PSDs, we calculate slope and biovolume for each grid cell, at a given export horizon, using the gridded dataset, and assume that these averages are representative of the climatological monthly PSD in each grid cell.

Although the gridded observations reduce data patchiness in well sampled regions, many grid cells only contain one observation. 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 correlation coefficient of less than 0.9, we remove the data point, as it likely does not follow a power law distribution. This quality control step removes less than 1% of data (Supplementary Information Fig. S1). The final processed UVP5 observation dataset contains 2,034 gridded observations at the export horizon, which together cover less than 10% of the ocean surface.

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 Random Forest (RF) algorithm to reconstruct 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 ensembled 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 extrapolation power outside of the training dataset. To mitigate the risk of overfitting, the RF does not use all data points for training. Instead, a bootstrapped sample of the data is selected for each tree in the forest. The degree of overfitting is determined by finding the error between the model and the data not used for training, i.e., the "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 "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 this RF (Supplementary Fig. S2). 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 results, we apply qualitative understanding of the predictors, combined with the recursive feature elimination, to determine if a predictor is should be included in the final regression.

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 our case, we use climatological monthly predictors at 1-degree spatial resolution. We include a variety of predictors that could be mechanistically related to particle production and export in the surface ocean, ranging from physical variables (e.g., temperature and salinity) to ecosystem-level quantities (e.g., primary production, euphotic zone depth). A list of all predictors used is shown in Table 1.

Some of these predictors are obtained from satellite products at high spatial and temporal resolution (e.g., surface chlorophyll and 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 spherical interpolation algorithm (D'Errico, 2016). 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). To process net primary production, we also calculate the critical depth, where light becomes too limiting to support photosynthesis, based on climatological chlorophyll concentration and incident shortwave radiation (Siegel et al., 2002). Net primary production is then set to zero at all points where the mixed layer depth exceeds the critical depth, before interpolating. We also include the standard deviation of the primary production, as a proxy for intermittency and sub-seasonal variability. Similarly, we restrict chlorophyll and net primary production based on climatological sea ice cover.

We use three-dimensional variables (e.g., temperature, nutrients) to generate two-dimensional surface predictors based on mixed layer averages. We also include predictors that quantify the change of variables with depth, by calculating the average of the variable from the base of the mixed layer to 100m below it. For surface-only variables (e.g., chlorophyll, net primary production) and nutrients we also include predictors that quantify change in time, calculated by applying a finite-difference time derivative, because change over time might be more indicative of export flux than the actual variable. We refer to these depth- and time-change variables as "variations" in Table 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 1). If a predictor is in the universal category, it is always included in all RF realizations. For all other categories, only one predictor is chosen for each realization, but if a predictor is chosen, all variations are included too.

Based on the categories listed in Table 1, we use a total of 29 predictors for each RF realization (Table 1). We generate 100 realizations, with variable hyper-parameters (the number of trees and their complexity) and randomly chosen predictors from each category, and take the ensemble average as the final product, with the inter-model spread representing the error. Generating an ensemble of 100 RFs, with varying hyper-parameters and predictors, reduces biases and overfitting, making the results robust with respect to parameter tuning, and the choice of different data products. Thus, our 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 considering 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.

2.2 Sinking Speed and Carbon Content

Particle sinking speed and carbon content have been empirically estimated using power law relationships analogous to Equations (3) and (4). Most of these studies measured a range of particles that does not wholly encompass the sizes detected by the UVP5. Also, these relationships are defined for specific particles types, which are not distinguished in the dataset used. Since estimates of total flux are sensitive to the sinking speed and

Table 1. Variables used to predict PSD parameters, their source and variations.The categories are organized based on their predictor type where the universal predictors are used in every random forest.

Category	Variable	Variations	Source
Universal			
	Topography		N.G.D.C (2006)
	Temperature below mixed layer	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	C.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 depth	Time Derivative	Johnson et al. (2012)
	Mixed Layer depth	Time Derivative	de Boyer Montégut et al. (2004)
Primary Production			
	Eppley VGPM	Time Derivative	Antoine and Morel (1996)
	VGPM	Time Derivative	Behrenfeld and Falkowski (1997)
	CBPM	Time Derivative	Westberry et al. (2008)
	CAFE	Time Derivative	Silsbe et al. (2016)
NPP Standard Deviation			
	Eppley VGPM		Antoine and Morel (1996)
	VGPM		Behrenfeld and Falkowski (1997)
	CBPM		Westberry et al. (2008)
Euphotic Zone Depth			
	VGPM		Morel et al. (2007)
	CBPM		Morel et al. (2007)
Iron			
	Soluble Iron Fraction	Time Derivative	Hamilton et al. (2019)
	Labile Iron Fraction	Time Derivative	Myriokefalitakis et al. (2018)

carbon content relationships, here 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 particle flux predictions (Equation 5) and co-located *in situ* carbon flux measurements (Bisson et al., 2018).

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 (5), selecting reasonable combination for the final estimate. Ultimately, when optimizing the sinking carbon parameters, the global export flux is insensitive to the size range; however the resulting empirical relationships are (see Section 3.2 and Supplementary Fig. S5). The insensitivity of the carbon flux to the size range of that flux indicates a compensatory effect between the sinking carbon parameters and the size range selected. Thus, choosing different size combinations would result in a similar total flux, although it may slightly alter spatial or temporal patterns.

Our final choice of size range is informed by average sinking speeds and carbon content previously reported (Kriest, 2002). We assume the slope and intercept (calculated from the biovolume; Equation 7) of the particle size distribution to be constant for sinking particulate matter, and expand the minimum size to include all sinking particle sizes. Based on this 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), a size that is likely rapidly remineralized, making its contribution to the sinking flux negligible (Riley et al., 2012). Similar to the observation limitations, we choose 5 mm as the maximum size of sinking particles.

2.3 Flux reconstruction and evaluation

Using the PSD reconstructions and the optimized sinking carbon parameters, we calculate particle export fluxes following Equation (5). We evaluate these reconstructions by comparing them to *in situ* flux observations and previous global reconstructions. Specifically, we compare total fluxes, meridional 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. We evaluate seasonal cycles by analyzing temporal correlations between reconstructions, and by introducing a seasonality index defined by the ratio between the seasonal range and annual mean flux in each region.

We first present results for fluxes estimated at the climatological euphotic zone depth, and then repeat the calculation at the maximum mixed layer depth. For the latter, we keep the same sinking speed and carbon content parameters as determined for the euphotic zone depth. Thus, the only methodological difference between the two estimates is the depth used to calculate the export flux.

3 Results

3.1 Particle Size Distribution reconstructions

Our global reconstructions capture most of the variability of UVP5-based PSD slope and biovolume data (Figs. 2 and 3), and robustly reproduce observations, with global average values of 0.6 ppm for biovolume ($r^2 = 0.91$) and 3.9 for slope ($r^2 = 0.86$). Observations that are not used in the training (out-of-bag) are also robustly predicted, with a RMSE of 2.1 ppm for biovolume ($r^2 = 0.74$) and 0.33 for slope ($r^2 = 0.68$). Some of the remaining uncertainty in our reconstructions can be attributed to the episodic nature of particle production and export. Our method operates under the assumption that

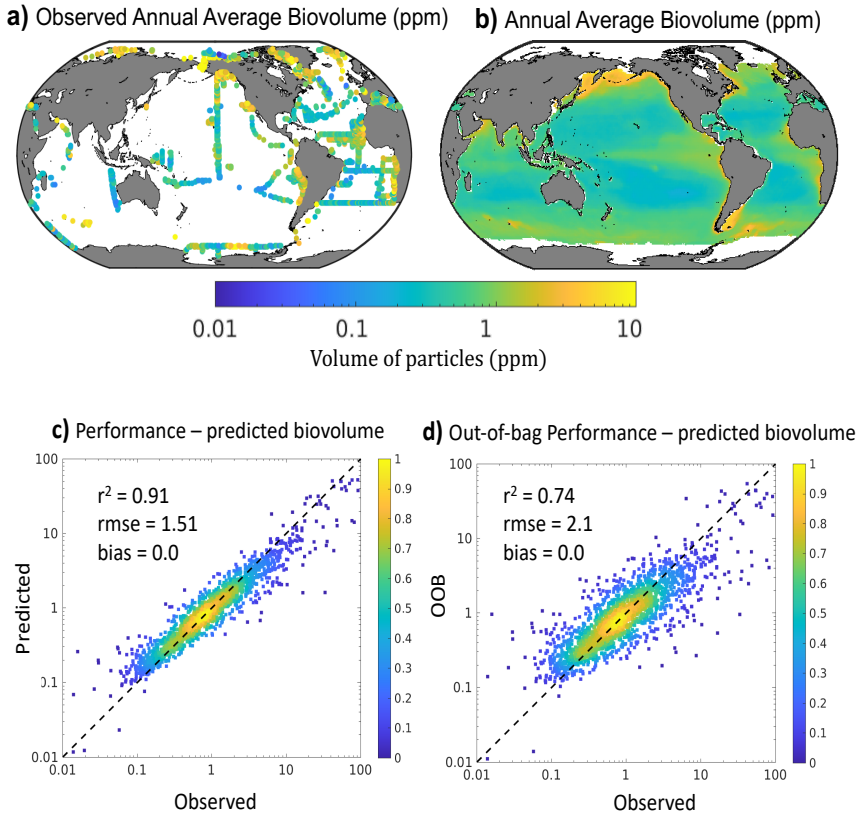

Figure 2. Observed and reconstructed Particle biovolume at the base of the euphotic zone. (A) Map of the observations of intercept, for locations with observations in multiple months the average is shown (B) Map of the biovolume reconstructions. (C) Performance of the RF reconstruction shown as density scatter plots of predicted vs. observed Biovolume (colors indicate the normalized density of grid points surrounding the given cell). (D) Same as B, but using out-of-bag (OOB) predictions, i.e., predictions vs. observations withheld from training. Annotations in B and C show the square of the correlation coefficient (r2), the RMSE and the global bias.

the input data (i.e., the UVP5 measurements) are monthly climatological averages, rather than snapshots. By ensembling these snapshot measurements into 2,034 monthly observational data points, we reduce part of the episodic nature of these observations; however some patchy behavior may still exist in the gridded data. Overall, the reconstructions show slight underestimates of extreme values (i.e., a reduced range), but negligible mean biases for both variables (Figs. 2 and 3).

We find high biovolume in productive regions such as high latitudes, coastal waters, and upwelling systems, and low biovolume in the oligotrophic suptropical gyres. These patterns mirror satellite-derived primary productivity and chlorophyll measurements (Supplementary Fig. S3), suggesting that phytoplankton and photosynthesis ultimately control the total abundance of particles in any given region. Reconstructions of PSD slope show a similar but negative correlation with primary production and chlorophyll, with smaller slopes (i.e., "flatter" PSD) in more productive regions, and larger slopes (i.e., "steeper" PSD) in oligotrophic waters.

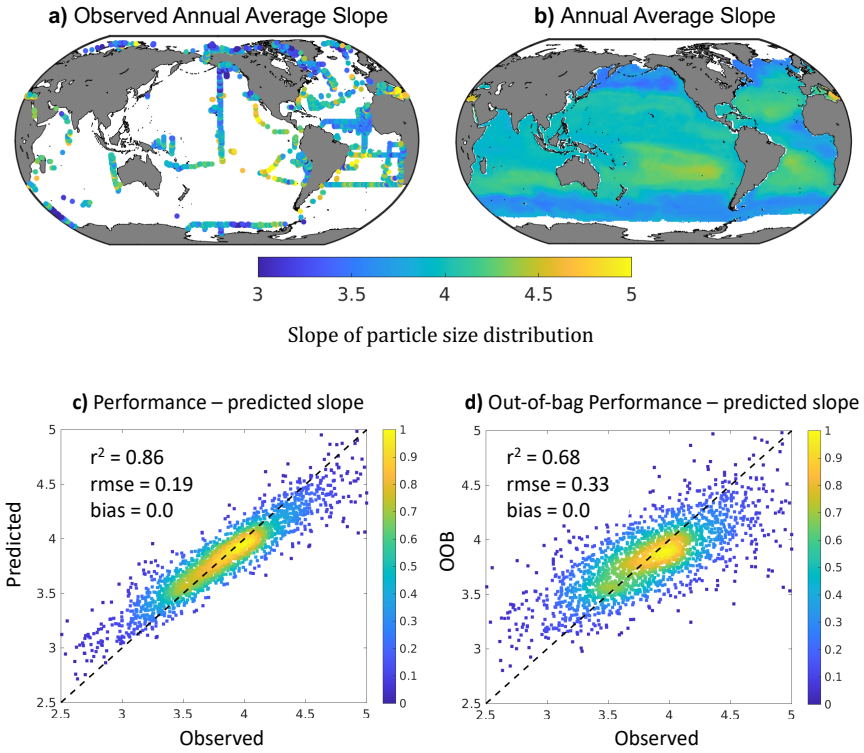

Figure 3. Observed and reconstructed PSD slope at the base of the euphotic zone. (A) Map of the observations of PSD slope, for locations with observations in multiple months the average is shown (B) Map of the PSD slope reconstructions, colored dots show the observations from A. (C) Performance of the RF reconstruction shown as density scatter plots of predicted vs. observed particulate slope (colors indicate the normalized density of grid points surrounding a given cell). (D) Same as B, but using out-of-bag (OOB) predictions, i.e., predictions vs. observations withheld from training. Annotations in B and C show the square of the correlation coefficient (r2), the RMSE and the global bias.

Consistently, we find that slope and biovolume are negatively correlated ($r^2 = 0.4, p < 0.01$, Supplementary Fig. S3), roughly indicating 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 conditions. Since large particles contribute proportionally more than smaller particles to export fluxes, because of their faster sinking speed, this relationship suggests that biovolume and slope will synergistically enhance export fluxes in particle-rich regions, and depress them in particle-poor regions.

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. 3), 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 system, 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.

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 (Supplementary Fig S4). 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 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. As discussed above, spatial and temporal decoupling of the biovolume-slope relationship could have important consequences for the patterns of particle export flux.

A recursive feature elimination suggests that multiple variables are required for a robust reconstruction of PSD (Supplementary Fig. S2). Among these, we highlight chlorophyll, mixed layer depth, and oxygen, each with 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 can not be directly tied to these rankings, highlighted predictors can provide insights into the role of different processes that may be affecting PSDs. We note that certain predictors with mechanistic links to particle export, for example silicate, which serves as a proxy for diatom production, were deemed unimportant and are not included in the final RF regressions. It is likely that information contained in these predictors is shared by correlations with other variables, and is picked up by the main predictors used by the method.

Multiple variables are significantly correlated with biovolume and slope (Supplementary Fig. S6). In particular, we find that biovolume correlates positively and robustly with chlorophyll ($r^2 = 0.25, p < 0.01$, Supplementary Fig. S3). This is not surprising, because phytoplankton are ultimately the main source of organic matter and sinking particles in the ocean. However, we find that chlorophyll is not as strong a predictor of slope, when the whole ocean is considered ($r^2 = 0.04, p < 0.01$, Supplementary Fig. S3), and that additional predictors are needed for robust slope reconstructions. This result contrasts with 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. Although likely not directly related to the PSD, 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 mesopelagic community responsible for this respiration. We note that the PSD slope is an emergent property that reflects the interaction of physical and biological processes that are still poorly understood. Not surprisingly, slope is harder to reconstruct than biovolume, and shows overall weaker correlations with other individual predictors (Supplementary Fig. S6).

Spatial patterns in slope share several features with estimates of phytoplankton size spectra from observations and models (Kostadinov et al., 2009; Roy et al., 2013; Barton et al., 2013; Ward et al., 2014), reflecting the importance of phytoplankton size structure and composition for particle export. Satellite-based reconstructions of phytoplankton functional groups (e.g., Mouw et al., 2017) could be included as predictors. However, methodological shortcomings and disagreement between different approaches currently limit their applicability — something that may be mitigated by future advances. It is also likely that information related to phytoplankton composition implicitly enters the RF regression via relationships with predictors such as surface chlorophyll and temperature.

3.2 Particle export flux

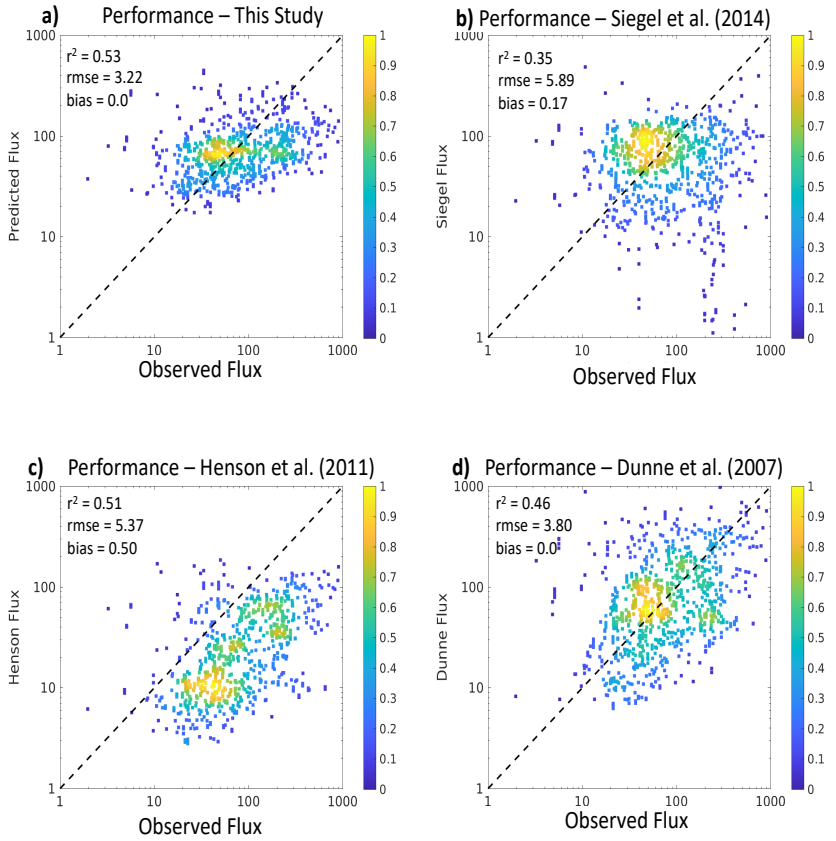

Figure 4. Density scatter plots showing the relationships between *in situ* flux observations and global flux reconstructions from (A) This study, (B) Siegel et al. (2014)l, (C) Henson et al. (2011), (D) (Dunne et al., 2007). Colored dots represent the relative density of grid points surrounding the data point, dashed line indicates a 1:1 ratio. Annotations show the correlation coefficient (r^2) RMSE, and average bias.

Based on the PSD reconstructions and Equation 5, we optimize for the particle sinking speed and carbon content parameters (m_0 and μ) that produce export fluxes in best agreement with *in situ* observations (see Section 2.2). This approach results in a value of 27.65 mgC m s^{-1} for m_0, and of 2.90 for μ, both in the range suggested by *in situ* observations (Kriest, 2002), and comparable to values adopted by previous studies (Kriest, 2002; Stemmann et al., 2004; Guidi et al., 2008; Kiko et al., 2017; Bianchi et al., 2018).

The resulting carbon fluxes compare well with sediment trap and thorium-based observations (Fig. 4), performing in a similar way or better than previous estimates (Henson et al., 2011; Dunne et al., 2007; Siegel et al., 2014). Our estimate reduces the overall uncertainty (here expressed by the RMSE) compared to previous work, and shows negligible bias. However, our method also reduces the overall range of reconstructed fluxes, i.e., it overestimates the flux at low values and underestimates it a high values compared to observations. This bias may be related to the tendence of the RF algorithm to underpredict extremes in both biovolume and slope (Figs. 2 and 3).

Extrapolated to the whole ocean, our method reveals spatial patterns of export fluxes in broad agreement with previous studies, with some notable differences (Fig. 5). 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 relatively high fluxes even 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 3.2.1). We also reconstruct substantially stronger export than previously found in the Southern Ocean, in particular south of 50S (see discussion in Section 3.3).

Globally integrated, we estimate a particle export flux of 6.7 $\pm$ 0.4 PgC/y, in excellent agreement with the range of observational and model-based estimates of the biological gravitational pump (4-9 PgC/y, Boyd et al. (2019)). Compared to other spatially-resolved reconstructions, our global flux sits between the low-value estimate of Henson et al. (2011) (3.0 $\pm$ 0.3 PgC/y) and the high-value estimate of Dunne et al. (2007) (9.8 $\pm$ 0.4 PgC/y).

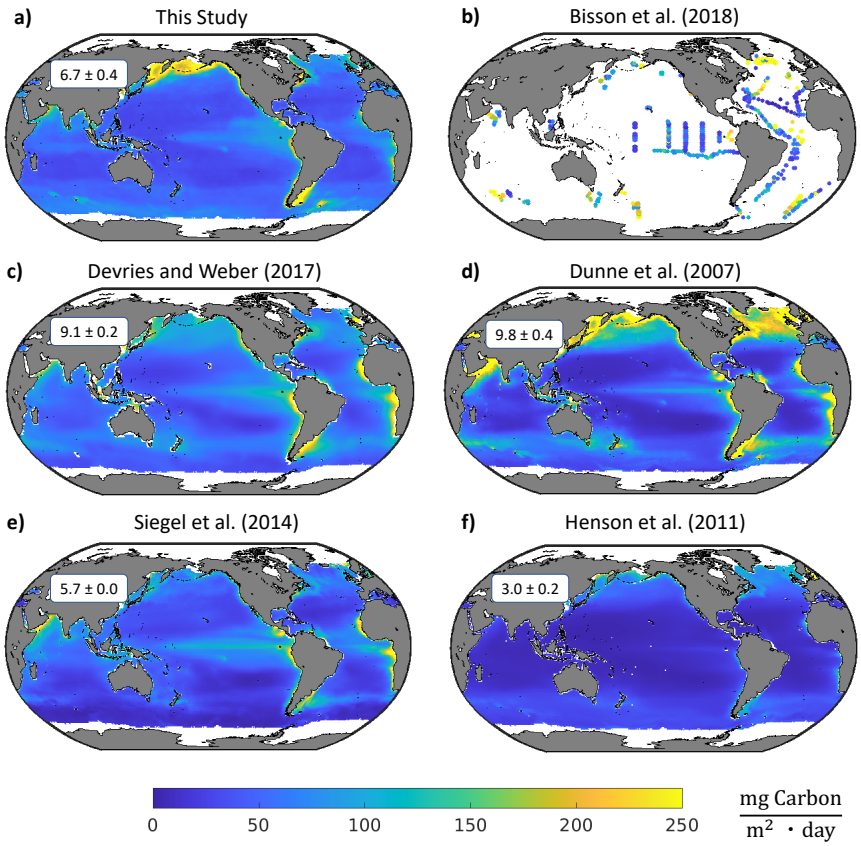

Figure 5. Annual average particle export from the euphotic zone for the (a) Random forest derived compared to the *in situ* data of (b) Bisson et al. (2018), the steady state satellite driven model (c) SIMPLE-TRIM of DeVries et al. (2017), empirical models of (d) Dunne et al. (2007) and (f) Henson et al. (2011) and the satellite-driven euphotic zone food web model of (e) Siegel et al. (2014). Annotated in each figure is the calculated total export and uncertainty reported by each study.

3.2.1 Meridional variability

We illustrate the main spatial differences between our and other reconstructions by considering zonally averaged export fluxes (Fig. 6). 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 exports are 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. 6a), in the Atlantic Ocean it dramatically increases moving from the Northern to the Southern Hemisphere (Fig. 6b). These patterns reflect a combination of open-ocean and shelf enhanced particle export. Specifically, high exports in the Northern Pacific and Southern Atlantic Oceans are partly driven by large fluxes in the Bering Sea, the Sea of Okhotsk, and the Patagonian shelf. At lower latitudes, coastal upwelling systems sustain high export in the northern Indian Ocean and the tropical to subtropical Atlantic Ocean.

Variations in export patterns reflect a combination of varying particle biovolume (Fig. 2) and PSD slope (Fig. 3). These two quantities generally correlate in such a way as to increase export fluxes in particle-rich waters, where large, fast-sinking particles tend to be relatively more abundant than small particles, and decrease them in particle-poor waters where small particles dominate (see Section 3.2, and Supplementary Fig. S3). High export in the eastern equatorial and tropical Pacific can be attributed to a substantial increase in biovolume, with a minor contribution from PSD slope, which appears to be more uniform across the region. The picture is somewhat different in the equatorial Atlantic Ocean, where a more substantial flattening of the PSD suggests a more important role of large particles in driving elevated export fluxes. A similar interaction of particle abundance and size-structure dramatically intensify fluxes in the subpolar North Pacific and Southern Ocean, and to a lesser extent the subpolar Atlantic, where a relative increase in particle abundance is followed 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.

Our reconstruction shows broad meridional patterns similar to previous estimates (Fig. 6); however, significant regional-level discrepancies remain. For example, in the low latitudes, we predict somewhat less intense equatorial export peaks and subtropical lows, compared to the estimates of Dunne et al. (2007) and Siegel et al. (2014). In this respect, our reconstruction is more in line with the results of DeVries and Weber (2017). In the subpolar Pacific, our estimate shows a northward shift of maximum export 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 iron leakage from shelves and marginal seas (Nishioka et al., 2020). 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 (see also Section 3.3).

3.2.2 Seasonal cycle

The seasonal cycle of particle export is comparable to previous studies (Fig. 7), when averaged over large-scale coherent biomes (Weber et al., 2016). 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.

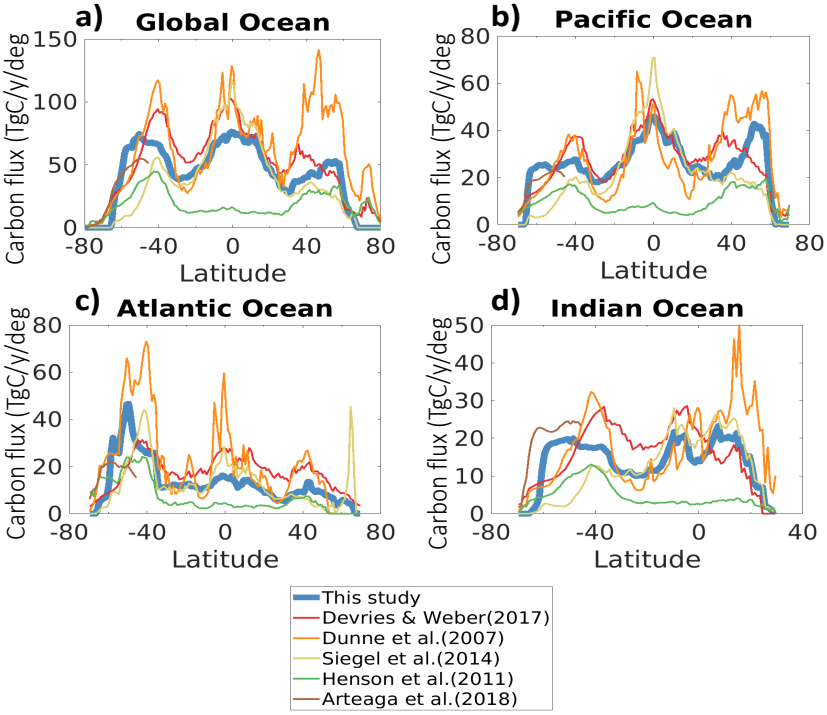

Figure 6. Zonally integrated annual export for the (a) Global Ocean, (b) Pacific Ocean, (c) Atlantic Ocean, and (d) Indian Ocean. Each color represents a different study, as listed in the legend (bottom).

Similar to other estimates, we capture well-known seasonal export pulses associated with spring phytoplankton blooms in the North Atlantic and North Pacific Oceans. Over most of the tropics, our reconstruction reveals nearly constant seasonal cycles, and a slight asymmetry about the equator with more pronounced seasonality in the Northern Hemisphere compared to the Southern Hemisphere. 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 detail in Section 3.3.

To better quantify seasonality and provide a more objective comparison to other studies, we define a seasonality index as the range of monthly export divided by the annual mean flux in each region (Supplementary Fig. S10). A higher seasonality index is indicative of a more dynamic export cycle, with more dramatic variations between low and high export periods. As expected, seasonality is larger in high latitudes, and decreases toward the tropics; the highest values are reached in the mid-latitude to subpolar North Pacific and Atlantic, and higher variability is confirmed in the tropics and subtropics of the Southern Hemisphere compared to the Northern Hemisphere.

The relatively muted seasonality of our reconstruction, compared to previous work, is consistent with the weaker spatial gradients discussed in the previous sections, and suggests less dramatic gradients in net community production and export than previously assumed. The machine learning approach used in this study relies on non-linear relationships with multiple ocean variables to reconstruct particle export fluxes, which may reveal 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

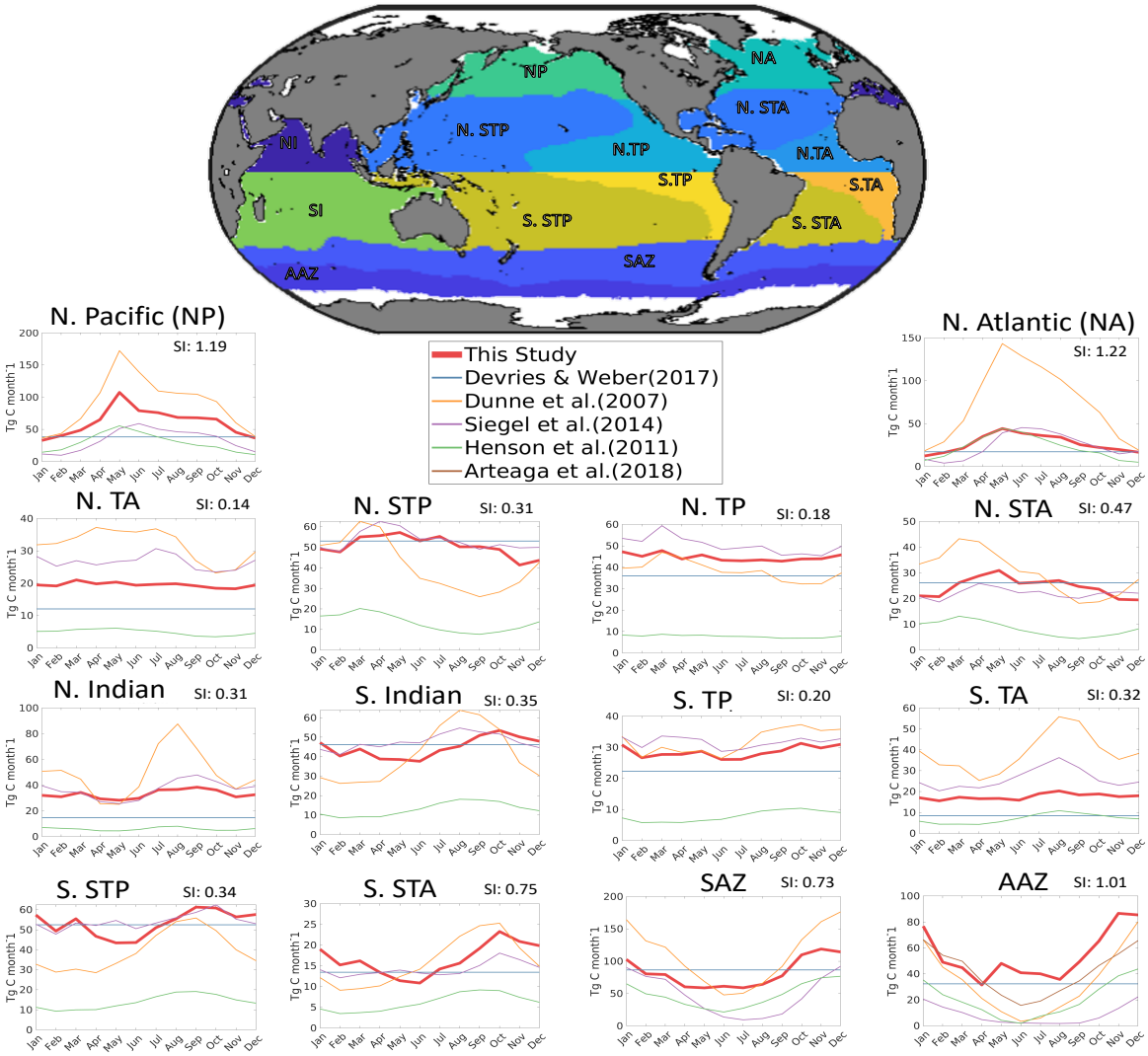

Figure 7. Annual seasonal cycle of particle flux from the euphotic zone for the regions specified in the map (top). Each line corresponds to a study shown in the legend (top right). The same seasonal spatial mask was applied to each study.

with other variables that may be important in modulating spatial and seasonal export patterns. It is also possible that our method somewhat underestimates variability compared to previous work. As previously noted (see Section 3.1), RF ensemble reconstructions appear to reduce extremes in both biovolume and PSD slope, potentially underestimating variability in particle export fluxes derived from these quantities. When compared with other studies, our results show overall similar patterns in seasonal variability, but lower seasonality in most regions, particularly at higher latitudes (Fig. 7 and Supplementary Fig. S10).

3.3 Southern Ocean Export

Export flux in the Antarctic Zone are substantially larger than other global reconstructions, especially during winter (Fig. 7). A regional study based on 10 years of bio-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. 8). 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 reveals substantially higher export near landmasses, for example the Patagonian Shelf, 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 winter 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 in winter (Fig. 7).

This discrepancy with prior global estimates in Antarctic Zone export could be due to a variety of factors. First, UVP5 observations in the Southern Ocean, particularly in winter, are extremely scarce. Similarly, satellite-based observations of predictors based on ocean color dwindle in wintertime. Other climatological variables, such 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 our model result in strong inter-model variability, highlighting the spread between different RF realizations. Second, our reconstruction reveals significant export primarily next to land masses in the Atlantic and Indian sectors of the Southern Ocean. Proximity to land masses has been shown to increase productivity and carbon flux (Jouandet et al., 2014), presumably via iron fertilization from terrestrial and sedimentary sources in otherwise high-nutrient low-chlorophyll waters. It is possible that other flux reconstructions underestimate this rapid aggregation of particles and increased particle export, in particular during winter, when observations are scarce. Increasing the number of particle flux and size distribution observations from the Antarctic Zone, in particular downstream of major land masses and in wintertime, together with a better regional understanding of export processes, could help shed more light on the patterns of export in the region.

3.4 Mixed layer versus euphotic zone export

By exploiting the high vertical resolution of UVP5 measurements, our approach allows reconstruction of particle export at arbitrary horizons. We illustrate this capability by estimating particle fluxes at the depth of the maximum climatological wintertime mixed layer, and compare them to export from the climatologically-varying euphotic zone, shown in Fig. 5. We find that, globally integrated, the particulate carbon export from the mixed layer is 9.4 $\pm$1.1 PgC/y, i.e., about 3 Pg/year larger than the global export from the euphotic zone. This estimate is slightly lower than observational estimates of carbon export and sequestration from the same depth horizon based on ANCP and POC export estimates (Emerson, 2013).

Overall, export at the wintertime mixed layer follows broad spatial patterns similar to the export from the euphotic zone (Fig. 9a). However, tropics and subtropics show larger export fluxes from the mixed layer (locally, up to a few times), while high latitudes show overall weaker export fluxes (Fig. 9b). The low-latitude intensification of mixed layer particle fluxes is similar in all ocean basins, and more than compensates for the reduction at high latitudes (Supplementary Fig. S12), 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,

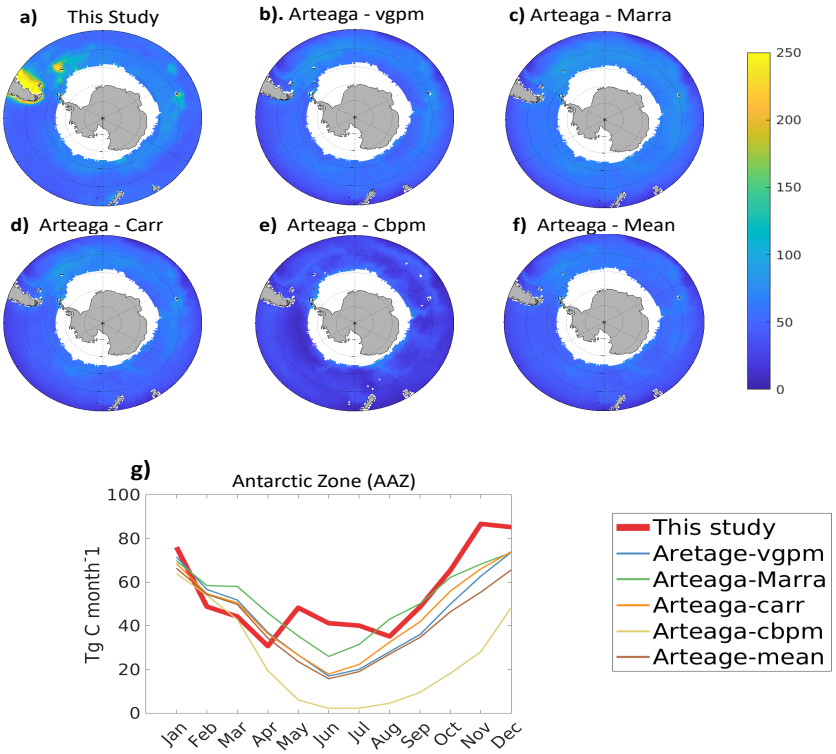

Figure 8. Southern Ocean specific export for (a) This study, and (b-e) the various models from Arteaga et al. (2018), and (f) the mean from that study. Each model from Arteaga et al. (2018) represents a different net primary productivity algorithm used to derive export. (g) Seasonal cycle of export for each model in the Antarctic zone (shown in figure 7).

export from the mixed layer in the Southern Ocean is substantially depressed compared to export from the euphotic zone.

Differences between euphotic zone and mixed layer export can be best interpreted by considering the different depth of these horizons (Palevsky & Doney, 2018). The depth of the maximum mixed layer is shallower than the euphotic zone in the tropics and subtropics, and is deeper in high latitudes (Fig. 9c). This suggests that shallower export horizons are generally characterized by higher fluxes than deeper export horizons, likely because of remineralization of particles in the upper layers of the ocean. 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. 9c). Here, particle remineralization between the wintertime mixed layer and the euphotic zone depth reduces export from the latter horizon, indicating net heterotrophy in the deeper layers of the euphotic zone. Over subpolar regions, the wintertime mixed layer is deeper on average than the climatological euphotic zone. Here, export fluxes reach maximum values within the euphotic zone, and decrease below it following remineralization. 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 depth are very similar in magnitude, suggesting a close seasonal compen-

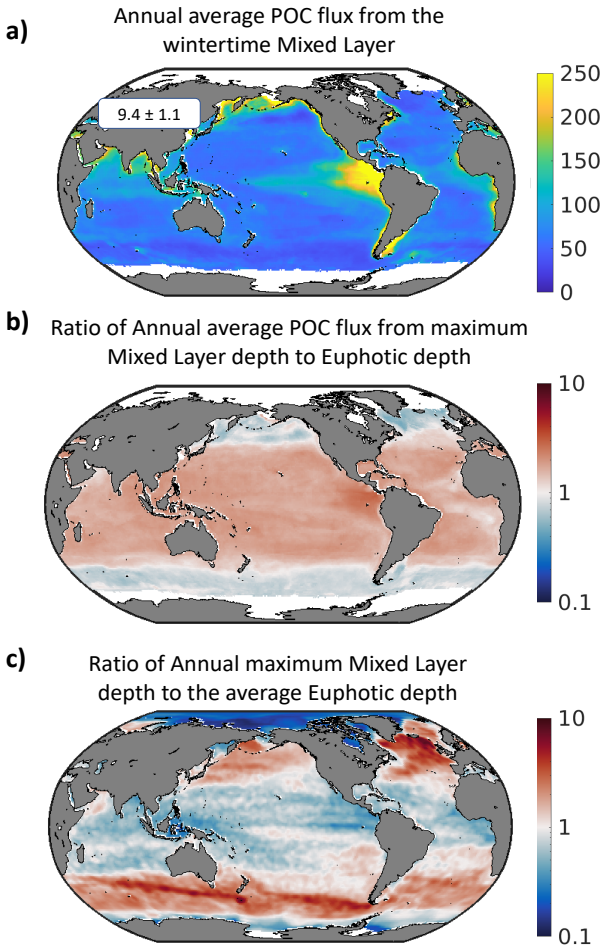

Figure 9. (a) Annual mean particle export from the maximum mixed layer depth. (b) Ratio of the export from the maximum mixed layer depth to the export from the euphotic zone. (c) Ratio of the maximum mixed layer depth to the euphotic zone depth.

583 sations between enhanced euphotic zone fluxes when this horizon is found above the win-
584 tertime mixed layer, and reduced euphotic zone fluxes when it is found below it.

585 Ultimately, differences in export between the euphotic zone and the wintertime mixed
586 layer are important when considering the role of the biological pump for carbon seques-
587 tration (Palevsky & Doney, 2018). Our results suggest that more carbon is sequestered
588 below the wintertime mixed layer than leaves the euphotic zone. The sensitivity of these
589 fluxes to climate variability and change is also likely to differ (Palevsky & Doney, 2021).

590 ### 3.5 Caveats to our approach

591 There are multiple sources of uncertainty and inherent limitations that could af-
592 fect our estimates and call for further work. First, expanding the coverage of observa-
593 tions with UVP5 and similar instruments, in particular in under-sampled regions char-
594 acterized by large variability, such as the Southern Ocean, would improve the robust-

ness of our estimates, and shed additional light on regional export 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.

Second, supervised learning methods are only as reliable as the data used for training; 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.

Third, 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 fluxes. Different machine learning methods that are better suited to capture extreme values should be tested to address this limitation, in conjunction with more detailed analyses of particle flux observations, including at time-series stations.

Finally, we use globally-averaged relationships between particle size, sinking speed, and carbon content, which we optimize against observations. However, these quantities are likely to depend on region and time of the year, reflecting variable underlying ecological processes. More work combining *in situ* and optical measurements should focus on constraining these quantities and their regional and temporal variability.

4 Conclusions

In this paper, we provide a new, data-constrained estimate of particle fluxes based on global UVP5 observations of PSDs. It captures regional and seasonal variability in observed PSD properties and export fluxes, and demonstrates the ability of statistical machine learning methods to extrapolate these quantities globally. Our approach also allows reconstruction of export fluxes from multiple depths, highlighting the importance of the choice of export horizon (Palevsky & Doney, 2018), and paving the way to fully three-dimensional particle flux reconstructions.

We obtain a global particle export flux of 6.7 $\pm$ 0.4 PgC/year from the euphotic zone, in line with previous work, although with significant regional and temporal differences. Our results suggest weaker spatial and seasonal variability in particle fluxes compared to previous studies, in particular in the open ocean, while highlighting the importance of coastal waters and marginal seas for high latitude export. We also capture similar patterns of high latitude seasonal blooms in the Northern Hemisphere as previous studies, but less variable flux in the tropical to subtropical ocean, and substantially higher year-round export in the Southern Ocean, in better agreement with regional estimates (Arteaga et al., 2018). Results from the Southern Ocean suggest that processes that sustain elevated fluxes, in particular in wintertime, 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 model (Dunne et al., 2007).

The statistical nature of our machine learning approach does not directly reveal mechanisms behind PSD and export fluxes. However, we are able to highlight globally coherent patterns, and the relative roles of particle abundance and size structure in driving export fluxes. Specifically, we show that the total particle biovolume and the PSD slope show similar patterns, and are in fact correlated in such a way to act synergistically on particle fluxes (Supplementary Fig. S3). We also suggest distinct deviation from these patterns, for example in the tropical and northern subtropical Pacific Ocean, where high availability of particles, rather than dominance of large vs. small particles, tends to drive elevated export.

We illustrate the ability of our method to reconstruct particle fluxes at multiple depth horizons by reconstructing and comparing carbon export from the euphotic zone and the wintertime mixed layer depth. Export from the base of the wintertime mixed layer is overall stronger than export from the euphotic zone in low and mid latitudes, and weaker in high latitudes, driving a significantly larger global particle export of 9.4 ±1.1 PgC/year. This suggests that more carbon is sequestered for decadal to centennial timescales than is available for remineralization in the twilight zone of the ocean.

We attribute the disparity in flux from the euphotic zone and mixed layer to differences in their depths: maximum mixed layers are typically shallower than the euphotic zone over large swathes of the low latitude ocean, but deeper in high latitudes. Future three-dimensional reconstructions of particle fluxes would allow a closer investigation of the processes controlling transfer of carbon between the surface and the deep ocean, providing further insights into particle transformation processes, transfer efficiency, and ocean carbon sequestration.

Acknowledgments

This material is based upon work supported by the U.S. National Science Foundation under Grant No. OCE-1635632. D.B. acknowledges support from the Alfred P. Sloan Foundation. A.M.P.M acknowledges support from NSF Award No. 1654663. T.W. was supported by NSF award OCE-1635414. RK acknowledges support via the BMBF funded project CUSCO, the EU project TRIATLAS (European Union's Horizon 2020 programme, grant agreement No 817578) and a "Make Our Planet Great Again" grant of the ANR within the "Programme d'Investissements d'Avenir"; reference "ANR-19-MPGA-0012". Data generated by this analysis has been uploaded to BCO-DMO, with a DOI pending. The individual UVP5 profiles used to generate the reconstructions can be obtained on the EcoTaxa website https://ecotaxa.obs-vlfr.fr/part/.

References

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

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

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., & 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

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

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

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

Boyd, P. W., Sherry, N. D., Berges, J. A., Bishop, J. K., Calvert, S. E., Charette, M. A., ... Wong, C. S. (1999). Transformations of biogenic particulates from the pelagic to the deep ocean realm. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *46*(11-12), 2761–2792. doi: 10.1016/S0967-0645(99)00083-1

Brun-Cottan, J. (1971). Etude de la granulométrie des particules marines, mesures effectuées avec un compteur coulter. *Cah. Oceanogr*, *23*, 193–205.

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

C.C.C.S. (2017). *Era5: Fifth generation of ecmwf atmospheric reanalyses of the global climate.* Copernicus Climate Change Service Climate Data Store (CDS). Retrieved from `https://cds.climate.copernicus.eu/cdsapp#!/home` (accessed: 11-13-2019)

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

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

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., & 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

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

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

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

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.

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

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., 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

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

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.

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

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

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

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

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.

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

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

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

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

Palevsky, H. I., & Doney, S. C. (2021). Sensitivity of 21st century ocean carbon export flux projections to the choice of export depth horizon. *Global Biogeochemical Cycles*, e2020GB006790.

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

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

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

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

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., ... Steinberg, D. K. (2016). Prediction of the export and fate of global ocean net primary production: The exports science plan. *Fron-*

tiers in Marine Science, *3*(MAR), 1–10. doi: 10.3389/fmars.2016.00022

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., 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.

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., Jackson, G. A., & Ianson, D. (2004). 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., 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

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., 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

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

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.