Constraining the Particle Size Distribution of Large Marine Particles in the Global Ocean With In Situ Optical Observations and Supervised Learning

D. J. Clements, S. Yang, T. Weber, A. M. P. Mcdonnell, R. Kiko, L. Stemmann, D. Bianchi

Global Biogeochemical Cycles®

RESEARCH ARTICLE
10.1029/2021GB007276

Key Points:

- We use optical observations of marine particle size distribution to reconstruct global climatological particle biovolume (BV) and spectral slope
- We describe the importance of different biogeochemical variables on particle BV and spectral slope
- Spatial and seasonal variations of BV and slope are likely to have synergistic effects on carbon export

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

D. J. Clements and D. Bianchi,
dclements@atmos.ucla.edu;
dbianchi@atmos.ucla.edu

Citation:

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 particles in the global ocean with *in situ* optical observations and supervised learning. *Global Biogeochemical Cycles*, *36*, e2021GB007276. https://doi.org/10.1029/2021GB007276

Received 16 DEC 2021
Accepted 28 APR 2022

Constraining the Particle Size Distribution of Large Marine Particles in the Global Ocean With *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], and 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, NY, USA, [3]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK, USA, [4]Sorbonne Université, CNRS, UMR 7093, Institut de la Mer de Villefranche sur mer, Laboratoire d'Océanographie de Villefranche, Villefranche-sur-Mer, France

Abstract The abundance and size distribution of marine particles control a range of biogeochemical and ecological processes in the ocean, including carbon sequestration. These quantities are the result of complex physical-biological interactions that are difficult to observe, and their spatial and temporal patterns remain uncertain. Here, we present a novel analysis of particle size distributions (PSDs) from a global compilation of in situ Underwater Vision Profiler 5 (UVP5) optical measurements. Using a machine learning algorithm, we extrapolate sparse UVP5 observations to the global ocean from well-sampled oceanographic variables. We reconstruct global maps of PSD parameters (biovolume [BV] 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 BV and flatter PSD slope, that is, a high relative abundance of large versus small particles. The resulting negative correlations between particle BV and slope further suggests synergistic effects on size-dependent processes such as sinking particle fluxes. Our approach and estimates provide a baseline for an improved understanding of particle cycles in the ocean, and pave the way to global, three-dimensional reconstructions of PSD and sinking particle fluxes from the growing body of UVP5 observations.

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 & Peng, 1982; Kwon et al., 2009; Sarmiento & Gruber, 2006). Particles can be remineralized by microbial degradation and grazing by zooplankton (Karl et al., 1988; Giering et al., 2014; Steinberg et al., 2008), aggregate and disaggregate following physical and biological interactions (Briggs et al., 2020; Burd & Jackson, 2009; Dilling & Alldredge, 2000; Jackson, 1990; Kiørboe et al., 1990), and sink out of the surface layers to transfer carbon and nutrients to the ocean interior (Boyd et al., 2019; Buesseler et al., 2007; Turner, 2015). Marine particles are hotspots of microbial activity and diversity (Church et al., 2021; DeLong et al., 1993; Karl et al., 1984), provide the bulk of energy that sustain mesopelagic and abyssal ecosystems (Burd et al., 2010; Giering et al., 2014), and eventually supply materials to the ocean sediment, where burial sequesters them for geological timescales (Dunne et al., 2007; Sarmiento & Gruber, 2006).

Most processes that drive particle cycles in the ocean depend on particle size (Stemmann & Boss, 2012). To first order, size controls particle elemental composition (Alldredge & Gotschalk, 1988), aggregation and disaggregation rates (Briggs et al., 2020; Burd & Jackson, 2009), and the ability of particles to sink (Alldredge & Gotschalk, 1988; Cael et al., 2021; McDonnell & Buesseler, 2010), 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 (Bianchi et al., 2018; Jackson, 1989; Kiørboe et al., 2002; Nguyen et al., 2022), and coupling with seawater chemistry via regeneration of elements (Broecker & Peng, 1982; Sarmiento & Gruber, 2006), adsorption, and scavenging processes (Ohnemus et al., 2019; Turekian, 1977).

Therefore, the abundance of particles of different sizes, that is, 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 (Burd & Jackson, 2009; Turner, 2015). Conversely, disaggregation and consumption by microorganisms tend to reduce the size of particles and their sinking speed, facilitating shallow remineralization (Briggs et al., 2020; Goldthwait et al., 2005; Karl et al., 1988).

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 (Burd & Jackson, 2009; Cram et al., 2018; Devries et al., 2014; Gehlen et al., 2006; Kriest & Evans, 1999; Nguyen et al., 2022; Omand et al., 2020; Stemmann et al., 2004; Weber & Bianchi, 2020). 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, 2010). 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; 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; Huete-Ortega et al., 2014; Reynolds & Stramski, 2021; Stemmann & Boss, 2012). 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 (DeVries & Weber, 2017; Siegel et al., 2014).

Recent advances in ocean optical observations enable direct determination of in situ PSD throughout the water column (Boss et al., 2015; Lombard et al., 2019; Stemmann & Boss, 2012, 2020). 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., 2022).

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 (Alldredge & Gotschalk, 1988; Guidi et al., 2008; Kiko et al., 2017; Kriest, 2002). 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 paper 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 our results to previous studies, discussing the uncertainties and caveats inherent to our approach. Section 4 summarizes the main findings and discusses future directions.

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 (Huete-Ortega et al., 2014; Kiørboe et al., 1993; Reynolds & Stramski, 2021; Sheldon et al., 1977), 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:

$$n(s) = n_0 \cdot s^{-\beta}, \tag{1}$$

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 fist 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; Reynolds & Stramski, 2021). 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; Huete-Ortega et al., 2014; Karp-Boss et al., 2007; Organelli et al., 2020; Runyan et al., 2020), or the influence of mineral phases from riverine inputs and exchange with sediment in coastal waters (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, disaggregation (Briggs et al., 2020; Huete-Ortega et al., 2014; Kiørboe et al., 1993; Reynolds & Stramski, 2021; Sheldon et al., 1977), and zooplankton grazing dynamics (Moscoso et al., 2022; Schartau et al., 2010). Accordingly, alternative methods have been proposed to approximate PSD observations, including sums of log-normal functions (Jonasz & Fournier, 1996) and non-parametric approaches based on cumulative PSD (Reynolds & Stramski, 2021). 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, 2016; Roullier et al., 2014; Stemmann & Boss, 2012).

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 1 to observed normalized particle concentrations, that is, 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

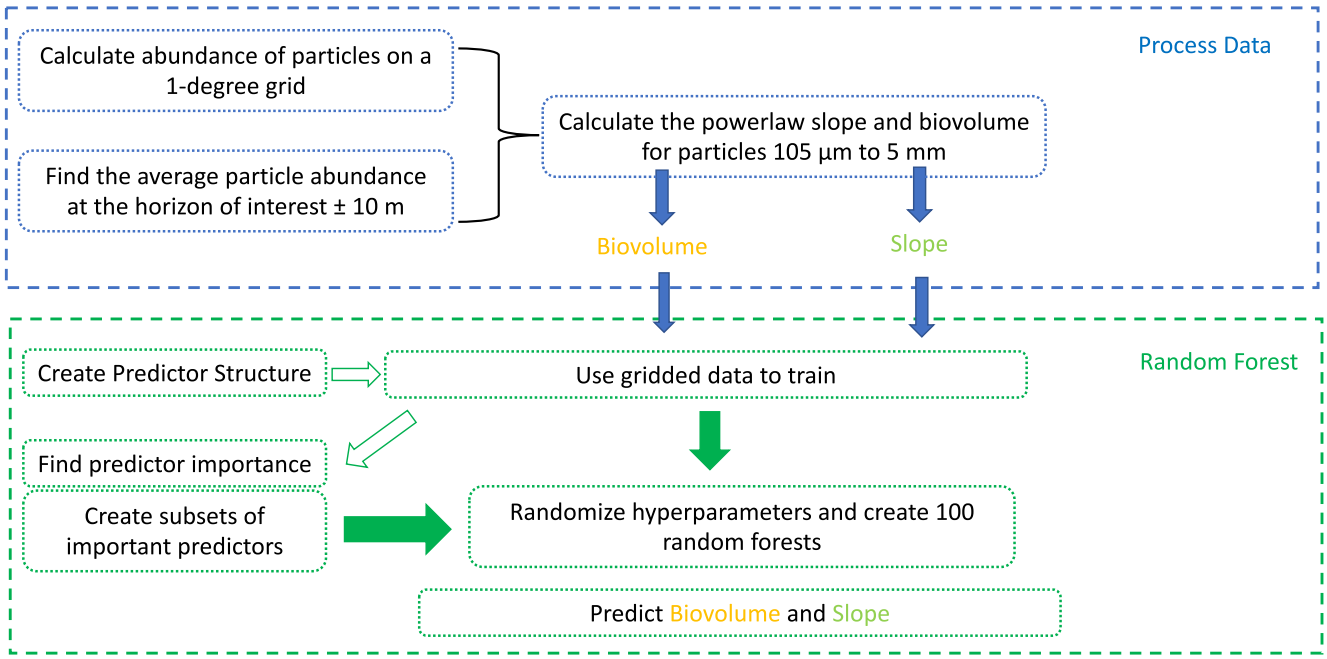

Figure 1. Schematic diagram illustrating the general workflow of processing UVP5 observations into a global particle size distribution (PSD) data set. Observations are ensembled onto a standard 1° grid, with observation representing an average of a 20 m vertical bin about the euphotic zone depth. PSD observations (power law slope and biovolume [BV]) are calculated for the 105 µm to 5 mm size range. The PSD slope and BV are globally extrapolated using a bagged Random Forest algorithm.

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 Figure 1, and are discussed in the following sections.

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., 2022). The data set consists of over 6,700 profiles from 119 cruises, collected from 2008 to 2020 (Figure 2). These observations provide robust particle counts for the 105 µm–5 mm size range at each location and depth. Under the power law assumption (Equation 1), the two parameters n_0 and β are needed to capture the PSD (Bader, 1970; Devries et al., 2014; Sheldon et al., 1972; Stemmann & Boss, 2012; Stemmann et al., 2004).

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:

$$BV = \int_{s_{\min}}^{s_{\max}} n(s) \cdot \frac{\pi}{6} \cdot s^3 \, ds. \tag{2}$$

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 BV can also be expressed analytically as a function of the slope and power law constant, by substituting Equation 1 into Equation 2:

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

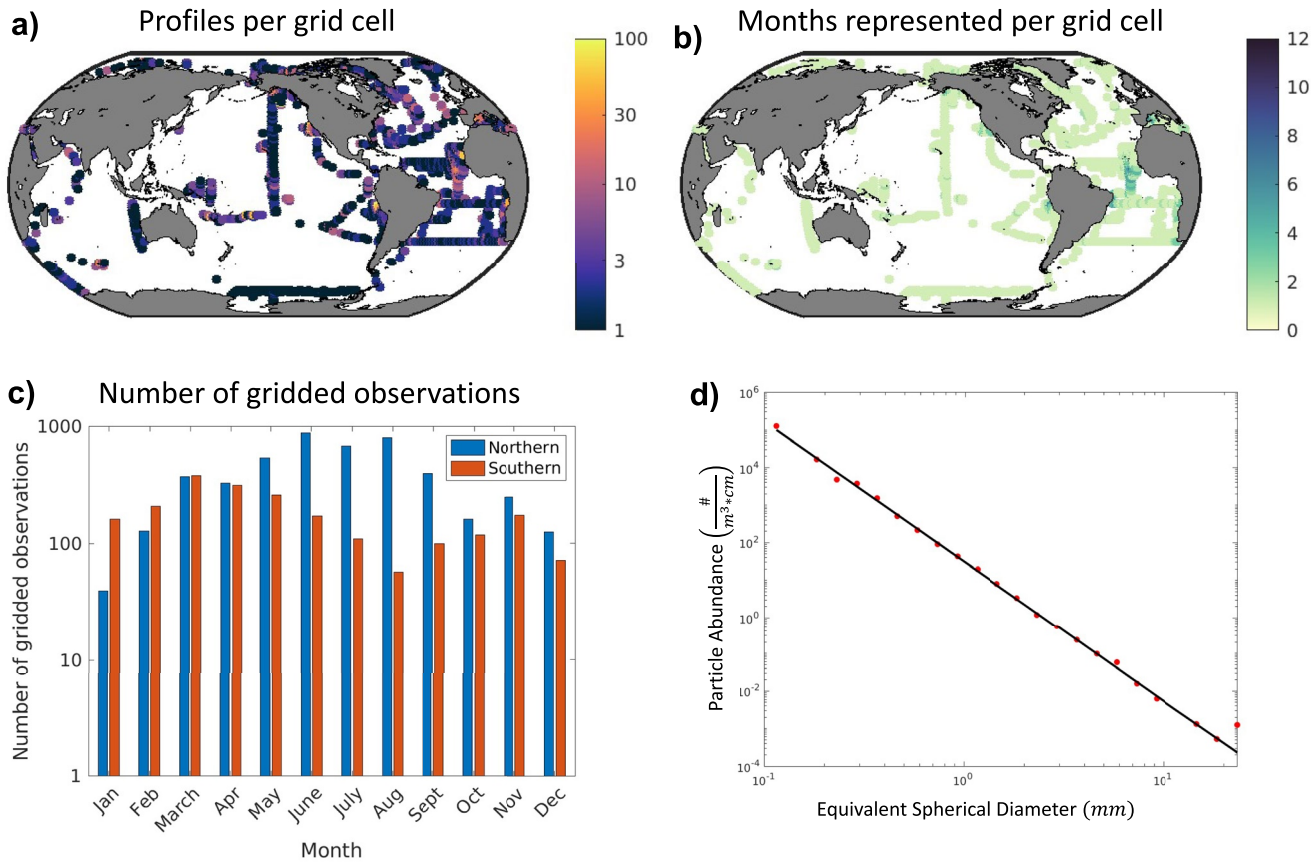

Figure 2. Global distribution of the UVP5 observations used in this study. (a) Number of profiles per one-degree resolution grid cell. (b) Number of months represented in each grid cell. (c) Histogram showing the monthly distribution of observations for the Northern Hemisphere (blue) and the Southern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

By fixing the size range, that is, the minimum and maximum particle size that can be robustly derived from UVP5 instruments ($s_{\min}$ and $s_{\max}$ respectively), we solve Equation 3 for the power law constant n_0 as a function of the PSD slope and the observed BV:

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

We set the minimum and maximum size for this equation to the same values used to estimate the slope and BV 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 BV at a variety of locations sampled by UVP5 (Forest et al., 2012; Stemmann & Boss, 2012; Stemmann et al., 2008).

We coarsen the temporal and spatial resolution of the UVP5 profiles by binning them onto the standard monthly 1°-resolution grid of the World Ocean Atlas (Garcia et al., 2018; 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° 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° resolution enables us to utilize the native resolution of many predictors with minimal post-processing, for example, 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 m-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 BV 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 (~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 $R^2 = 0.98$ (Figure S1 in Supporting Information S1). 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 (Figure S1 in Supporting Information S1), and likely reflect a combination of allochthonous mineral particle supply and highly productive or variable conditions (Reynolds & Stramski, 2021; Reynolds et al., 2010, 2016). The results of this quality check indicate that, for the size range considered here (105 μm–5 mm), the majority of open ocean UVP5 observations can be well approximated by power laws, in line with previous work (Cram et al., 2018; Guidi et al., 2009; Roullier et al., 2014; Stemmann & Boss, 2012). 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 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.

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 BV 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 BV) 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 (~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, that is, the so-called "out-of-bag" (OOB) data.

The rank of predictors is given by the OOB 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 OOB error to describe the contribution of each predictor to the final solution. When there is relatively no change in the OOB error for every additional predictor, these predictors are considered not important for the RF (Figure S2 in Supporting Information S1). 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.

2.1.2. 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° 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 and

Table 1
Variables Used to Predict Particle Size Distribution Parameters, Variations (i.e., Vertical or Temporal Changes), and Data Sources

Category	Variable	Short name	Variations	Source
Universal				
	Topography	topo		N.G.D.C (2006)
	Temperature below MLD	temp_deep	Time derivative	Locarnini et al. (2019)
	Chlorophyll	Chlorophyll_modis	Time derivative	NASA G.S.F.C (2014)
	Oxygen	o2_ml	ML/ML + 100 m	Garcia et al. (2019)
		o2_deep	Time derivative	
	Shortwave radiation	Shortwave	Time derivative	Copernicus Climate Change Service (2017)
	Nitrate	no3_ml	ML/ML + 100 m	Garcia et al. (2018)
		no3_deep	Time derivative	
	Phosphate	po4_ml	ML/ML + 100 m	Garcia et al. (2018)
		po4_deep	Time derivative	
	Salinity	salt	ML/ML + 100 m	Zweng et al. (2019)
Mixed layer				
	Mixed layer	MLD_MIMOC	Time derivative	Johnson et al. (2012)
	Mixed layer	MLD_DBM	Time derivative	de Boyer Montégut et al. (2004)
Primary production				
	Eppley VGPM	Eppvgpm	Time derivative	Antoine and Morel (1996)
	VGPM	vgpm	Time derivative	Behrenfeld and Falkowski (1997)
	CBPM	cbpm	Time derivative	Westberry et al. (2008)
	CAFE	cafe	Time derivative	Silsbe et al. (2016)
NPP standard deviation				
	Eppley VGPM	Eppvgpm_std		Antoine and Morel (1996)
	VGPM	vgpm_std		Behrenfeld and Falkowski (1997)
	CBPM	cbpm_std		Westberry et al. (2008)
Euphotic zone depth				
	VGPM	zeuph_vgpm		Morel et al. (2007)
	CBPM	zeuph_vgpm		Morel et al. (2007)
Iron				
	Soluble iron	HAM_SFE	Time derivative	Hamilton et al. (2019)
	Labile iron	LFE	Time derivative	Myriokefalitakis et al. (2018)

Note. The categories are organized based on predictor type, where universal predictors are used in every Random Forest realization.

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 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 (2022) 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 toward 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 our 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 our approach. Because we limit our final calculations to regions with at least eight 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 our final 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 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 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 Bisson et al. (2021). We do not make any specific attempt to directly propagate errors in the predictors through our reconstructions. Rather, we rely on the standard deviation of the large ensemble (100) of 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 BV 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 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, 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 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.

3. Results and Discussion

3.1. Particle Size Distribution Reconstructions

Figures 3 and 4 show the global reconstructions of PSD BV and slope. Our reconstruction method is able to capture most of the variability of the UVP5 observations, and robustly reproduce the gridded measurements, with

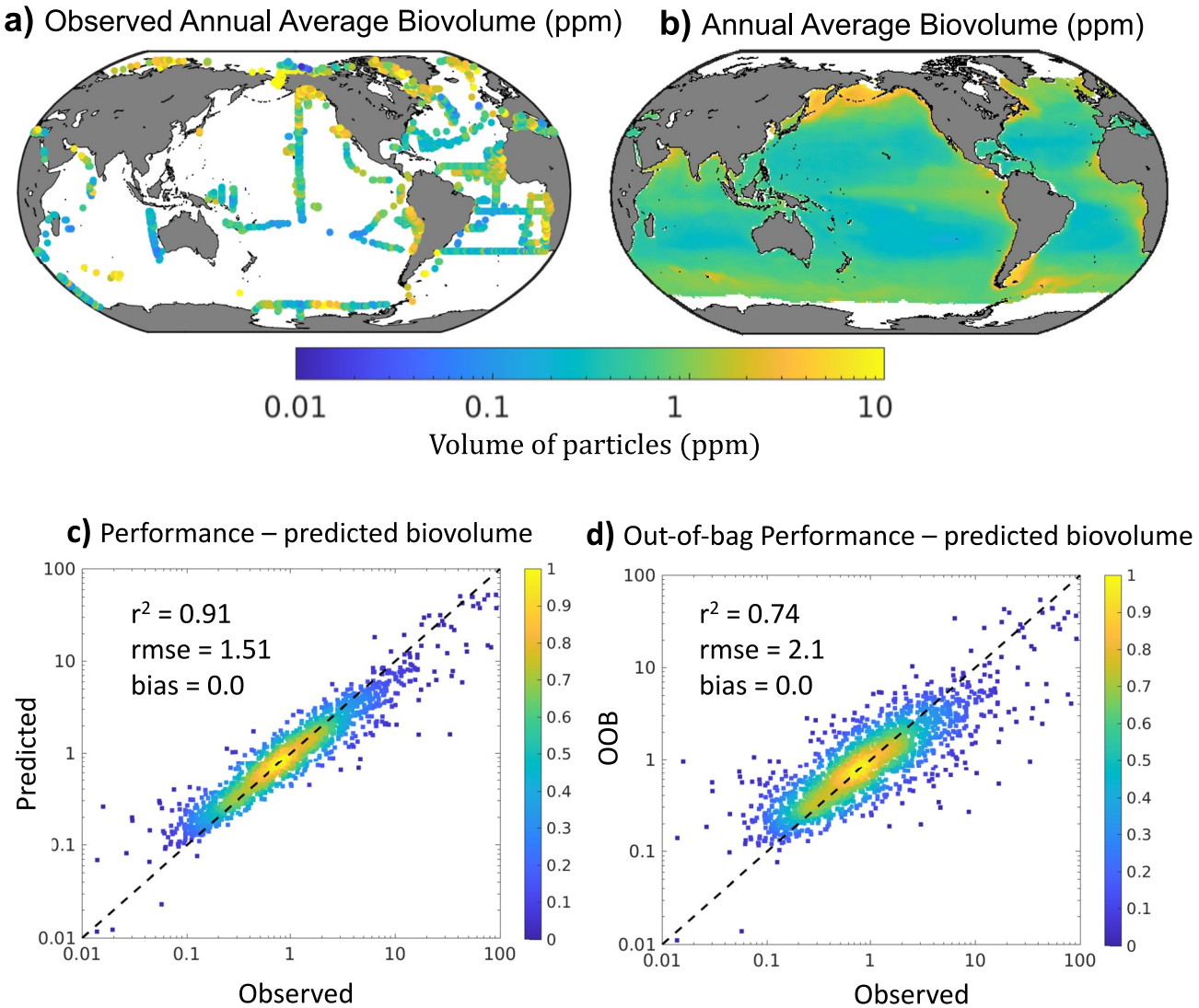

Figure 3. Observed and reconstructed particle biovolume (BV, in parts per million, ppm) at the base of the euphotic zone. (a) Observed average BV. (b) Annual mean BV reconstructions. (c) Performance of the Random Forest reconstruction shown as density scatter plots of predicted versus observed BV (colors indicate the normalized density of observations at each point). Panel (d) same as panel (c), but using out-of-bag (OOB) predictions, that is, predictions versus observations withheld from training. Annotations in panels (b and c) show the coefficient of determination (r^2), the RMSE, and the global bias.

global average values of 0.6 ppm for BV ($r^2 = 0.91$) and 3.9 for slope ($r^2 = 0.86$) when considering the entire data set. Observations that are not used in the training (OOB) provide a more stringent test for the method's robustness. As shown in Figures 3d and 4d, these OOB observations are also robustly predicted, with a RMSE of 2.1 ppm for BV ($r^2 = 0.74$) and 0.33 for slope ($r^2 = 0.68$). Relative to both the full data set and the OOB observations, our reconstructions show negligible biases. That is, there is an overall compensation between data points where our method overestimates observations, and data points where our 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 Figures 3c, 3d, and 4d. Some of this remaining uncertainty could be explained by the episodic nature and patchiness of particle production, transport, and export, and by factors not captured by our climatological predictors. Our 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

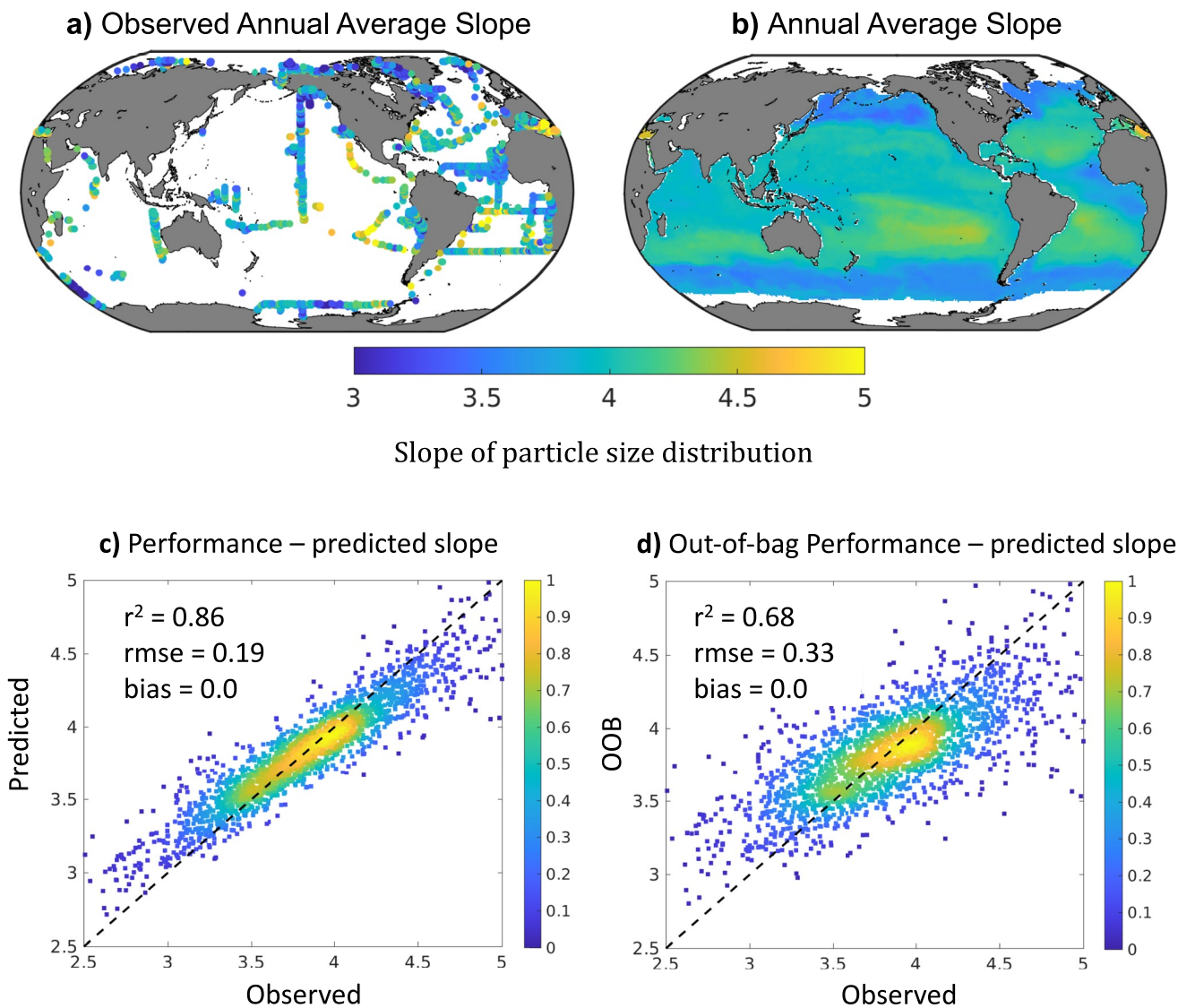

Figure 4. Observed and reconstructed particle size distribution (PSD) slope at the base of the euphotic zone. (a) Observed average PSD slope. (b) Annual mean PSD slope reconstructions. (c) Performance of the Random Forest reconstruction shown as density scatter plots of predicted versus observed particulate slope (colors indicate the normalized density of observations at each point). Panel (d) same as panel (c), but using out-of-bag (OOB) predictions, that is, predictions versus observations withheld from training. Annotations in panels (b and c) show the coefficient of determination (r^2), the RMSE, and the global bias.

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, that is, our reconstructions have a slightly reduced range compared to observations (Figures 3c, 3d, and 4d). 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 BV 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 3.5.

3.2. Global Patterns in Particle Size Distribution

Our reconstructions of the PSD for the time frame 2008 to 2020, reveal high BV in productive regions such as high latitudes, coastal waters, and upwelling systems, and low BV in the oligotrophic subtropical gyres (Figure 3b and Figure S3 in Supporting Information S1). 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

(Figure 4b and Figure S4 in Supporting Information S1), although with somewhat less pronounced variations compared to BV. Consistent with this, we find that slope and BV are negatively correlated ($r^2 = 0.4$, $p < 0.01$ Figures 5a and 5b). Spatial patterns in BV and slope roughly follow the distribution of satellite-derived chlorophyll and primary production estimates, suggesting that phytoplankton and photosynthesis exert a strong control on the total abundance of particles in any given region (Cram et al., 2018; Kostadinov et al., 2009, 2017). Accordingly, we find a positive correlation between BV and surface chlorophyll ($R = 0.49$ for observations, and $R = 0.68$ for reconstructions, both with $p < 0.01$, Figures 5a and 5b) and a negative correlation for slope ($R = -0.18$ for observations, and $R = -0.37$ for reconstructions, both with $p < 0.01$, Figures 5c and 5d).

The negative correlation between particle BV and slope ($R = -0.40, -0.64$ Figures 5e and 5f) indicates that particle-rich regions (higher BV) are also characterized by an excess of large particles over small particles (i.e., flatter slope), relative to average oceanic conditions. Such co-variation between slope and BV 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 BV do not co-vary as closely as expected. For example, in the North Pacific subpolar gyre, flatter slopes are found in the open ocean (Figure 4b), in particular close to the subpolar-subtropical transition, while the highest BVs 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 BVs 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, 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 our 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., 2010). 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.

3.3. Seasonal Variability in Particle Size Distribution

The seasonal dynamics of BV and slope confirms the general anti-correlation of these two variables, and reveals significant seasonal cycles, with maximum BV and minimum slope generally found in spring, and minimum BV and maximum slope in late fall to winter (Figure 6). Similar to the spatial distribution, we find significant deviations from the general anti-correlation between BV and slope. For example, in the North Atlantic, the peak in BV (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 BV 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 BV) 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, and Antarctic Zone), display more marked asymmetries, possibly reflecting seasonal variations in mixing (Behrenfeld & Boss, 2014).

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

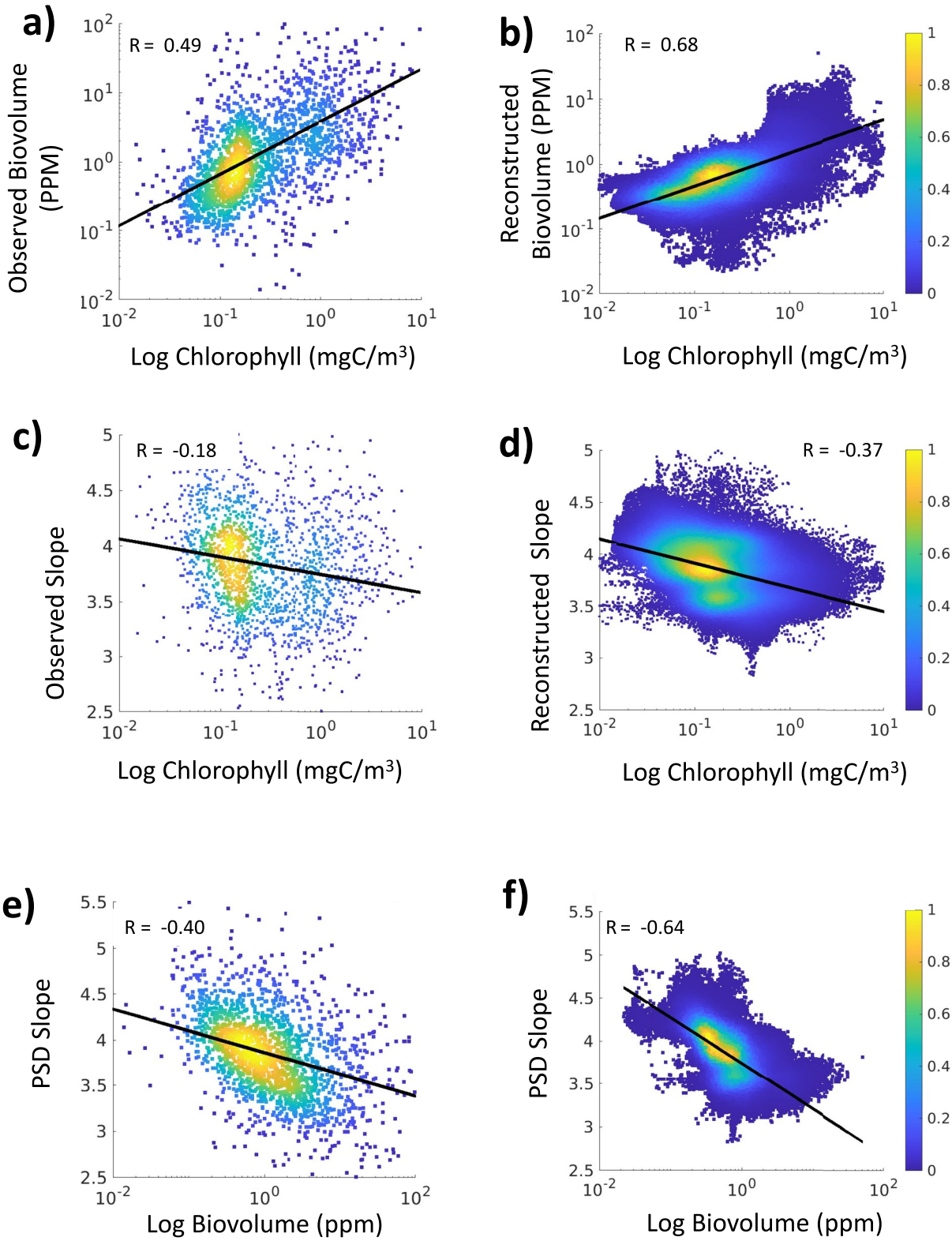

Figure 5. Relationships between particle size distribution (PSD) parameters and surface chlorophyll. (a and b) Relationship between PSD slope and chlorophyll for (a) observed and (b) predicted data. (c and d) Relationship between particle biovolume (BV) and chlorophyll for (c) observed and (d) predicted data. (e and f) Relationships between PSD slope and particle BV, 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.

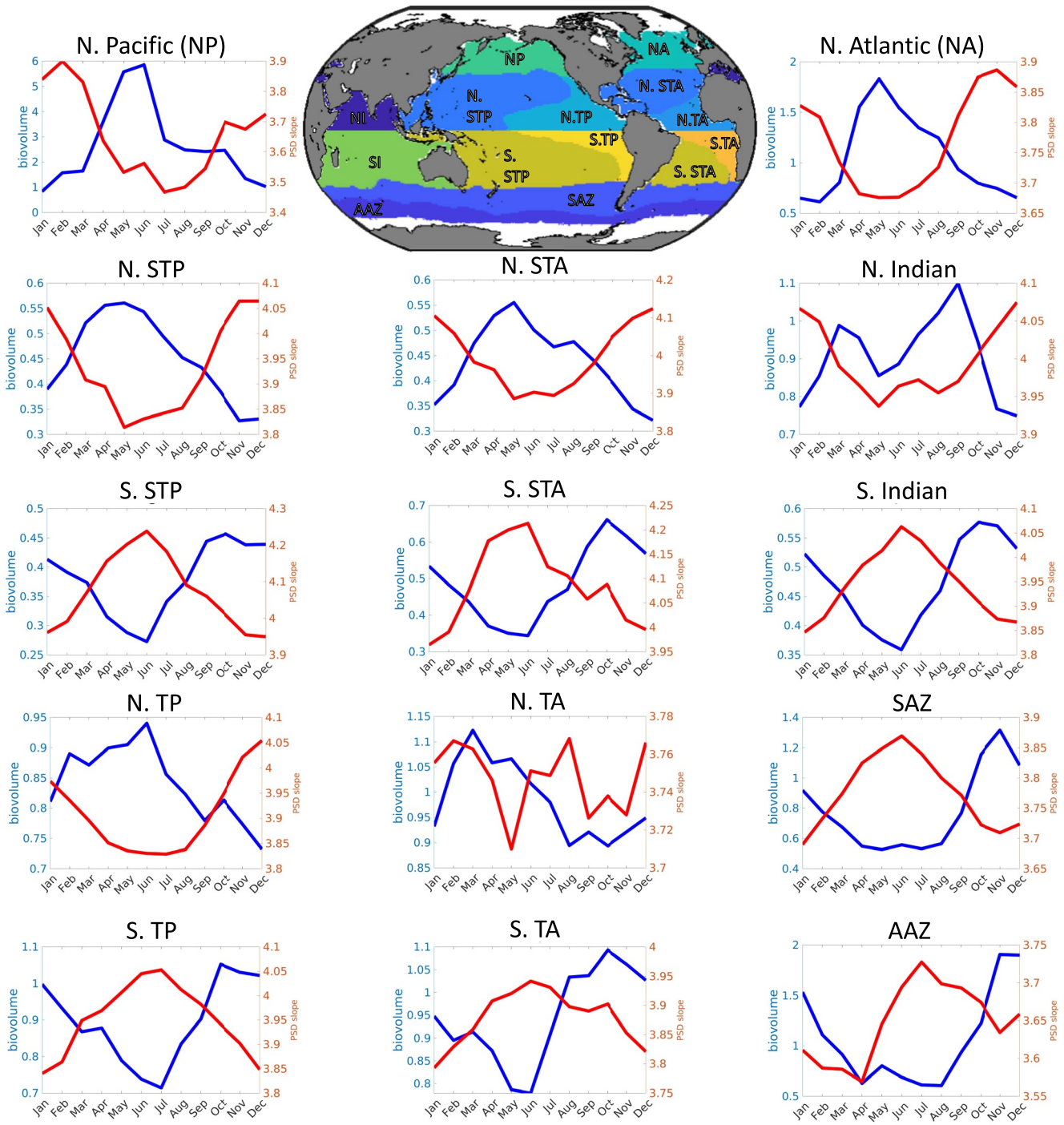

Figure 6. Annual seasonal cycle of particle biovolume (blue lines, in ppm) and slope (red lines) from the Random Forest reconstructions. Each seasonal cycle is from the euphotic zone for the regions specified on the map (top).

co-vary. Likewise, we expect high seasonality in BV 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 (Bisson et al., 2021), that could introduce spurious

patterns in BV and slope. Specific caveats related to the seasonality of our reconstructions are discussed further in Section 3.5.

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 (Figure S2 in Supporting Information S1). Among the important features, we highlight chlorophyll, mixed layer depth, and oxygen, although each has a somewhat different importance for explaining BV 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 BV 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 $p < 0.01$, Figures 5a and 5b). This is not surprising, since chlorophyll is an indicator of phytoplankton, the main source of organic matter and sinking particles in the ocean (Falkowski et al., 1998; 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 $p < 0.01$, Figures 5c and 5d), 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, 2006). 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 BV share several features with estimates of particulate backscattering, and phytoplankton size spectra and composition from observations and models (Barton et al., 2013; Kostadinov et al., 2009; Roy et al., 2013; Ward et al., 2014). Regions with higher BV and flatter slope are dominated by larger phytoplankton, while the subtropics, with lower BV 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 (Burd & Jackson, 2009; Jackson, 1990; Kiørboe et al., 1990). 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 affecting the abundance and size of fecal pellets and other aggregates that are the byproduct of zooplankton feeding (Turner, 2015).

Phytoplankton functional groups (e.g., Mouw et al., 2017) and abundance should be considered as important controlling factors on both BV 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 data sets—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 surface chlorophyll and temperature (Kostadinov et al., 2017; Mouw et al., 2017).

3.5. Caveats to Our 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 our 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 our estimates, and shed additional light on regional PSD 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.

Our 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; Reynolds & Stramski, 2021), 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 (Reynolds et al., 2010; Reynolds et al., 2016). While our tests suggest that a power law assumption is globally robust for the range of sizes sampled by UVP5 instruments (Figure S1 in Supporting Information S1), other statistical models, for example, based on different distributions (Jonasz & Fournier, 1996) or non-parametric descriptors (Reynolds & Stramski, 2021) may be more appropriate, and could result in somewhat different patterns of PSD and BV 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 our 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 (Boss et al., 2015; Karp-Boss et al., 2007; Lombard et al., 2019; Reynolds et al., 2010; Stemmann & Boss, 2012). Lastly, our 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 our reconstructions in coastal regions where inputs of mineral particles from rivers and sediment may be important (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 our 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; 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 (Bisson et al., 2021).

Similar to previous work (DeVries & Weber, 2017; Siegel et al., 2014), limiting our reconstruction to regions with more complete satellite coverage leads to gaps 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 is likely to occur in these regions and times of the year (e.g., Bisson & Cael, 2021; Lowry et al., 2018; 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 RF 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 was excluded from the final reconstructions (Figure S2 in Supporting Information S1). It is possible that our RF 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).

4. Conclusions

In this paper, 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 (Briggs et al., 2020; Burd & Jackson, 2009), interactions with microbial communities (Church et al., 2021; DeLong et al., 1993) 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 our 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 BV and the PSD slope are characterized by similar but inverse patterns, with regions of high particle BV generally characterized by flatter slopes, that is, relatively more abundant large particles. Similarly, the seasonal cycles of the particle slope and BV are inversely correlated over time through most of the ocean. Importantly, because of this anti-correlation, BV and slope variations would act synergistically on sinking particle fluxes, by enhancing them in region of higher BV and flatter slope, and reducing them in regions of low BV and steeper slope. We also show that BV 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 BV 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. Our findings in turn support observational 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.

Acknowledgments
This material is based upon work supported by the U.S. National Science Foundation under Grants No. OCE-1635632 and OCE-1847687. D. Bianchi acknowledges support from the Alfred P. Sloan Foundation, and computational support by the Extreme Science and Engineering Discovery Environment (XSEDE) through allocation TG-OCE17001. A. M. P. McDonnell acknowledges support from the NSF Award No. 1654663. T. Weber was supported by the NSF award OCE-1635414. R. Kiko acknowledges support via the BMBF funded project CUSCO, the EU project TRIATLAS (European Union's Horizon 2020 Program, 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". L. Stemmann was supported in data collection by the Chaire Vision by CNRS and Sorbonne University.

Data Availability Statement

Data generated by this analysis has been uploaded to BCO-DMO, https://doi.org/10.26008/1912/bco-dmo.856942.1. The individual UVP5 profiles used to generate are discussed in a separate manuscript https://doi.org/10.5194/essd-2022-51.

References

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

Antoine, D., & Morel, A. (1996). 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. https://doi.org/10.1029/95GB02831

Bader, H. (1970). The hyperbolic distribution of particle sizes. *Journal of Geophysical Research*, *75*(15), 2822–2830. https://doi.org/10.1029/JC075i015p02822

Barton, A. D., Pershing, A. J., Litchman, E., Record, N. R., Edwards, K. F., Finkel, Z. V., et al. (2013). The biogeography of marine plankton traits. *Ecology Letters*, *16*(4), 522–534. https://doi.org/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*(1), 167–194. https://doi.org/10.1146/annurev-marine-052913-021325

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology & Oceanography*, *42*(1), 1–20. https://doi.org/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*, *11*(4), 1–6. https://doi.org/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. https://doi.org/10.5194/bg-13-3109-2016

Bisson, K., & Cael, B. (2021). How are under ice phytoplankton related to sea ice in the Southern Ocean? *Geophysical Research Letters*, *48*(21), e2021GL095051. https://doi.org/10.1029/2021gl095051

Bisson, K. M., Boss, E., Werdell, P. J., Ibrahim, A., Frouin, R., & Behrenfeld, M. J. (2021). Seasonal bias in global ocean color observations. *Applied Optics*, *60*(23), 6978–6988. https://doi.org/10.1364/AO.426137

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., et al. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, *10*(10), 6225–6245. https://doi.org/10.5194/bg-10-6225-2013

Boss, E., Guidi, L., Richardson, M. J., Stemmann, L., Gardner, W., Bishop, J. K., et al. (2015). Optical techniques for remote and in-situ characterization of particles pertinent to geotraces. *Progress in Oceanography*, *133*, 43–54. https://doi.org/10.1016/j.pocean.2014.09.007

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. https://doi.org/10.1038/s41586-019-1098-2

Briggs, N., Dall'Olmo, G., & Claustre, H. (2020). Major role of particle fragmentation in regulating biological sequestration of CO_2 by the oceans. *Science*, *367*(6479), 791–793. https://doi.org/10.1126/science.aay1790

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

Buesseler, K. O., Lamborg, C. H., Boyd, P. W., Lam, P. J., Trull, T. W., Bidigare, R. R., et al. (2007). Revisiting carbon flux through the ocean's twilight zone. *Science*, *316*(5824), 567–570. https://doi.org/10.1126/science.1137959

Burd, A. B., Hansell, D. A., Steinberg, D. K., Anderson, T. R., Arístegui, J., Baltar, F., et al. (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. https://doi.org/10.1016/j.dsr2.2010.02.022

Burd, A. B., & Jackson, G. A. (2009). Particle aggregation. *Annual Review of Marine Science*, *1*(1), 65–90. https://doi.org/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. https://doi.org/10.1029/2020gl091771

Cavender-Bares, K. K., Rinaldo, A., & Chisholm, S. W. (2001). Microbial size spectra from natural and nutrient enriched ecosystems. *Limnology & Oceanography*, *46*(4), 778–789. https://doi.org/10.4319/lo.2001.46.4.0778

Church, M. J., Kyi, E., Hall, R. O., Jr., 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 & Oceanography*, *66*(9), 3255–3270. https://doi.org/10.1002/lno.11877

Copernicus Climate Change Service. (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

Cram, J. A., Fuchsman, C. A., Duffy, M. E., Pretty, J. L., Lekanoff, R. M., Neibauer, J. A., et al. (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. 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. https://doi.org/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. https://doi.org/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 & Oceanography*, *38*(5), 924–934. https://doi.org/10.4319/lo.1993.38.5.0924

D'Errico, J. (2022). Inpaint nans, MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans

Devries, T., Liang, J. H., & Deutsch, C. (2014). A mechanistic particle flux model applied to the oceanic phosphorus cycle. *Biogeosciences*, *11*(19), 5381–5398. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/s0967-0637(99)00105-3

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. https://doi.org/10.1029/2006GB002907

Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science*, *281*(5374), 200–206. https://doi.org/10.1126/science.281.5374.200

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. https://doi.org/10.5194/bg-9-1301-2012

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

Garcia, H. E., Weathers, K., Paver, C. R., Smolyar, I., Boyer, T. P., Locarnini, R. A., et al. (2019). World Ocean Atlas 2018, volume 3: Dissolved oxygen, apparent oxygen utilization, and oxygen saturation. *NOAA Atlas NESDIS*, *3*(83), 38.

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. https://doi.org/10.5194/bg-3-521-2006

Giering, S. L., Sanders, R., Lampitt, R. S., Anderson, T. R., Tamburini, C., Boutrif, M., et al. (2014). Reconciliation of the carbon budget in the ocean's twilight zone. *Nature*, *507*(7493), 480–483. https://doi.org/10.1038/nature13123

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

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. https://doi.org/10.3354/meps305059

Gregor, L., & Gruber, N. (2021). 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*(2), 777–808. https://doi.org/10.5194/essd-13-777-2021

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470. https://doi.org/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. https://doi.org/10.1016/j.dsr.2008.05.014

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

Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., et al. (2009). Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis. *Limnology & Oceanography*, *54*(6), 1951–1963. https://doi.org/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. https://doi.org/10.5194/bg-18-25-2021

Hamilton, D. S., Scanza, R. A., Feng, Y., Guinness, J., Kok, J. F., Li, L., et al. (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. https://doi.org/10.5194/gmd-12-3835-2019

Huete-Ortega, M., Rodríguez-Ramos, T., López-Sandoval, D., Cermeño, P., Blanco, J. M., Palomino, R. L., et al. (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. https://doi.org/10.3354/meps11007

Jackson, G. A. (1989). Simulation of bacterial attraction and adhesion to falling particles in an aquatic environment. *Limnology & Oceanography*, *34*(3), 514–530. https://doi.org/10.4319/lo.1989.34.3.0514

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. https://doi.org/10.1016/0198-0149(90)90038-w

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*, *117*(4). https://doi.org/10.1029/2011JC007651

Jonasz, M., & Fournier, G. (1996). Approximation of the size distribution of marine particles by a sum of log-normal functions. *Limnology & Oceanography*, *41*(4), 744–754. https://doi.org/10.4319/lo.1996.41.4.0744

Karl, D., Knauer, G., Martin, J., & Ward, B. (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature*, *309*(5963), 54–56. https://doi.org/10.1038/309054a0

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. https://doi.org/10.1038/332438a0

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. https://doi.org/10.4319/lom.2007.5.396

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

Kiko, R., Picheral, M., Antoine, D., Babin, M., Berline, L., Biard, T., et al. (2022). A global marine particle size distribution dataset obtained with the Underwater Vision Profiler 5, *Earth System Science Data Discussions*. [preprint]. https://doi.org/10.5194/essd-2022-51

Kiørboe, T., Andersen, K., & Dam, H. (1990). Coagulation efficiency and aggregate formation in marine phytoplankton. *Marine Biology*, *107*(2), 235–245. https://doi.org/10.1007/bf01319822

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. https://doi.org/10.1016/s0065-2881(08)60129-7

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. https://doi.org/10.1128/aem.68.8.3996-4006.2002

Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J., et al. (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. 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*, *114*(9), 1–22. https://doi.org/10.1029/2009JC005303

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2010). Global variability of phytoplankton functional types from space: Assessment via the particle size distribution. *Biogeosciences*, *7*(10), 3239–3257. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/s0967-0637(99)00032-1

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., et al. (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. https://doi.org/10.5194/bg-17-3439-2020

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. https://doi.org/10.1038/ngeo612

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

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

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, *6*, 196. https://doi.org/10.3389/fmars.2019.00196

Lowry, K. E., Pickart, R. S., Selz, V., Mills, M. M., Pacini, A., Lewis, K. M., et al. (2018). Under-ice phytoplankton blooms inhibited by spring convective mixing in refreezing leads. *Journal of Geophysical Research: Oceans*, *123*(1), 90–109. https://doi.org/10.1002/2016jc012575

McDonnell, A. M., & Buesseler, K. O. (2010). Variability in the average sinking velocity of marine particles. *Limnology & Oceanography*, *55*(5), 2085–2096. https://doi.org/10.4319/lo.2010.55.5.2085

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. https://doi.org/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. https://doi.org/10.1016/j.ecolmodel.2022.109907

Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J., Bricaud, A., et al. (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Frontiers in Marine Science*, *4*, 41. https://doi.org/10.3389/fmars.2017.00041

Myriokefalitakis, S., Ito, A., Kanakidou, M., Nenes, A., Krol, M. C., Mahowald, N. M., et al. (2018). Reviews and syntheses: The GESAMP atmospheric iron deposition model intercomparison study. *Biogeosciences*, *15*(21), 6659–6684. https://doi.org/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. https://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. https://doi.org/10.7289/V5J1012Q

Nguyen, T. T., Zakem, E. J., Ebrahimi, A., Schwartzman, J., Caglar, T., Amarnath, K., et al. (2022). Microbes contribute to setting the ocean carbon flux by altering the fate of sinking particulates. *Nature Communications*, *13*(1), 1–9. https://doi.org/10.1038/s41467-022-29297-2

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. https://doi.org/10.1186/1471-2105-11-110

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. https://doi.org/10.1029/2018gb006145

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. https://doi.org/10.1038/s41598-020-60424-5

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. https://doi.org/10.1364/oe.408439

Picheral, M., Catalano, C., Brousseau, D., Claustre, H., Coppola, L., Leymarie, E., et al. (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. https://doi.org/10.1002/lom3.10475

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. https://doi.org/10.4319/lom.2010.8.462

Reynolds, R. A., & Stramski, D. (2021). 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. 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 & Oceanography*, *61*(5), 1869–1890. https://doi.org/10.1002/lno.10341

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*, *115*(C8), C08024. https://doi.org/10.1029/2009JC005930

Roullier, F., Berline, L., Guidi, L., Durrieu De Madron, X., Picheral, M., Sciandra, A., et al. (2014). Particle size distribution and estimated carbon flux across the Arabian Sea oxygen minimum zone. *Biogeosciences*, *11*(16), 4541–4557. https://doi.org/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. https://doi.org/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. https://doi.org/10.1029/2020jc016218

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

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. https://doi.org/10.1093/plankt/fbq072

Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., et al. (2020). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. *Current Climate Change Reports*, *6*(3), 1–25. https://doi.org/10.1007/s40641-020-00160-0

Sheldon, R., Prakash, A., & Sutcliffe, W. H., Jr. (1972). The size distribution of particles in the ocean 1. *Limnology & Oceanography*, *17*(3), 327–340. https://doi.org/10.4319/lo.1972.17.3.0327

Sheldon, R., Sutcliffe, W., Jr., & 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. https://doi.org/10.1139/f77-314

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. https://doi.org/10.1002/2013GB004743

Siegel, D. A., Cetinić, I., Graff, J. R., Lee, C. M., Nelson, N., Perry, M. J., et al. (2021). An operational overview of the EXport Processes in the Ocean from Remote Sensing (EXPORTS) Northeast Pacific field deployment. *Elementa: Science of the Anthropocene*, *9*(1), 00107. https://doi.org/10.1525/elementa.2020.00107

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. https://doi.org/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. https://doi.org/10.1002/2016GB005521

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 & Oceanography*, *53*(4), 1327–1338. https://doi.org/10.4319/lo.2008.53.4.1327

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. https://doi.org/10.1146/annurev-marine-120710-100853

Stemmann, L., Gorsky, G., Marty, J.-C., Picheral, M., & Miquel, J.-C. (2002). Four-year study of large-particle vertical distribution (0–1000 m) 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. https://doi.org/10.1016/S0967-0645(02)00032-2

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. https://doi.org/10.1016/j.dsr.2004.03.001

Stemmann, L., Youngbluth, M., Robert, K., Hosia, A., Picheral, M., Paterson, H., et al. (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. https://doi.org/10.1093/icesjms/fsn010

Sverdrup, H. (1953). On conditions for the vernal blooming of phytoplankton. *Journal du Conseil/Conseil Permanent International pour l'Exploration de la Mer*, *18*(3), 287–295. https://doi.org/10.1093/icesjms/18.3.287

Turekian, K. K. (1977). The fate of metals in the oceans. *Geochimica et Cosmochimica Acta*, *41*(8), 1139–1144. https://doi.org/10.1016/0016-7037(77)90109-0

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Progress in Oceanography*, *130*, 205–248. https://doi.org/10.1016/j.pocean.2014.08.005

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. https://doi.org/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 of Earth Science*, *8*, 376. https://doi.org/10.3389/feart.2020.00376

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. https://doi.org/10.1029/2007GB003078

Yang, S., Chang, B. X., Warner, M. J., Weber, T. S., Bourbonnais, A. M., Santoro, A. E., et al. (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), 11954–11960. https://doi.org/10.1073/pnas.1921914117

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

Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Antonov, J. I., Locarnini, R. A., et al. (2019). World ocean atlas 2018, volume 2: Salinity. *NOAA Atlas NESDIS*, *2*(82), 50.