UNIVERSITY OF CALIFORNIA

Santa Barbara

Linking global carbon export processes from the ocean's surface into the deep

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Marine Science

by

Kelsey Marie Bisson

Committee in charge:

Professor David Siegel, Chair

Professor Timothy DeVries

Professor Mark Brzezinski

Professor Craig Carlson

December 2018

The dissertation of Kelsey Marie Bisson is approved.

_____

Mark Brzezinski

_____

Timothy DeVries

_____

Craig Carlson

_____

David Siegel, Committee Chair

December 2018

Linking global carbon export processes from the ocean's surface into the deep

Copyright © 2018

by

Kelsey Marie Bisson

ACKNOWLEDGEMENTS

Kelsey Marie Bisson

UC Santa Barbara
Phone: +1 (440) 503-4212
bisson@eri.ucsb.edu

Education	**UC Santa Barbara** Ph.D., Marine Science, expected 2018. Fields: Biological oceanography + biogeochemistry **The Ohio State University** B.S., Geological Sciences, 2013. Mathematics minor, *honors research distinction*, magna cum laude
Dissertation	"Linking global carbon export processes from the ocean's surface into the deep " Committee: David Siegel (chair), Tim DeVries, Mark Brzezinski, Craig Carlson
Awards and Fellowships	National Science Foundation Graduate Research Fellow, 2013-2018 (96K + tuition) Eugene Cota Robles Fellow, UCSB, 2016-2017 (24K) UCSB Graduate Division Dissertation Fellowship, 2018-2019 (8K + tuition) PI grant for ROAM, 2017-2018 (National Academies Keck Futures Initiative, 50K) PI (co), grant for R/V Sally Ride, December 15-23, 2017 (UC Marine Council, 350K) PI, grant for 'Sisters of the Blue' Ocean Science Radioshow, 2017-2018 (Coastal Fund, 5K) Earth Research Institute Travel Grant, 2018,2019 The Oceanography Society Travel Grant, 2018 UCSB Senate Doctoral Travel Grant, 2017 Best Poster Presentation, Natural and Mathematical Sciences, 2012 (OSU) Willis E. Rector Scholarship in Earth Science, 2011-2013 (6K) Shell Oil Company Technical Scholarship, 2011-2013 (20K)

Publications

[8] **Bisson, K.**, D.A. Siegel, and T. DeVries. Assessing mechanistic uncertainties of the biological pump by integrating data from satellites, experiments, and in situ observations, *in prep*.

[7] Cael, B.B. and **Bisson, K.**. Particle flux parameterizations: quantitative and mechanistic similarities and differences. 2018. *Frontiers in Marine Science*,5, 395.

[6] **Bisson, K.**, D. Siegel, T. Devries, B.B. Cael, and K. Buesseler.2018. How data set characteristics influence ocean carbon export models. 2018. *Global Biogeochemical Cycles*.

[5] Cael, B.B., **K. Bisson**, and Chris Follett. 2018. Can rates of ocean primary production and biological carbon export be related through their probability distributions?. 2018. *Global Biogeochemical Cycles*, 32 (6) 954-970.

[4] Cael, B. B., **K. Bisson**, M.J. Follows. How have recent temperature changes affected the efficiency of ocean biological carbon export?. 2017. *Limnology and Oceanography Letters*, 2(4), 113-118.

[3] Cael, B.B., Lambert, Bennett, and **Kelsey Bisson**. Pond Fractals in a Tidal Flat. 2015. *Physical Review E*, 92(5).

[2] **Bisson, Kelsey M.**, K.A. Welch, S.A. Welch, W.B. Lyons, J.M. Sheets, J.S. Levy, A.G. Fountain. Patterns and processes of salt efflorescences in the McMurdo Region, Antarctica. 2015. *Journal of Arctic, Antarctic, and Alpine Research*. 47(3), 407-425.

[1] W.B. Lyons, D.L. Leslie, R.S. Harmon, K. Neumann, K.A. Welch, **K.M. Bisson**, D.M. McKnight. The carbon stable isotope biogeochemistry of streams, Taylor Valley, Antarctica. 2012. *Applied Geochemistry*. 2012, 10.1016/j.apgeochem.2012.08.019.

Training and Expeditions	Co-Chief Scientist on R/V Sally Ride, Dec 15-23, 2017 NASA EXPORTS field campaign, 5 weeks aboard R/V Sally Ride, Summer 2018 NASA/U of Maine Ocean Optics Class, Darling Marine Center, 2015 CLIVAR P16N Cruise on NOAA R/V Ronald H. Brown, 5 weeks, June -July 2015 Moorea Coral Reef-LTER cruise aboard R/V Kilo Moana, 3 weeks, August 2014 Plumes and Blooms (NASA) aboard NOAA R/V Shearwater, monthly, 2013-present McMurdo Dry Valleys Long Term Ecological Research, Antarctica, 2012-13 Geological field mapping, Ephraim, Utah, 2011
Workshop Participation	National Academies Keck Future Initiative (NAKFI) 'Beyond Boundaries,' Irvine, 2017 Trait-Based approaches to Ocean Life Meeting, Bergen, Norway, 2017 (NAKFI) 'Discovering the Deep Blue Sea,' Irvine, 2016 Trait-Based approaches to Ocean Life Meeting, Waterville Valley, New Hampshire, 2015
Teaching	**Geography Department, UC Santa Barbara** Teaching Assistant, Remote Sensing of the Environment, 2016 Teaching Assistant and Guest lecturer, Ocean waves and tides, 2017 Guest lecturer, Introduction to Marine Science, 2016, 2017
Community Outreach + Service	**LASER** (Leonardo Art and Science Evening Rendezvous) at UCSB, co-planner, 2017 **Aquatic Cathartic**, community event, UCSB, Host UCSB Women in Stem Mentorship Program, mentor, 2016 **STEAMroom, Art + Ocean Science Show**, Red Barn Gallery, Curator UCSB Marine Science Spring Speaker Series, Seminar Organizer, 2016 Women in Geographical Science (UCSB), founding board member **Coastal Fund** Board of Directors, Vice Chair, 2013-2015 **Girls in Ocean Science Conference**, Dana Point, California. Invited Speaker. Led several 30 min satellite science workshops. **Ocean GEMS** (Go Explore Marine Science) Featured Mentor. Planned and hosted **Demystifying Grad School** event at UCSB. **Women in Math and Science Club** (OSU) President, 2009-2013. **Outreach to local schools**, Conducted presentations for Garden Street Academy, Anacapa School, Tremont Elementary School, MESA day, and Indianola Alternative K-8 school to present on topics related to ocean health, fresh and salty water bodies, and the broader natural sciences.
Media	3D fLUX data visualization in collaboration with Mark Hirsch displayed at UCSB's *Allosphere*, Spring 2017 KCSB 91.9 FM Sisters of the Blue Ocean Science radio show, co-host and producer, 2017-2018 'Contrast and Cadence' authored article in LTER blog, https://ssalter.blog/2017 *Master Variable*, A video installation showing in situ temperature products over the last 3 decades. Shown at STEAMroom UCSB, Oct 2016. Antarctic Blog, "Going the distance for the data" b509m.wordpress.com

Presentations & Posters

[2018] **Bisson, Kelsey**, Sasha Kramer, Alexis Fischer, Dylan Catlett, James Allen, and David Siegel. 'Spatial patterns and optical analysis of wildfire-derived ash in the Santa Barbara Channel.' Ocean Optics Meeting, Dubrovnik, Croatia. (poster)

[2015 - 2018] **Bisson, Kelsey**, Annual presentation to all students and faculty of the Marine Science Program, Seminar Series, UCSB. (talk)

[2018] **Bisson, Kelsey** and Eleanor Arrington. 'Oceanography in the Thomas Fire: Preliminary Results from a Graduate Student Led Expedition.' UCSB Library Lecture series, April 17, 2018. Pacific View Room. (talk)

[2018] **Bisson, Kelsey** and David Siegel. 'Mechanistic modeling of the biological pump by integrating satellite and in situ observations.' BN14D: Multiscale Interdisciplinary Observations of Marine Ecosystems and the Biological Pump Posters. Ocean Sciences Meeting, Portland, Oregon. (poster)

[2018] **Bisson, Kelsey**, Sasha Kramer, Gad Girling, and Sarah Amiri. 'Ride Sally Ride: A student-led approach to ocean science education aboard R/V Sally Ride.' ED013: 'Ship-to-Shore': Ocean Sciences in a Changing World. Ocean Sciences Meeting, Portland, Oregon. (poster)

[2018] **Bisson, Kelsey** and David Siegel. 'Towards a consensus of biological carbon export from the euphotic zone using ecological modeling, remote sensing, and in situ data analysis.' Trait Based Approaches to Ocean Life Workshop. Bergen, Norway. (poster).

[2016] **Bisson, Kelsey** and David Siegel. 'Satellite derived estimates of global export flux composition and attenuation in the mesopelagic.' Ocean Sciences Meeting. New Orleans, Louisiana. (poster)

[2016] **Bisson, Kelsey** and David Siegel. 'Linking shifts in remotely sensed planktonic community structure to changes in carbon export flux from the surface ocean to the mesopelagic.' Trait-Based approaches to Ocean Life Meeting, Waterville Valley, New Hampshire. (poster)

[2013] **Bisson, Kelsey** 'Distribution of cold desert salts in the McMurdo Region, Antarctica'. University of California Santa Barbara, Biogeochemical Brown Bag Seminar. (talk)

[2013] **Bisson, Kelsey** 'Reflections on a rookie season in Antarctica'. Byrd Polar Research Center, Ohio. (talk)

[2013] **Bisson, Kelsey** 'The Bossier Carbonate story: An analysis of composition, genesis, and anticipated significance.' Presented to Magnolia Field Development and Planning Team. Shell Oil Company, Texas. (talk)

[2013] **Bisson, Kelsey**, K.A. Welch, S.A. Welch, W.B. Lyons, J.M. Sheets, J.S. Levy, A.G. Fountain. 'Description and distribution of cold desert salts in the McMurdo region, Antarctica.' Geological Society of America National Meeting. Denver, Colorado. (poster)

[2012] **Bisson, Kelsey**, K.A. Welch, and W.B. Lyons. Temporal Variability of Dissolved Inorganic Carbon in Streams at the McMurdo Dry Valleys, Antarctica. Natural and Mathematical Sciences Forum, Ohio State University. Winner, 'Best Poster Presentation.' (poster)

ABSTRACT

Linking global carbon export processes from the ocean's surface into the deep

by

Kelsey Marie Bisson

The ocean contains reservoirs of carbon that influence atmospheric CO_2, impact climate, fuel deep-sea metabolism, and sustain fisheries worldwide (Volk and Hoffert, 1985; Doney et al., 2011). Despite the ocean's substantial role in the global carbon cycle, much remains to be known about the fundamental influences for carbon transport over climatically relevant scales. These details remain elusive because 1) available data is sparse, 2) the collection of new data is expensive, time consuming, and logistically difficult, and 3) the system itself is complex and highly variable. Overall, this dissertation confronts sources of uncertainty for data and models associated with differences in technology, incongruous spatiotemporal sampling schemes, and model complexity in order to extract as much as possible given what is currently available to use.

The chapter specific objectives are as follows:

(1) Examine how intrinsic and extrinsic sources of variability affect model calibration
through a compilation of data sets spanning method, space and time

(2) Assess mechanisms of the biological pump given available data products, and diagnose mechanistic parameter sensitivities as they relate to experimental/observational work

(3) Test and recalibrate parameters to predict the flux of carbon from a rapidly growing technology (Underwater Vision Profiler data) from the surface to depth.

In chapter 1 carbon export predictions from a mechanistic model are compared with observations of POC fluxes from several datasets compiled from the literature spanning different space, time, and depth scales as well as using different observational methodologies. Model parameters are optimized to provide the best match between model-predicted and observed POC fluxes, explicitly accounting for sources of error associated with each dataset. Model-predicted globally integrated values of POC flux at the base of the euphotic layer range from 3.8 to 5.5 Pg C yr^{-1}, depending on the dataset used to optimize the model. Results from chapter 2 reveal that modeled carbon export pathways also vary depending on the dataset used to optimize the model, and depending on the processes explicitly represented within a quite of nested models. Chapter 3 tests the relationship between the particle size distribution and flux in the global ocean using abundant observations of particle size spectra in tandem with high-resolution model outputs. Results motivate the need for improved parameterizations that link particle size to flux based on depth and/or surface ecosystem characteristics. Altogether these findings highlight the importance of collecting field data that average over the substantial natural temporal and spatial variability in carbon export fluxes, and of advancing satellite algorithms for ocean NPP, in order to improve predictions of biological carbon export.

x

TABLE OF CONTENTS

$+$

I. How data set characteristics influence ocean carbon export models

1 Introduction

Earth's oceans contain diverse planktic ecosystems that influence the global carbon cycle, and ultimately global climate, through processes that transport organic carbon to the deep ocean after it is fixed through photosynthesis in the surface (Volk and Hoffert, 1985), commonly called the 'biological pump.' The exported carbon feeds mesopelagic communities and is sequestered from the atmosphere on timescales ranging from months to centuries (DeVries et al., 2012). Of the many pathways for carbon export, the sinking flux of particulate organic carbon (POC) is especially complex and highly variable. Locally, interannual monthly variations in sinking particle export fluxes can be as large as the long-term monthly mean (Fig S1). The same is true for multiple, near-simultaneous flux measurements over large regions (>100 kilometers wide, Buesseler et al., 2007; Estapa et al., 2015; Black et al., 2018), highlighting large spatial variability in POC export fluxes. The high degree of variability in POC export fluxes stands as a significant barrier to constraining data-driven global-scale models of particulate carbon export (which are typically formulated as time-averaged models), and may contribute to the wide range of estimates of globally-integrated POC export which vary from roughly 5 to 12 Pg C yr^{-1} (Dunne et al., 2005; Boyd and Trull 2007; Henson et al., 2011; Siegel et al., 2014).

The aforementioned variability of POC fluxes is driven by both intrinsic and extrinsic factors. We define intrinsic factors as those driven by ecosystem process, while extrinsic factors as those that are introduced through sampling and data processing. Intrinsic variability of export will be

driven by variability in nutrient supply, phytoplankton growth, predator behavior and aggregate formation that creates spatial heterogeneity and temporal intermittency in the POC flux (e.g., Abraham, 1998; Karl et al. 2003; Zehr et al. 2017). POC export is also influenced by particle size, density, morphology, lability, ecological interactions, and physical factors, all of which exhibit substantial spatial and temporal heterogeneity (Alldredge and Silver, 1988; Burd and Jackson, 2009; Armstrong et al., 2001; Steinberg et al., 2000; Mahadevan et al., 2012; Cram et al., 2018). The time lag between production of POC in the surface ocean and its subsequent export to depth (e.g. Karl et al. 2003; Estapa et al. 2015) introduces an additional source of variability to POC fluxes.

POC fluxes are also observed in ways that introduce extrinsic variability. Shallow (<200m) sediment traps, the only direct measurement of POC flux, have known collection biases attributed to the hydrodynamics of flows within the trap itself (Gardner 1985), motions of the trap relative to its surrounding waters (Valdes and Price, 2000), vibrations of surface moorings (Gust et al., 1994), zooplankton 'swimmers' that enter the trap (Knauer et al., 1979; Michaels et al., 1990), and the solubilization of material once collected in the trap (Gardner et al., 1983; Antia, 2005). These factors can result in both under- and over-sampling biases, depending on the specific trap design and the environment in which it is deployed (see Buesseler et al. 2007a, and refs therein). The interpretation of sediment trap data is further confounded by advection, because material collected in the trap may have been transported laterally several 10's to 100's of kilometers before being deposited in the trap (Siegel et al. 2008). Several of these trap issues can be overcome by using untethered, or neutrally buoyant sediment traps (NBSTs) instead of surface-tethered drifting tethered traps or bottom-moored traps (e.g., Marsay et al., 2015; Buesseler et al., 2007), but NBSTs have not yet been deployed across a global scale. Finally, the depth at which sediment traps are deployed often introduces another source of variability, since POC flux measurements must be normalized to a common reference depth for comparison and the calibration of carbon export models. Normalization to a reference depth introduces significant errors since attenuation of the POC flux is

most rapid just below the base of the euphotic layer, Z_{eu} (Buesseler and Boyd, 2009). Here we

choose Z_{eu} as a reference depth

because we are interested in assessing the role that surface ecosystem structure and function has on

POC export fluxes; we discuss the influence of reference depth on our results in section 4.3.

In addition to direct measurement of POC fluxes using sediment traps, POC fluxes can be

estimated by measuring the depletion of particle-reactive [234]Th from its long-lived parent, [238]U in the

water column. POC fluxes are calculated by multiplying the measured [234]Th flux by the measured

POC: [234]Th ratio of sinking particles. This calculation has sources of uncertainty related to the

conversion of [234]Th flux to C flux via the particulate [234]Th:C ratio, and other model assumptions

used to compute [234]Th flux (e.g., the neglect of physical transport and the assumption of steady-state

in the Th isotope system; Buesseler et al., 2008). However, one advantage of this method is that

samples can be collected at high vertical resolution, from which estimates of POC flux at or very

near Z_{eu} can be obtained, eliminating the need to normalize observations to a common reference

depth.

These intrinsic and extrinsic factors that drive variability in observed POC fluxes pose a

problem for model validation. We hypothesize that a single POC flux measurement is at best a

snapshot of an instance of the ecosystem and will not capture the mean state needed to calibrate

global-scale models. Thus an individual POC flux measurement can be thought of as a random draw

from a highly variable distribution (Fig. 1, left inset). However, large-scale satellite and numerical

models aim to predict the mean of this distribution. When assemblages of random samples are

averaged over a sufficient number of observations, these averages approach the appropriate mean for

the ecosystem, and can be meaningfully compared to the output of climatological models (Fig. 1,

right inset). This simple illustration emphasizes the importance of assembling model calibration data

in such a way so as to capture the scales of the system that are represented by models. Of course, the

sampling choices made by the oceanographic community are not random draws at different places and times around the world, but rather are motivated by special interest and/or convenience. This introduces systematic spatiotemporal biases that also contribute to mismatches between models and data (the spatial bias is illustrated in Figure 2).

Despite the scale mismatches between POC flux observations and models, numerous studies have used POC flux observations to constrain models that represent mean states. Global satellite-based carbon export models have been developed from empirical relationships with environmental variables, or by tuning mechanistic models to fit POC flux data (e.g., Laws et al, 2000; Henson et al., 2011, 2012; Dunne et al., 2005; Siegel et al., 2014; Li and Cassar, 2016). One way to reduce the spatiotemporal discrepancy between flux observations and models is to account for known lag times between production and export (as in Henson et al., 2015; Giering et al., 2017); however this requires net primary production (NPP) products of sufficient resolution and additional assumptions about the timescale of particle settling. Though the formulae for the aforementioned models vary, all of them are constructed using satellite estimates of NPP. For this study we choose a model that builds on previous work by routing NPP through a two-size class food web model to determine POC flux (Siegel et al., 2014; illustrated in Fig. 3).

Although model fidelity is generally determined by how well the models predict the observations, the degree to which these observations represent a scale appropriate for comparison to climatological models has been little addressed. Here we address this issue by comparing modeled POC export fluxes using a suite of POC export data compilations with varying degrees of intrinsic and extrinsic variability. We use this model-data comparison to investigate the following questions:

- How do inconsistencies between the scale of POC flux observations and POC export models affect our ability to optimize these models using field observations?

- How do uncertainties in satellite NPP affect model-predicted POC export fluxes?

    The paper is organized as follows: Section 2 describes the carbon export model, the satellite

data inputs needed to drive the model, the datasets that are used to optimize the model, and our optimization procedure. Section 3 presents the optimized parameter values, the model outcomes, and the summary statistics from the model optimizations with each dataset. Section 4 discusses how different satellite NPP products influence the results, as well as implications for global carbon export, the importance of the depth horizon used to calculate flux, and the ideal sampling strategy needed for model development.

2 Data and Methods

2.1 Satellite-driven food web model

We use the Siegel et al. (2014) carbon export model (hereafter denoted as S14), which routes net primary production (NPP, mg C m^{-2} d^{-1}) through a simple two-size-class food web model to calculate particulate organic carbon (POC) flux at Z_{eu} (Fig. 3). The model includes two pathways for sinking particle export (TOT_{EZ}) from the euphotic layer: the sinking of phytoplankton algal aggregates (ALG_{EZ}) and the export of zooplankton fecal pellets (FEC_{EZ}):

$$TOT_{EZ} = ALG_{EZ} + FEC_{EZ}. \tag{1}$$

All terms in equation (1) are in mg C m^{-2} d^{-1}.

Estimates of ALG_{EZ} are modeled as a fraction (f_{Alg}) of the microplankton size class of phytoplankton NPP (NPP_M, following energy transfer efficiencies in Michaels and Silver, 1988; Boyd and Stevens, 2002):

$$ALG_{GZ} = f_{Alg} + NPP_M. \tag{2}$$

The baseline (S14) value for f_{alg} is 0.1.

Fecal fluxes (FEC_{GZ}) are calculated by multiplying the phytoplankton herbivory rates by

fixed egestion efficiency parameters based on size (f_{fecM} and f_{fecS}, where the size classes are small

(S = pico + nano, 0.2 -20 μm) and large (M = micro, 20 -50 μm) phytoplankton), and the S14

parameter values for f_{fecM} and f_{fecS} are 0.3 and 0.1, respectively, following work by Michaels and

Silver (1988) and Boyd and Stevens (2002). The higher efficiency of export production for

herbivory on large phytoplankton compared with herbivory on small phytoplankton accounts for the

fewer trophic steps:

$$FEC_{GZ} = [(f_{fecM} + G_M) + (f_{fecS} + G_S)] + Z_{eu}. \tag{3}$$

Herbivory rates G_i for both phytoplankton size classes (P_i; where i refers to the two size

classes) are calculated by mass balance that attributes the change in phytoplankton biomass (dP_i/dt,

as calculated from satellite-derived particulate backscatter using the algorithms of Kostadinov et al.,

2009 and Graff et al., 2015)) to gains from net primary production (NPP_i) and losses to herbivory

(G_i), non-grazing biological mortality (given a specific rate m_{ph} which is assumed by S14 to be 0.1

d[-1]), and reduction of biomass as the mixed layer (Z_{ml}) deepens (i.e., when $\frac{dZ_{ml}}{dt} > 0$).

$$\frac{dP}{dt} = \frac{NPP_i}{Z_{eu}} - G_i - m_{ph}P_i - \delta_{i,M}\frac{Alg_z}{Z_{ml}} - \frac{P_i}{Z_{ml}}\frac{dZ_{ml}}{dt} \,]] \left(\frac{dZ_{ml}}{dt}\right). \tag{4}$$

$H(x) = 1$ if x > 0 and 0 otherwise and $\delta_{i,M} = 1$ when $i = M$ and 0 otherwise so that when Z_{ml}

shoals or is unchanged, there is no loss of biomass.

Summarily, the model routes satellite-derived NPP through a food web model with four ecologically-relevant parameters ($f_{Alg}, f_{recir}, f_{fecs}, m_{ph}$) to calculate the total POC flux exiting the euphotic layer. We note three model assumptions:

1) Satellite estimates of biomass and NPP are uniform over the mixed layer, and the algal flux impacts on large phytoplankton are uniform over the mixed layer.
2) Phytoplankton biomass below the mixed layer is negligible.
3) The amount of large and small NPP is equivalent to their fractional contributions to planktic biovolumes.

The limitations and implications of these, and of other implicit assumptions (i.e., the satellite data products are of suitable quality), are reviewed in the Discussion.

2.2 Model input datasets

The model is driven by satellite data products from the Sea-viewing Wide-Field-of-view Sensor (SeaWiFS) mission (1997-2010) following data processing and calculation details described

in S14. The specific satellite-derived data products used are net primary production (NPP), the slope of an assumed power law particle size distribution, surface chlorophyll concentrations, and particulate backscatter at 443nm. The relationship between particulate backscatter at 443 nm and total phytoplankton carbon is updated from Siegel et al. (2014) using the relationship from Graff et al. (2015). The euphotic depth (Z_{eu}) is defined here as the depth of the 1% of the surface photosynthetically available radiation (PAR) isolume, and is modeled as a function of surface chlorophyll concentration (Morel et al., 2007). We assume that the uncertainty about (Z_{eu}) is smaller than the others. The input data climatologies span the period 1997-2008 at monthly temporal resolution on a one-degree spatial grid.

We test three different NPP inputs (Carbon-based Productivity Model, CbPMv2, Westberry et al., 2008; Vertically Generalized Productivity Model, VGPM, Behrenfeld and Falkowski, 1997; and the Carbon, Absorption, and Fluorescence Euphotic-resolving model, CAFE, Silsbe et al., 2016) in the export model to test how the optimal model parameters vary with the NPP product used. We select the CbPM for its mechanistic construction, the VGPM for its simple structure and wide use in the field, and the CAFE for its improved validation compared to previous models. Spatial maps of annually averaged NPP estimates for the three satellite data products and their differences are included in the supplement (Fig. S2). Though all three NPP models have similar globally integrated mean values (CbPM: 52 Pg C yr^{-1}, VGPM: 51 Pg C yr^{-1}; CAFE: 52 Pg C yr^{-1}), their NPP predictions vary widely on regional and seasonal scales. The CAFE model predicts higher NPP in tropical and subtropical environments relative to the VGPM and CbPM, while exhibiting especially different seasonality in the South Pacific Gyre (see Silsbe et al., 2016 for additional information). Compared to the VGPM, the CbPM predicts increased NPP in oligotrophic environments and reduced NPP at high latitudes. In many instances this difference exceeds a factor of three (Figure S2). Implications of the choice of NPP product are discussed in Section 4.1.

2.3 Synthesis of export flux observations

The model is fit to 5 different POC flux datasets representing various sampling methodologies and various spatiotemporal scales to assess how the type and scale of observations affects model calibration. First we compile from the literature a global dataset of export flux observations from sediment traps and the [234]Th method ('All Data'). This dataset is then subsampled to create the four other datasets: an extension of the data presented in Buesseler and Boyd (2009) of well-sampled (multi-station averages) [234]Th -estimated fluxes at Z_{eu} ('Climatological Data'), a

dataset consisting of [234]Th profiles at from transect cruises ('[234]Th Profiles'), and datasets from sediment traps at the HOT and BATS time-series sites ('HOT' and 'BATS'). Because much of the field data were not sampled within the time period when satellite data are available, monthly climatologies are used.

2.3.1 Global Flux dataset, 'All Data'

Existing databases, papers, and data from personal communication are assembled into a global dataset of shallow (<200m) POC flux from sediment traps and from the [234]Th technique, totaling 1719 observations (available in the supplementary section).

To make the data comparable to the model output, we apply the following analysis steps:

1) eliminate all data observations poleward of 65° latitude, where satellite NPP observations (which are used to drive the carbon export model) are confounded by cloud cover and

limited daylight;

2) choose only observations that are in sufficiently deep water (> 500m) as to not be affected by terrestrial input;

3) bin all data into 1-degree latitude-longitude bins, the spatial resolution of the S14 model;

4) normalize the POC flux measurements to the base of the climatologically-averaged monthly Z_{eu} for each site by applying a power law for the POC attenuation with depth below Z_{eu} (following Martin et al., 1987; Primeau 2006 where 'b' = 0.7; Marsay et al., 2015); and

5) eliminate outliers by excluding data that lie 1.5 interquartile ranges (IQRs) above the 75th percentile or 1.5 IQRs below the 25th percentile of the log-transformed POC flux. This excludes a few rare high flux events (> 600 mg C m^{-2} d^{-1}).

Steps 1, 2, and 5 above reduce the total number of data points from 1721 to 1343 (Fig. 2). Step 3 further organizes the 1343 data points into 591 unique bins (Fig. 2). This dataset has the largest global coverage but also the most uncertainty associated with it (see Table 1).

The 'All Data' dataset has greater spatial coverage than those presented in Mouw et al. (2016) and Henson et al. (2011). 'All Data' is smaller in total number of observations compared to the dataset presented in Mouw et al. (2016) because it does not include deep traps or those at high latitudes, but it has greater spatial coverage because it includes recently published GEOTRACES [234]Th observations. The 'All Data' dataset also differs from that used in Henson et al. (2011) because the Henson et al. (2011) dataset exclusively comprises [234]Th-estimated POC fluxes normalized to a nominal depth (100m) and only includes data up to the year 2009.

2.3.2 'Climatological Data'

The 'Climatological Data' set was curated to be consistent with model assumptions. 'Climatological Data' consists of [234]Th-derived POC flux averages from multiple sites consistently

analyzed for POC flux at the base of the euphotic layer (see supplementary material for compilation). We require at least 4 observations per lat/lon/month bin at the euphotic layer depth. This dataset eliminates (or substantially reduces) sources of natural (intrinsic) variability in the POC flux measurements, thereby matching the scales of the climatological model (Table 1). This dataset expands on the dataset presented in Table 1 of Buesseler and Boyd 2009 (BB09) to include 165 total observations from 14 sites (Fig. S3). We eliminated the two Ocean Station Papa observations from the original BB09 dataset because each of those were single observations, as well as the Southern Ocean observations because there is not sufficient satellite data to evaluate the model at those latitudes. Observations from Estapa et al. (2015), which consists of multiple (>8) flux measurements along ~30 km transects, were added because they were thought to be sufficient to reduce the natural, intrinsic sources of variability.

2.3.3 '234Th Profiles'

We also compile a dataset consisting of [234]Th profiles with sufficient vertical sampling resolution in the upper 200m in order to recalculate fluxes at specific depths.   This '234Th Profiles' subset uses [234]Th-derived POC flux profiles that were collected along transects as part of the US and Dutch GEOTRACES programs (Owens et al., 2015; Black et al., 2018) in addition to a regional study (Estapa et al., 2015), which estimated export flux values at Z_{eu} (Fig. S3, supplementary material). These studies have additional data points above and below Z_{eu} with which to calculate fluxes at varying depths, and we do so to assess the errors related to normalizing the POC fluxes to a specific depth horizon, and to evaluate how model performance varies across data with a dynamic range from one measurement technique. However, all other sources of uncertainty, including the intrinsic uncertainties associated with natural variability, affect this dataset (Table 1). There are 91

unique lat/lon/month bins totaling 217 measurements. The mean number of observations per bin is 2, but the majority of bins consist of a single observation of flux ('[234]Th profiles' is available in the supplementary material).

2.3.4 'HOT' & 'BATS'

The Hawaii Ocean Time Series (HOT) and Bermuda Atlantic Time Series (BATS) are two locations that provide sufficient long-term observations of upper ocean POC flux from which a monthly climatology of POC export can be constructed. This makes these locations ideal for comparing with the climatological export model, and thus we create a third and fourth subdataset from 'All Data' for the climatological POC export at HOT and BATS. The HOT dataset is from January 1998 to October 2010, totaling 115 measurements (Fig. S1), and the BATS dataset is from October 1997 to November 2010, totaling 135 measurements (Fig. S1). All traps are cylindrical particle interceptor traps deployed at 150m for 3 days (Steinberg et al., 2001; Karl and Lukas, 1996). These data are normalized to the depth of the euphotic layer, as is done in the global dataset, and monthly climatologies of POC export are calculated spanning the time periods above. 'HOT' and 'BATS' are climatologically averaged fluxes of localities at 150m, rather than regional averages of flux at Z_{eu}, which makes them inappropriate for inclusion in the 'Climatological Data' set.

2.3.5 Uncertainty estimates for data and model

Each dataset has associated uncertainties due to a combination of intrinsic variability from a varying ecosystem and extrinsic uncertainty due to instrument and sampling biases, and the

normalization of data products to Z_{eu}. Additionally, the carbon export model has a substantial

uncertainty due to uncertainty in the satellite-based NPP product used to drive the model. Table 1

quantifies these uncertainties, where uncertainty values were calculated as follows:

1. The temporal uncertainty estimate of 43% is the average ratio of standard deviation (s) to mean

    monthly flux value ($\bar{x}_i$) at HOT and BATS, where i is the number of observations and m is the

    month, or $uncertainty_{time} = mean\left(\frac{s_m}{\bar{x}_{i,m}}\right)$. The conservative estimate is reported (BATS had

    43% uncertainty, HOTS had 30% uncertainty) (Fig. S1). This approach is an imperfect way of

    addressing temporal error because it generalizes the variability at BATS and HOT to other

    locations; however, it is nonetheless provides a reasonable error for the 'low frequency' (~1

    month[-1]) flux data worldwide.

2. The spatial error estimate of 63% was calculated from the Estapa et al. (2015) dataset, which is

    composed of ~10 samples per 1° grid. Similar to the temporal estimate, the standard deviation

    was divided by the flux for each grid and the overall average is reported. We report these

    spatiotemporal estimates for uncertainty as 40% and 60% (table 1), because both of them carry

    error from spatiotemporal bias that cannot be quantified given the currently available data

    products.

3. The normalization to Z_{eu} uncertainty is calculated for each data point in the global dataset as

    well as in the time series datasets because the data were not collected at the base of Z_{eu}. For

    each data point, 1000 random sample 'b' values were drawn from a normal distribution with

    mean 0.7 and standard deviation 0.35 (to represent a range of Martin 'b' values, Martin et al.,

    1987). Then 1000 flux values were calculated for each data point to represent the range of flux

    attenuation with depth scenarios. The ratio of standard deviation relative to the mean is the

    reported error ($error_z = {}^{\sigma}/_{\mu}$). For observations close to Z_{eu}, the error is minimal, but for

    observations sampled well beneath Z_{eu}, the error is larger (up to 60%, see Fig. S4). Our reported

uncertainty is a lower bound on the normalization uncertainty because there is additional certainty introduced with the formula choice for flux attenuation (e.g., exponential, rational, ballast). We choose the simplest and most commonly used parameterization for consistency with previous work.

4. There are many other sources of uncertainty that fall under instrumental/sampling bias, including the assumed $C/^{234}Th$ values used in the ^{234}Th method, and differences in trap designs for sediment trap observations. These errors are at least 30% (Buesseler et al., 2007), and can be much larger depending on when and where a sample is collected. As stated in the introduction, reported flux values vary depending on the sampling and analysis methods used, even within a single sampling type (e.g. tethered traps vs. neutrally buoyant traps). These errors could not be directly quantified among the different methods and environmental conditions.

2.4 Model optimization procedure

The model was optimized with each dataset using a Bayesian approach that maximizes a log-likelihood function (Neal, 2003),

$$\log \text{likelihood} = -\sum_i^{N_{bin}} \left[W_i + \log \left(\frac{Model}{Observations} \right)^2 \right]. \tag{5}$$

In Equation (5) i is the index of each lat/lon/month bin, N_{bin} is the number of bins, and W_i is the weight for each bin, calculated as the inverse of the variance,

$$W_i = \frac{1}{\sigma_i^2}. \tag{6}$$

The negative coefficient in Equation 5 means that the log likelihood reaches a maximum when the model-data misfit is minimized. For each parameter, log-prior values were assigned that penalize

14

large parameter deviations from what is expected. The log prior uses the four (indexed by j) baseline parameters ($f_{alg} = 0.1$, $f_{recM} = 0.3$, $f_{recS} = 0.1$, $m_{ph} = 0.1$) from S14 as the prior information.

$$\log \text{prior} = \sum_1^4 (\log(parameter_i) - \log(expected_i))^4 . \tag{7}$$

The log prior values (quantifying the departure from the expected value of the parameter) and the log likelihood PDF were summed together, collectively forming the log posterior probability function,

$$\log \text{posterior} = \log \text{likelihood} + \log \text{prior} . \tag{8}$$

The uncertainty associated with each observation is explicitly accounted for in each dataset. The quantifiable source of error, as described in Section 2.3.5, for each dataset (v_i) was calculated by:

$$v_i = \left[(0.43 * data_1 - 0.63 * data_i)^2 \Big/ N_{obs,i} \right] - (data_i * error_{z,i})^2. \tag{9}$$

where $error_{z,i}$ is the error associated with the Martin curve correction for each bin, with larger errors associated with greater deviations between the sampling depth and Z_{eu}. The intrinsic variability (first term in Equation 9) was normalized to the number of observations within a given bin so that if multiple measurements were taken within the same lat/lon/month bin, the effective error from natural variability was reduced. From this the standard deviation (sigma) was calculated as

$$\sigma_i^2 = \log(1 + {}^{v_i}\!/_{\log{(data)}^2}).$$ (10)

In this way, the quantifiable uncertainty associated with the observations is explicit within the standard deviation formulation. This affects the weight for each bin within the log-likelihood function (Equation. 7) so that the values of the optimized parameters are influenced by the magnitude of the observational uncertainty.

We used a slice-sampling scheme (Neal, 2003) that generates a set of samples from the posterior probability density function (PDF). First, four initial parameter values (to represent $f_{alg}, f_{feca}, f_{recs}, m_{ph}$) were sampled from a uniform distribution within [0,1]. The model was run with these parameters and the value of the likelihood function (Equations 5, 6) was calculated. This process was repeated iteratively to generate a "chain" of of 2000 samples of each parameter from the posterior PDF. We ran 50 chains independently and terminated sampling when the chains had converged to a solution, quantified by a 'Potential Scale Reduction Factor' near 1 (Brooks and Gelman, 1998). This factor confirms that the chains, and thus the final parameter values, are independent of the randomly drawn initial values. The first 50% of samples from each chain were discarded as part of the 'burn-in' phase when the simulated chains were exploring the posterior function for solutions. Finally the remaining parameter values were saved, and their means and standard deviations are reported.

This procedure was repeated using 5 different datasets for the optimization, and using 3 different satellite NPP products (CbPM, VGPM, and CAFE) to drive the carbon export model. This yielded a total of 15 possible combinations of NPP product and optimization dataset.

3.0 Results

The optimized parameters and their associated uncertainties using the different datasets are presented in Fig. 4 a - d, alongside the correlation coefficients and log likelihoods for each optimization run (e,f), model mean export fluxes (g), and the average ez-ratio (h, export flux at base of the euphotic layer divided by NPP) for the corresponding locations and times in each dataset. The mean values discussed below are for outputs generated using the lat/lon/month bins specific to each dataset. The figure shows model performance using all three NPP satellite data products but the discussion below is only for the results using the CbPM, as is done in S14. The role of NPP algorithm choice on model outcomes is discussed in Section 4.1.

3.1 Data-model comparisons with the 'All Data' data synthesis

Both the S14 baseline model and the 'All Data' optimized model under-predict the mean C export flux (model mean of 68 mg C m^{-2} d^{-1} compared to the data mean of 113 mg C m^{-2} d^{-1}). This result is visually confirmed by the model-data comparisons (Fig. 5a) that show how the model is unable to reproduce both high flux and low flux observations (those that exceed 300 mg C m^{-2} d^{-1} or fall below 10 mg C m-2 d-1), which is due to the fact that individual observations are subject to intrinsic variability not captured by the climatological model.

The optimized model parameter values for the fraction of NPP_M in algal sinking (f_{Alg}), the non-grazing mortality parameter (m_{Ph}), the fecal fraction for big zooplankton (f_{FecM}), and the fecal fraction for small zooplankton (f_{FecS}) are 0.21± 0.11, 0.05±0.02 d^{-1}, 0.13±0.08, and 0.09±0.02, respectively (Fig. 4a-d) . This contrasts with the baseline case with parameter values f_{Alg} = 0.1,

17

$m_{ph} = 0.1$ d^{-1}, $f_{fecM} = 0.3$, and $f_{fecS} = 0.1$. Notably, the optimized algal fraction is double the

baseline case, and the non-grazing mortality parameter is half the baseline case. Further, the relative

contribution of small and large phytoplankton to the fecal flux is roughly equal in the 'All Data'

optimized models, whereas in the baseline case large phytoplankton herbivory contributes 3 times

more to fecal flux than herbivory on small phytoplankton.

The mean departure of the model from the data is roughly equivalent to the model mean (77

mg C m^{-2} d^{-1} compared to 68 mg C m^{-2} d^{-1}), and the correlation coefficient is weak (r = 0.22),

suggesting poor model performance against the 'All Data' global dataset.

3.2 Data-model comparisons with the 'Climatological Data' synthesis

Using 'Climatological Data' the optimized parameters more closely coincide with the

baseline case compared to those optimized using 'All Data', with values $f_{Alg} = 0.24 \pm 0.17$, $m_{ph} = $

0.07 ± 0.02 d^{-1}, $f_{fecM} = 0.21 \pm 0.13$, and $f_{fecS} = 0.08 \pm 0.03$, and the parameter uncertainties are about

the same as in 'All Data.' The departures from S14 are also similar to those observed in 'All Data'

with the exception of the ingestion parameter for large phytoplankton. The fecal fraction for big

zooplankton (0.21) is higher than in 'All Data' but is still reduced compared to the S14 value (0.3),

again indicating that the contribution of large phytoplankton to the fecal flux pathway is smaller

than previously prescribed. The 'Climatological Data' algal flux fraction is again roughly double the

18

S14 value, and the value for non-grazing mortality is slightly reduced.

The optimized model mean (62 mg C m^{-2} d^{-1}) is very close to the data mean (67 mg C m^{-2} d^{-1}) which, taken together with a good correlation coefficient (r = 0.79), suggests that the model predicts the observed climatological variations well. This result is visually confirmed by comparing Fig. 5a and 5b, which is analogous to the conceptual diagram presented earlier (inset panel in Fig. 1). A dataset with high noise results in a poor correspondence with the model, but a dataset of averaged observations has an improved agreement with the model, which is aimed to represent the mean of the system. The log likelihood is -4.8, compared to -25.2 with 'All Data,' which indicates a substantial improvement in model fit to the observations.

3.3 Data-model comparisons with '^{234}Th Profiles'

The optimized parameters using the '^{234}Th Profiles' dataset are roughly equal (Fig. 4), with values f_{Alg} = 0.07±0.04, m_{Ph} = 0.06±0.02 d^{-1}, f_{FeM} = 0.07±0.05, and f_{FeS} =0.08±0.02. This result implies an equal contribution of small and large phytoplankton to the fecal flux pathway, and also a diminished role of aggregation compared to the baseline case and to the other datasets. The optimized model represents improvement over the baseline case, which predicts a mean of 47 mg C m^{-2} d^{-1} compared to 69 mg C m^{-2} d^{-1} when confronted with a data mean of 53 mg C m^{-2} d^{-1}. Despite this result, the correlation is very weak (r = 0.08; Fig. 5c). For low flux environments, the '^{234}Th Profiles' optimized model over-predicts export flux.

3.4 Data-model comparisons with the 'HOT' and 'BATS' time series observations

The results from the model optimization with HOT and BATS are shown in Fig. 4 and 5d,e.

Optimizations with the climatological model and data, rather than the individual observations, are performed to ensure that the spatiotemporal scale of the model and data are similar. For the 'HOT' time series, the optimized parameters are $f_{Alg} = 0.11 \pm 0.07$, $m_{ph} = 0.11 \pm 0.06$ d^{-1}, $f_{egM} = 0.10 \pm 0.06$, and $f_{egS} = 0.11 \pm 0.05$. The optimized parameter values are nearly equivalent to S14 except for the egestion efficiency parameter for big zooplankton, which is much lower than the S14 baseline. At BATS, the optimized parameters using the CbPM are $f_{Alg} = 0.12 \pm 0.11$, $m_{ph} = 0.10 \pm 0.07$ d^{-1}, $f_{egM} = 0.13 \pm 0.12$, and $f_{egS} = 0.11 \pm 0.16$, which has substantially more uncertainty with each parameter compared to those estimated from the HOT dataset.

The optimized model using CbPM as an input does not predict the flux variation observed at HOT or BATS well (r values are 0.20 and -0.44 at HOT and BATS, respectively), though we note that the model was not designed to predict intrasite variations. We do so here because HOT and BATS provide climatological POC flux products. At BATS, the S14 POC export is 18 mg C m^{-2} d^{-1} while the optimized POC export is 30 mg C m^{-2} d^{-1}, which is closer to the observed mean of 47 mg C m^{-2} d^{-1}. Even so, the overall model performance for these sites is poor, which is unsurprising when comparing the poor correspondence of the seasonal signals in the CbPM model to the observed seasonality in the trap record for both sites (Fig. S1). Model performance is significantly improved when CAFE is used at HOT and BATS (r = 0.90 and r = 0.82, respectively), demonstrating the importance of the NPP algorithm.

4.0 Discussion

We compared the Siegel et al. (2014) POC export model with available field observations, and determined optimal parameter values and model predictions using five different export flux

datasets that represent different spatial and temporal scales. While the model can be optimized to fit POC observations from a specific place and time, no one set of model parameters can describe the totality of observed variability from local to global scales. These findings emphasize the need for both accurate satellite data products (especially NPP) for driving the POC export model, as well as well-sampled flux observations for optimizing model parameters. Here, we discuss how the magnitude and mechanisms of global biologically-driven carbon export in the model are influenced by the NPP product and POC flux datasets used to optimize the model.

4.1 Importance of the satellite NPP data products

The results of the model are sensitive to the choice of NPP, and some NPP products allow the model to better predict POC export in some locations better than others. NPP models are known to differ dramatically in their regional patterns of NPP, and their performance compared with observations needs improvement (Kahru, 2017; Lee et al., 2015; Doney et al., 2009), especially in high-nutrient low-chlorophyll conditions, and at extreme temperatures or chlorophyll concentrations (Carr et al., 2006 and refs therein). Thus far, the model-data results have been examined with the CbPM model. However, adjusting the choice of NPP product has large impacts on the magnitude of the optimized parameters (Fig. 4a-d). For example, with 'All Data', optimized parameters can vary by 100%, especially when comparing the VGPM generated parameters to those from CbPM and CAFE (Fig. 4). At HOT, the fraction of ingested small phytoplankton that drives export varies from 9% to 18% (Fig. 4d) depending on NPP product used. Lower NPP values from VGPM (Fig. S1) result in higher optimized fractions of small phytoplankton that enter fecal export, in order to produce final POC export values that are comparable in magnitude to the observations.

The choice of NPP product not only affects the optimized parameter values, but also affects model performance. Figure 4 shows how model performance is linked to NPP product. The most

21

pronounced case is at BATS, where using CbPM leads to a correlation coefficient of -0.44, but using VGPM leads to a correlation coefficient of 0.83. NPP estimates can vary by more than a factor of 3 at local to regional scales, which explains why the choice of NPP product leads to large variations in model outcome (Fig. S1). The various NPP products also differ in their predictions of the variability of NPP over a seasonal cycle at HOT and BATS (Fig. S2).

Other satellite data inputs, such as the slope of the particle size distribution and the backscattering coefficient as a proxy for phytoplankton carbon content, are only approximations of their intended representations, and they are also subject to large uncertainties. However compared to the other data inputs, it is the magnitude of NPP that largely sets the magnitude of the export flux; thus, the location at which the modeled and observed POC export fluxes are being compared greatly influences which NPP product will be most successful in producing the observed magnitude and variation of flux. No one NPP product succeeds in producing modeled POC export fluxes that match well with the data everywhere. This underscores the need for improving satellite NPP models if they are to be used in models of POC export (e.g. Siegel et al., 2014; DeVries and Weber, 2017).

4.2 Implications for the magnitude and mechanisms of POC export

The optimized parameters found using the three global datasets with observations from multiple locations ('All Data', 'Climatological Data', '[234]Th Profiles') can potentially provide meaningful information about ecological interactions on global scales. Of all the datasets used, 'Climatological Data' has the minimum amount of uncertainty associated with the observations, across depth, time, space, and method (Table 1). It is also the dataset that agrees best with the model predictions, as shown by the highest correlation coefficients and log likelihoods, independent of NPP product used. Despite the differences in the sources of error for 'All Data', 'Climatological

Data', and '^{234}Th Profiles', the general patterns in global flux are consistent. Maps of flux and the ez-ratio show that the highest ez-ratios occur in areas of upwelling and along coastal margins (Fig. 6). In those places, pulses of nutrients can create 'leaky' food webs where a greater fraction of production is exported to depth (Buesseler, 1998). The algal ratio (algal flux / total flux) is also highest in those high EZ-regions, where there is a high concentration of large phytoplankton. In the more dilute oligotrophic regions that are dominated by small phytoplankton, the role of direct algal sinking relative to total flux is reduced. We note that this result largely comes from the model construction, which explicitly parameterizes algal flux as a function of NPP.

The magnitude of the globally-integrated POC export predicted by the model is also sensitive to the dataset used to optimize the model. Using the CbPM, the annually-integrated global POC export is 5.5, 5.4, and 3.8 Pg C yr^{-1} when optimizing the model with 'All Data,' 'Climatological Data' and '^{234}Th Profiles,' respectively (Fig. 6; hereafter all global summary statistics are with the CbPM and optimized parameters). The model-predicted annually-averaged global ez-ratios (the fraction of NPP that becomes POC export) are 0.10, 0.10, and 0.07, respectively. Global POC export from the models optimized to fit 'All Data' and 'Climatological Data' is similar to that predicted by the baseline model (~6 PgC yr^{-1}; Siegel et al., 2014). However, the model that is optimized to match the '^{234}Th Profiles' dataset predicts a substantially smaller globally integrated flux value of 3.8 Pg C yr^{-1}. This number is similar to that derived by Henson et al. (2011), who used ^{234}Th-estimated POC fluxes to calibrate a temperature-dependent empirical model of POC export. One major distinction between the '^{234}Th Profiles' dataset and the others is that much of the data comes from oligotrophic regions (e.g., South Pacific and the Sargasso Sea), and is therefore skewed toward low-export environments (the mean flux of the '^{234}Th Profiles' dataset is 48 mg C m^{-2} d^{-1} compared to 67 mg C m^{-2} d^{-1} with 'Climatological Data' and 113 mg C m^{-2} d^{-1} with 'All Data'). The model's optimized parameters for '^{234}Th Profiles' are also similar to those

at BATS and HOT, two oligotrophic regions. This indicates that the global POC export is probably too low when the model is optimized using the '^{234}Th Profiles' dataset.

The pathways of POC export are also sensitive to the dataset used for optimization, and differ from the baseline S14 values. The most pronounced departure of parameter values from the S14 baseline model is in the values of the direct algal settling term. The contribution of sinking aggregates to global particulate carbon flux is highest in the 'All Data' and 'Climatological Data' cases, with algal ratios (the proportion of NPP that is exported as aggregates) of 25% and 29%, respectively, compared to the baseline case of 10%. There is a reduced fraction of large phytoplankton grazing losses routed to the fecal flux pool (13% and 21% for 'All Data' and 'Climatological Data', respectively, compared to 30% in the baseline), implying a smaller role of large phytoplankton to fecal flux than in the S14 model. The results with the 'All Data' and 'Climatological Data' differ from those of the '^{234}Th Profiles', which produces a reduced algal ratio of 7%. Again, because the observations in the '^{234}Th Profiles' dataset are predominantly from oligotrophic regions that are dominated by small particles in a diffuse environment, it follows that there would need to be a smaller proportion of sinking aggregates relative to total NPP than in high-NPP regions. Figure 6 shows higher algal ratios along coastal margins and areas of upwelling, which is because there is a greater fraction of large plankton in those areas.

One consistent departure from the S14 baseline model across all datasets is the smaller value of non-grazed mortality parameter, which is similar for all datasets (0.05, 0.07, 0.06 d^{-1} for 'All Data', 'Climatological Data', and '^{234}Th Profiles', respectively), compared with the baseline case of 0.1 d^{-1}. These values approach the range of values typically used in global ecosystem models (e.g. Aumont and Bopp, 2006 where phytoplankton mortality = 0.01 d^{-1}; Aumont et al., 2003 where mortality ranges from 0.008 d^{-1} up to a maximum of 0.2 d^{-1} depending on the type of phytoplankton; Le Quéré et al., 2005, where maximum mortality is 0.05 d^{-1}; Fasham et al., 1990 where mortality

varies from 0.045 to 0.06 d^{-1}). These optimized non-grazed mortality rates suggest that death by viruses and programmed cell death is less than 10% per day at the global scale.

The optimized results reveal that, while the magnitude of the globally integrated export flux is nearly unchanged from the S14 model to the optimized model using 'Climatological Data', the optimized model parameters suggest fundamentally different things about the magnitude of the processes contributing to flux. Compared to S14, the optimal model predicts a higher fraction of NPP that goes into algal flux, a reduced rate of non-grazing mortality, and a larger role of small plankton to total fecal flux. Of the total POC flux, the average direct algal sinking contribution is 29%, with the remainder coming from large fecal flux (13%) and small fecal flux (58%), using the parameters from 'Climatological Data' and CbPM. In all, the optimal model highlights the elevated importance of direct algal settling globally compared to what was expected based on previous food web models (Michaels and Silver, 1988; Boyd and Stevens, 2002).

4.3 Importance of the depth horizon used for export flux

Carbon export flux attenuates strongly with depth due to both biological and physical factors, making the choice of depth at which POC flux is considered 'export flux' critically important for interpreting both the overall magnitude of POC export and the ez-ratio. Different studies have used various definitions of the export depth, ranging from the mean mixed layer depth, the base of euphotic zone, the fixed depths of traps, arbitrary reference depths such as 100m, and the depth of winter mixing (e.g. Antia et al., 2001; Buesseler and Boyd, 2009; Henson et al., 2011; Li and Cassar, 2017). Here we use the '^{234}Th Profiles' dataset to recalculate model fluxes at several different reference depth: the Primary Production Zone (PPZ) (the depth where fluorescence reaches 10% of its maximum; Owens et al, 2015), the monthly mean mixed layer depth (MLD), and 100m (see supplementary material for methodology). The choice of these in any given study will have

different rationales (mechanistic or nominal) depending upon their usage (Palevsky and Doney, 2018).

Depending on the depth normalization used, the optimized global flux varies by ~40%, from 2.8 to 4.0 Pg C yr^{-1} (Table S2). These export fluxes are small compared to other global estimates because the dataset used is '^{234}Th Profiles.' The largest calculated flux corresponds to the shallower 1% PAR calculations, whereas the minimum integrated flux corresponds to the deeper PAR values (given by the PPZ, which is roughly equivalent to 0.1% PAR). Using the monthly mixed layer depth as the export depth yields a total POC export in between these extremes at 3.2 Pg C yr^{-1} (Table S2). The magnitude of these differences in POC export demonstrates how model comparisons can be confounded by the choice of export depth.

Adjusting the export depth also results in small changes in the model-predicted mechanisms responsible for POC export (Table S2). The biggest change is for the non-grazing mortality rate, which decreases with increasing depth (Figure S5). This is likely because the assumptions of the model (i.e. phytoplankton carbon biomass is negligible below the mixed layer) imply that a deeper production layer is more dilute. In general, larger flux observations associated with shallower reference depths generate higher optimized values for the algal flux fraction and the ingestion fraction of big phytoplankton.

4.4 What is needed to robustly compare POC export models with observations?

Calibrating POC export models with observations can help quantify the global and regional magnitudes of POC export, and improve our mechanistic understanding of biologically mediated carbon export. However, one should be careful to ensure that the observations to which the model is being compared are suitably-averaged in order to match the scale of the model. For global POC export models, these scales are typically climatological. It is not appropriate to calibrate climatological models with data that does not represent climatological conditions, as most export

data do not.

The datasets described herein span a spectrum of sampling methods, duration, locations, times, and depths; the ability of our POC export model to represent climatological POC export fluxes is confirmed with the 'Climatological Data' compilation because these data are averaged to minimize naturally occurring variability within the bounds of the space and time scales captured by the model. The global dataset compiled here reveals that spatial variability within a one-degree grid box is roughly equivalent to temporal variability within a given locality (63% and 43% uncertainty, respectively). Ideally, data would be sampled multiple times at several locations within each model grid box, in order to reduce uncertainty in flux estimates to a point where they can be meaningfully compared with the model output. It is not possible to assess the biases between sediment traps and the [234]Th technique here because there are insufficient observations from traps at Z_{eu}.

An ideal sampling scheme to calibrate food-web models for carbon export is one that quantifies activity in the surface ecosystem that produces POC flux, and the processes that regulate its transport to depth. The upcoming EXPORTS field campaign will accomplish this sampling on a limited scale (Siegel et al., 2016), which aims to identify shifts in ecosystem 'states' that regulate export. However if we want to calibrate global POC export models with POC flux data from sediment traps or [234]Th techniques, we will need long-term observations at many locations throughout the ocean. Ideally we would have a global array of long-term flux measurement sites that have minimal sampling errors, and that can carry out observations for 3-7 days per month throughout the year. Alternatively, in the absence of sediment traps or [234]Th techniques, optical proxies for flux that are available at comparable spatiotemporal scales should be investigated for potential incorporation into global models (i.e., Dall'Olmo and Mork, 2014; Estapa et al., 2017).
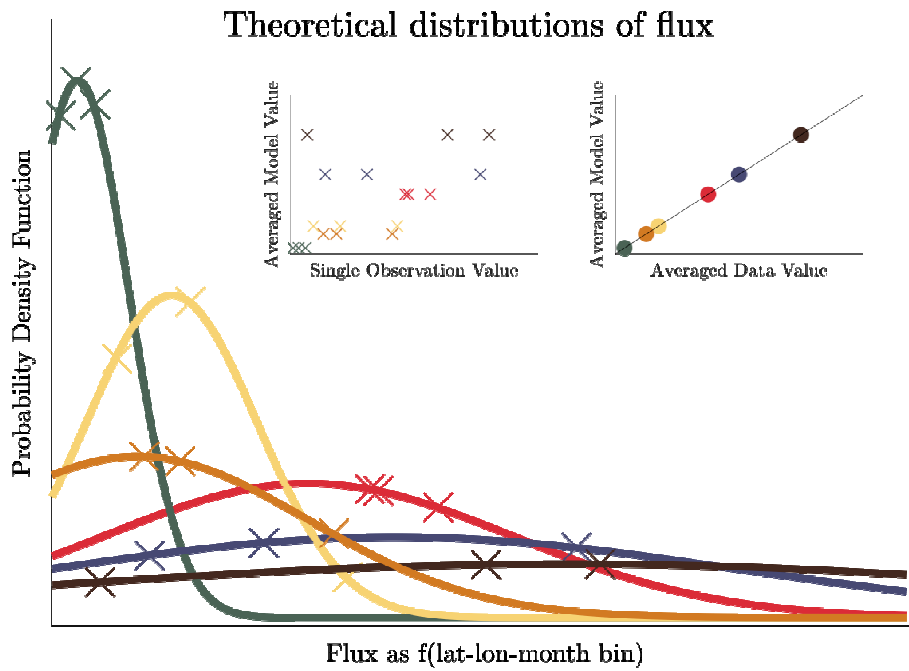
Until we have suitable long-term POC flux measurements of the kind needed to calibrate global climatological models, it will be necessary to compile datasets that average over long temporal and large spatial scales (e.g. 'Climatological Data') in order to minimize intrinsic and

extrinsic variability in the system so that model-data comparisons can be used to extract meaningful insights about carbon export. It will also be useful to calibrate POC export models with data that is not subject to the same intrinsic and extrinsic sources of variability that affect carbon flux estimations, such as ocean tracer data (e.g. DeVries and Weber, 2017).

This study has also shown that uncertainty in satellite NPP data products is a major weakness in carbon cycle models that rely on satellite NPP products. Models will benefit significantly from improvements in algorithms for estimating NPP from satellite observations, particularly if those algorithms are themselves constrained by climatological oceanographic observations such as long-term mean POC export fluxes and ocean tracer data.

Acknowledgments

Theoretical distributions of flux

Probability Density Function

Averaged Model Value

Single Observation Value

Averaged Model Value

Averaged Data Value

Flux as f(lat-lon-month bin)

Figure 1 caption: Theoretical distributions of biological carbon flux arising from complexities within lat-lon-month model bins are plotted. The inset shows how model – data comparisons look when data is either randomly sampled (left) or sufficiently averaged (right). The distributions are randomly sampled three times each and their values are plotted in the left inset. The overall sampling average of the distribution is plotted against the distribution's true mean in the right inset.
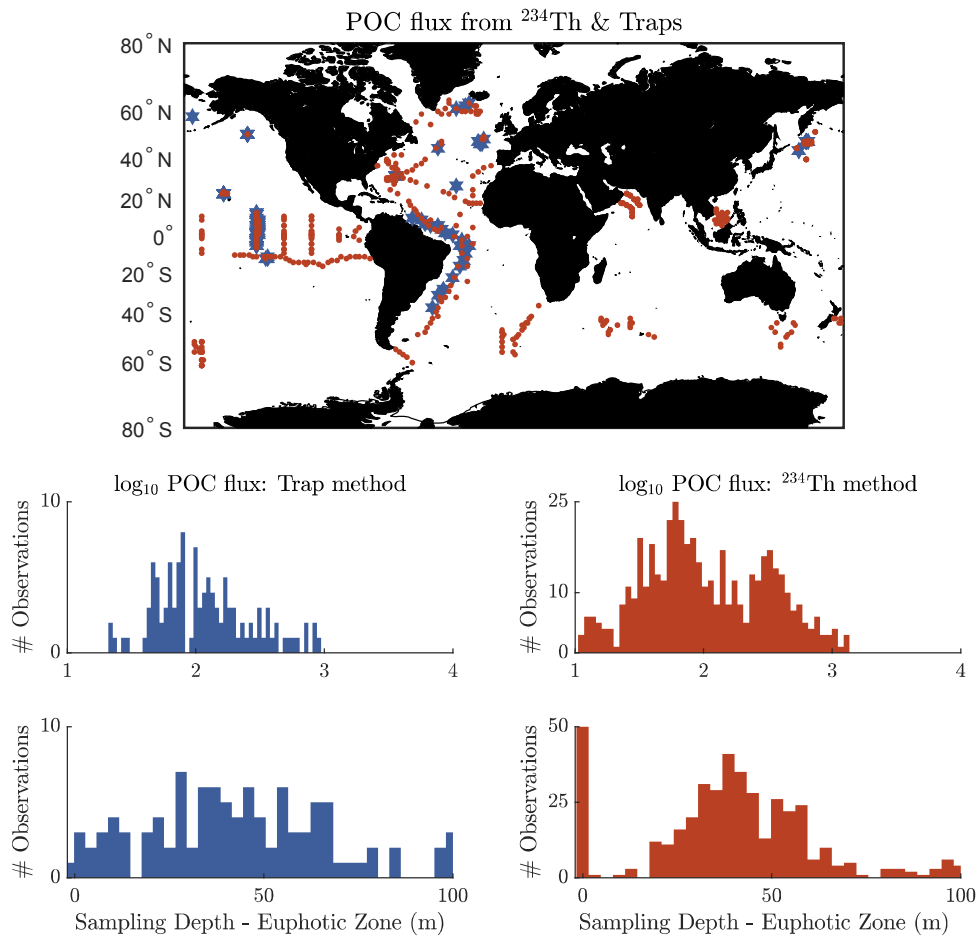
Figure 2 caption: Map showing the sediment trap (blue) and POC flux from ^{234}Th (red) locations of the bins (n=591). The bins are unique lat/lon/months in the complete dataset arising from the data treatment described in the methods section. The bottom panels show histograms of these data products after they are normalized to the euphotic zone (Z_{eu}), and of their sampling depth relative to the climatological Z_{eu}, which is the depth of 1% PAR . Note the scale difference on the y-axis for each product.
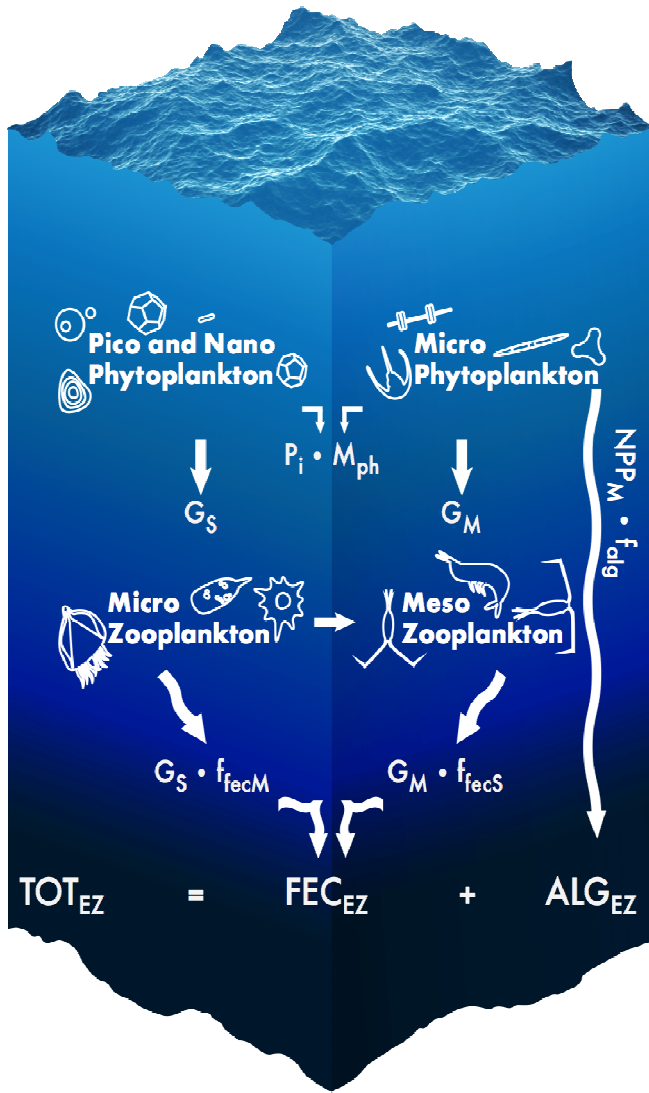
Figure 3 caption: Diagram showing the conceptual elements of the model. NPP is routed into two phytoplankton size classes (pico and nano phytoplankton (small), micro phytoplankton (large)) which are ingested at a rate given by micro and meso zooplankton herbivory (Gs and Gm, respectively). In the baseline model case, a lower ingestion efficiency is attributed to the smaller phytoplankton size class to implicitly account for higher trophic transfers. This is illustrated here by

the flux line between micro and meso –zooplankton. Fecal flux from both zooplankton sizes classes is added together with an aggregation flux from bigger phytoplankton, totaling the reported particulate organic flux (POC) at the euphotic layer, Z_{eu}. Image credit: Gad Girling.
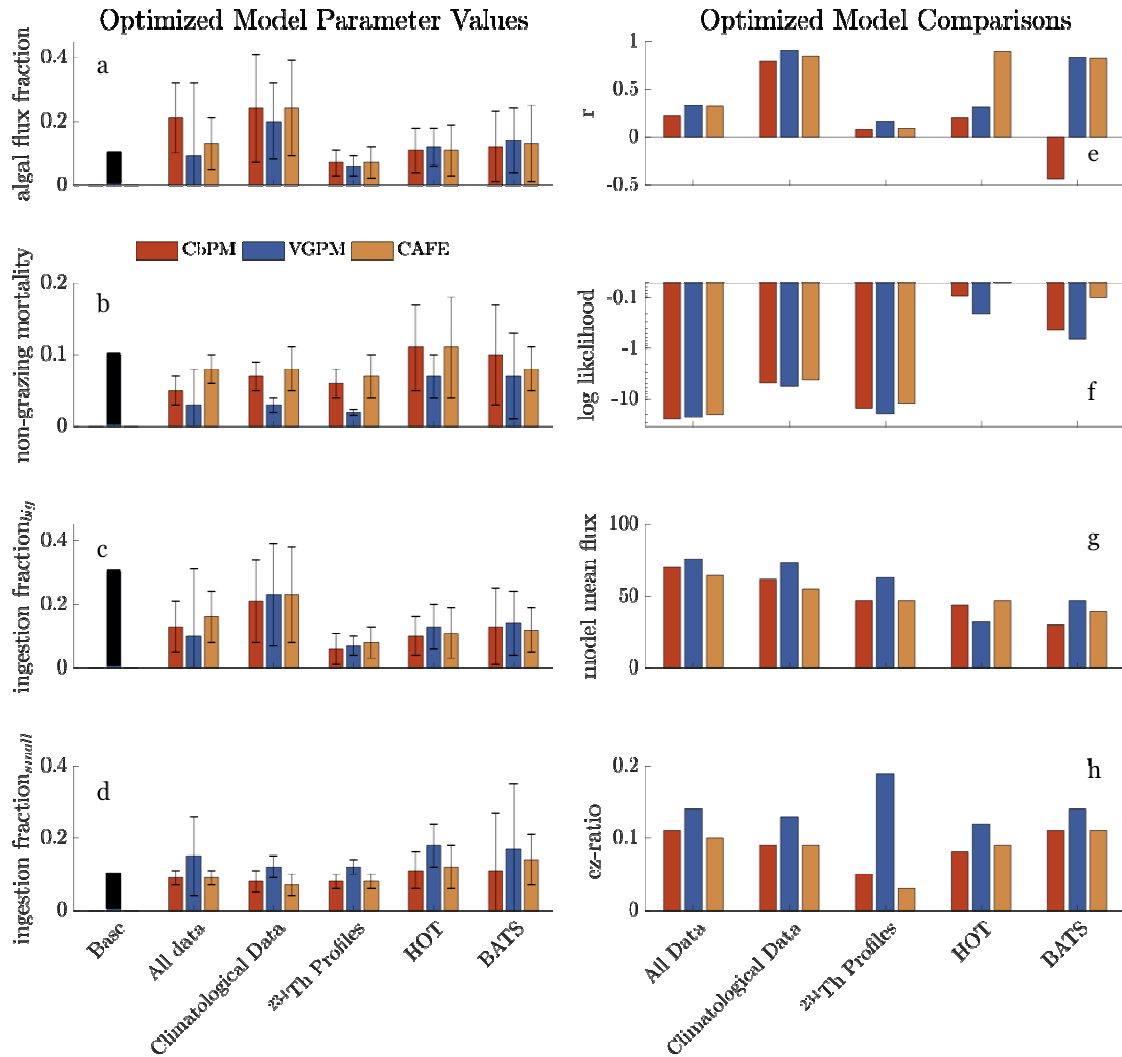
Figure 4 caption: Left panel: Bar graphs for optimized parameter values for the various NPP input products and the model development data sets are shown. NPP product comparisons are shown in color (red is CbPM, blue is VGPM, and tan is CAFE). The a priori values for each parameter are shown in cream on the far left. The x-axis shows the dataset used. Error bars indicate the

32

magnitude of the standard deviation for each product. Note the y-scale differences among parameter type. Right panel: The top panel shows the Pearson correlation coefficient, r, which compares the specified dataset against the optimized model. The 2nd panel shows values for the log likelihood, where a lower (more negative) log-likelihood equates to a higher cost, and consequently a higher model-data mismatch. The 3rd panel shows the values for the mean model flux (mg C m^{-2} d^{-1}), and the final panel shows values for the ez-ratio, the fraction of production that is exported.
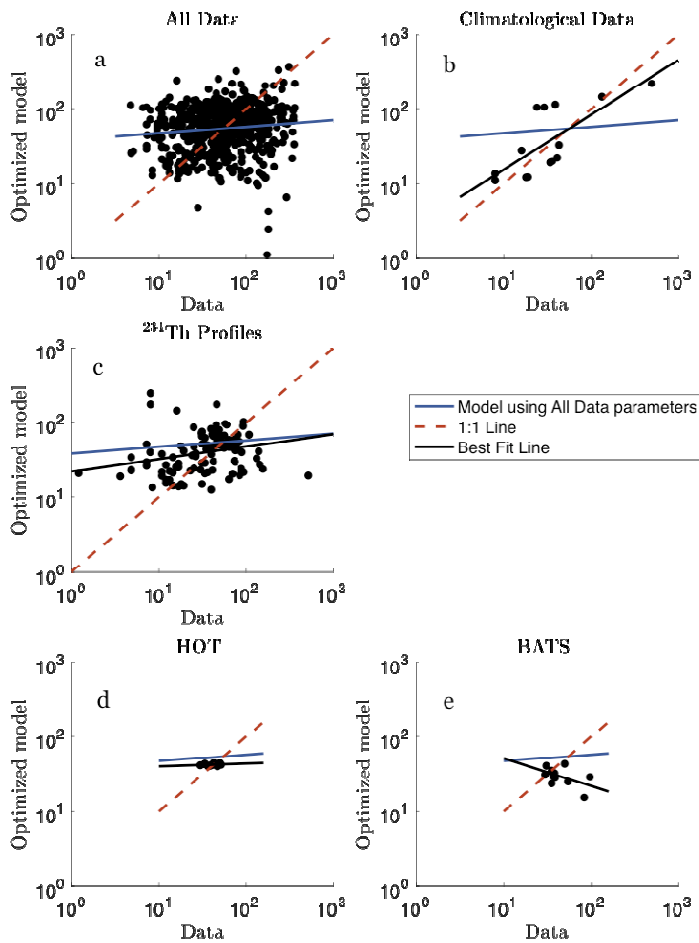
Figure 5 caption: Model vs data scattergraphs are compared among the different dataset types using the CbPM model as the NPP input. The navy line is the best fit line for 'All Data' case, which

is shown in the other graphs to illustrate how the model would look if the 'All Data' parameters were used in those different scenarios. The red line is the 1:1 line, and the black line is the best fit line for b) through d).
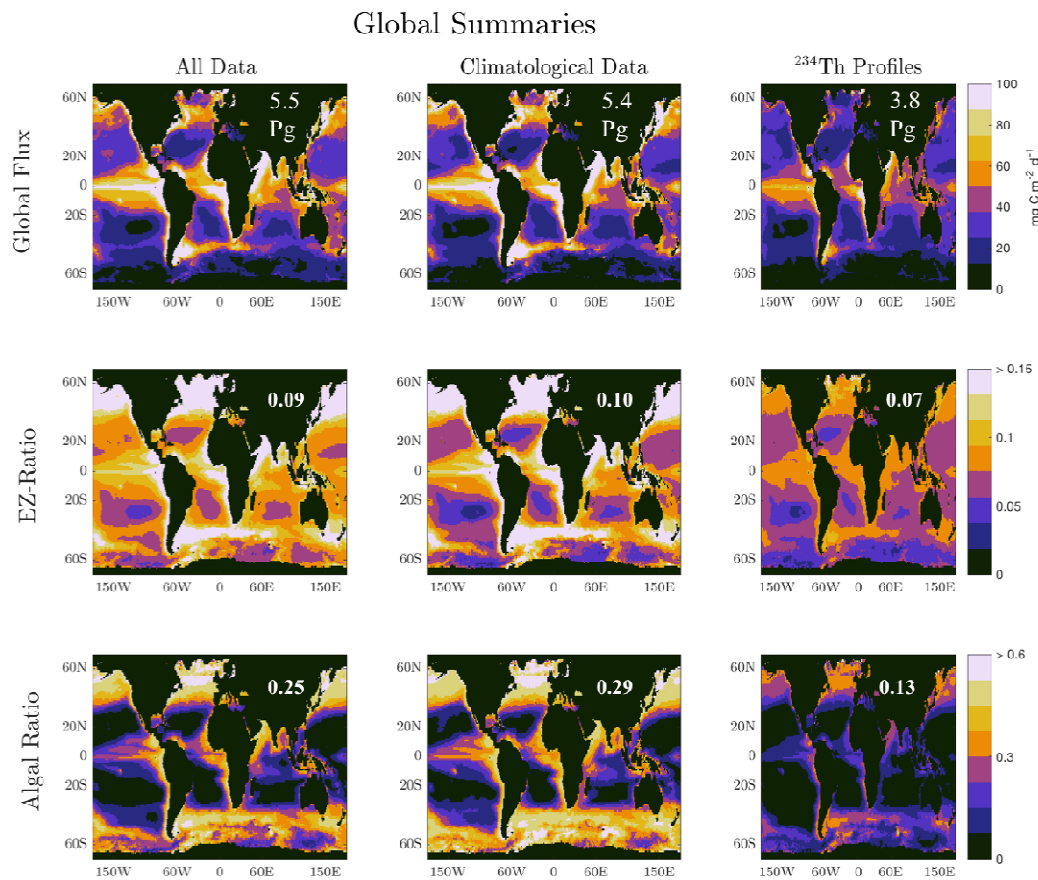
Figure 6 caption: Global summary maps and statistics are shown for the 3 datasets that span several ocean basins ('All Data', 'Climatological Data', '234Th Profiles'). The annually integrated POC

flux values are shown in white on the top panel, in Petagrams. The EZ-ratio (the proportion of production that is exported beneath the euphotic layer) is shown in the 2nd row, with global averages in white. The algal ratio (the fraction of flux from aggregates) is shown in the 3rd row with global averages in white. All of the reported statistics and maps are done with CbPM as an input production product.

SOURCES OF ERROR				
	Natural Variability		Instrumental/ sampling biases	Normalization to Zeu
	Temporal	Spatial		
All Data	✓	✓	✓	✓
Climatological Data		✓	✓	
^{234}Th Profiles	✓	✓	✓	
HOT & BATS		✓	✓	✓
Uncertainty	43% *	63% *	≥ 30%	0-60% *

Table 1 caption: Table showing sources of error and uncertainty for each dataset used. The starred columns (temporal, spatial, and Z_{eu} normalization) are used in the cost function to account for dataset-specific sources of error. Checked columns indicate that error is present within that dataset, which arises from insufficient sampling to overcome temporal, spatial, and/or depth biases. Temporal variability is not considered a problem for 'Climatological Data' because the timescale of interest is 1 month. A description of how the error values are calculated is in the methods section.

Supporting Information

Introduction

This supplement contains text that expands on the methods for section **'4.3 Importance of the depth horizon used for export flux'** as well as figures to support the main text, which follow.

The depth used to calculate flux has a substantial impact on the globally integrated flux. The euphotic layer (here defined as the 1% light level) is used as the depth where export flux to the deep initiates, but not all phytoplankton exist within the full depth of the euphotic layer (which can be much greater than the 1% light level) when there is a strong density gradient keeping them in the surface. The mixed layer depth constricts the position of plankton within a column of water and can be thought of as the production layer (Li and Cassar, 2017) during those instances when light is replete. On the other hand, it is reasonable to consider the primary production zone (PPZ) as the base of fluorescence (or where fluorescence reaches 10% of its maximum, as used by Black et al, 2018)), which is closer to the 0.1% light level. Because it's not possible to calculate this number from satellite data directly, we choose the 1% light level and add 100m to it, which is close to the PPZ depths reported in Black et al., (2018).Here we describe the methods we employ to test the importance of the export flux depth horizon.

Holding the model structure constant allows for the results to be evaluated in the context of a variable integration depth. The depths considered are the Mixed Layer Depth, 100m, and The Primary Production Zone (PPZ) as defined by Black et al, 2018. The '^{234}Th Transects' profiles were used to recalculate the fluxes at the depths specified above. MLD at the time of sampling was not readily available in all cases, so in those instances (i.e. stations within the Dutch Geotraces

Transect) it was calculated from CTD profiles using the Geotraces Data Extractor (<https://webodv.awi.de/geotraces_extractor_ctd>).

For each station, mixed layer depth was calculated as the depth where density is less than a density threshold (compared to the density at 10m where the threshold is defined as the density where temperature is < 0.2C from the temperature at 10m). This threshold is calculated as

$$\Delta Density = 0.2C * alpha \ (C^{-1}) \times 1025 \ (kg \ m^{-3})$$

where alpha is the coefficient of thermal expansion for a given salinity, temperature, and pressure. For the stations with available data across the depths of interest, the U-Th difference (in dpm/L) was multiplied by the decay rate for ^{234}Th to get the ^{234}Th at various depths above and below the MLD, 100m, and PPZ. From there we linearly interpolated across depth to get ^{234}Th flux at the specified depth, and multiplied that flux by the C/Th relationship specified in Owens et al, 2015 to get POC flux. This yielded 3 datasets of POC flux at the MLD, 100m, and PPZ. We do not recalculate the data according to the model value of Z_{eu} because those studies consider PPZ to be Z_{eu}. Therefor, for the model depths of 1% light and PPZ, we use the same flux data but change the depth of the production zone within the model.

Then the model was adjusted so that all NPP occurs within the boundaries of the depth of interest, where previously all NPP is assumed to be within the boundaries of the 1% light level. This approach is imperfect because it relies on several assumptions: production is limited to the specified depth (but not below), and the general food web framework is appropriate in all cases. Despite these assumptions it is useful to put bounds on global flux as a function of normalization depth. With in the model framework, equations 3 and 4 are adjusted to remove the entrainment term, and to account for the depth used (Z, where Z = {PPZ, 100m, MLD}), yielding

$$\frac{dP}{dt} = \frac{NPP_i}{Z} - G_i - m_{ph}P_i - \delta_{i,M}\frac{AllG_{PZ}}{Z_{:}}.$$

$$FCC_{BZ} = \left[(f_{fecM} + G_M) + (f_{fecS} + G_S)\right] * Z.$$

From this the model parameters were optimized in the same way as described in the main text.
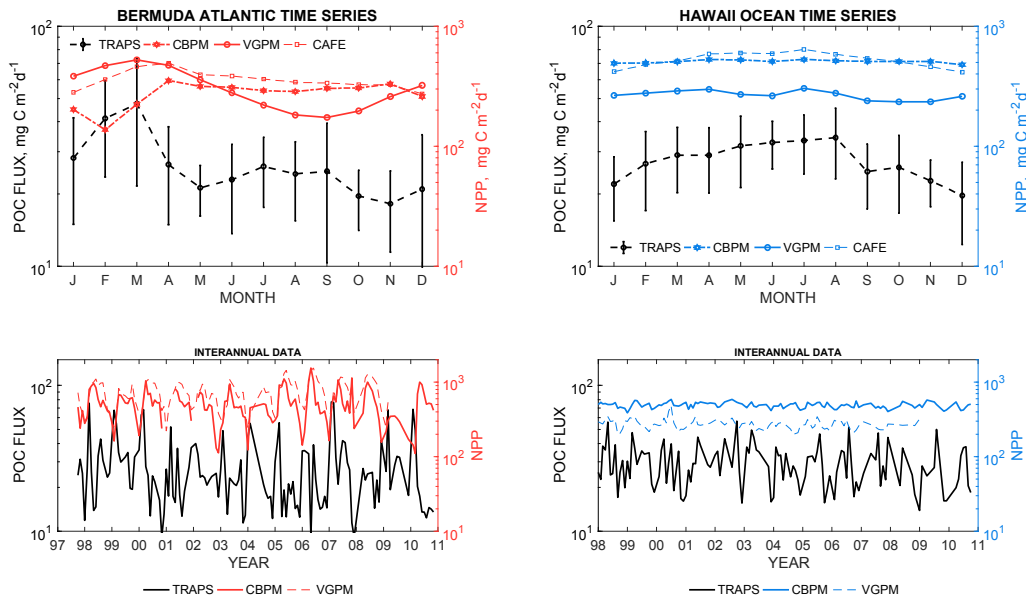
gure S1. Graphs of BATS and HOT climatological trap data are compared to climatological satellite NPP product on the top row. The three NPP products are the Carbon Based Productivity Model (CbPM), the Vertically Generalized Production Model (VGPM) and the Carbon, Absorbing, and Fluorescence Euphotic-Resolving model (CAFE). Error bars on the trap data functionally represent standard deviations from the mean, and conceptually illustrate the variations in sampling once monthly. The bottom row shows interannual trap data plotted with available satellite matches (monthly average for 0.25 x 0.25 degree grid nearest to each site) for each month during the SeaWiFS era.
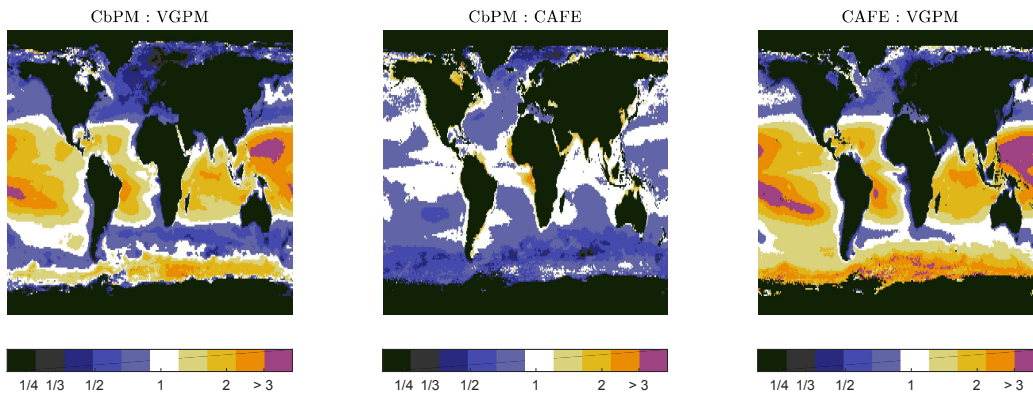
Figure S2. Ratios of in annually averaged climatologies for satellite NPP product (mg C m^{-2} d^{-1}) are compared across latitude for the 3 algorithms considered. The white shading is bounded by the contours (0.89, 1.2), meaning that the two products are roughly equivalent.

POC flux from ^{234}Th Profiles & Climatological Data
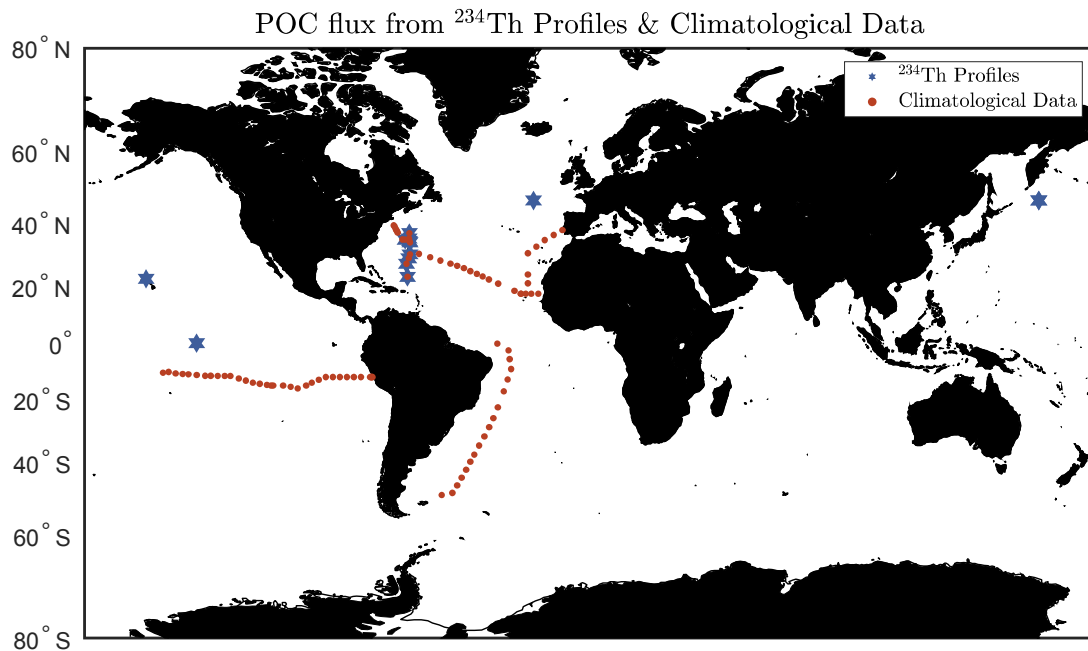
Figure S3. Map showing the locations for the 'Climatological Data' locations (red) and those for the '^{234}Th Profiles' data (blue).
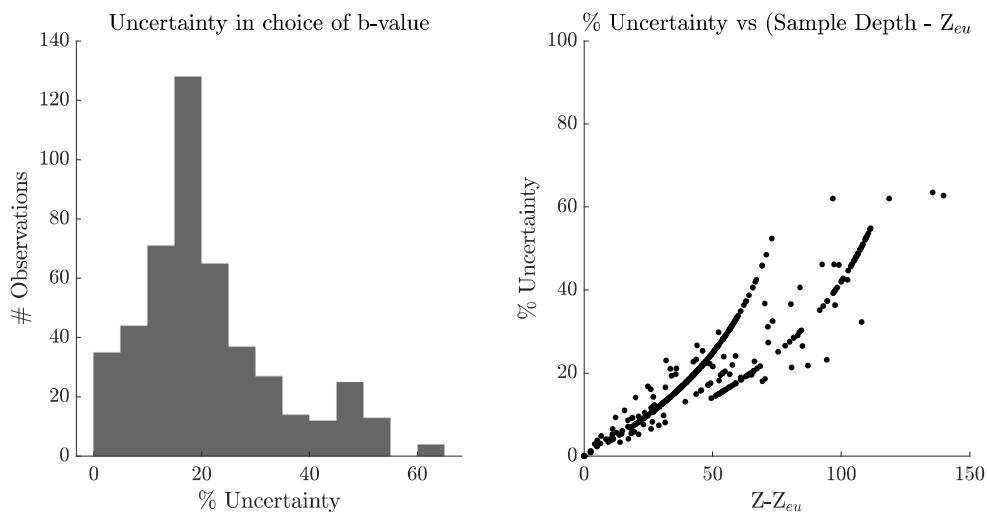
Figure S4. A histogram of the percent uncertainty arising from the martin b power law assumption for depth normalization is shown. The bins (N=591) for the 'All Data' dataset are used for baseline flux data. First we produce a normal curve with 'b' values of mean 0.7 and standard deviation 0.35 and sample it 1000 times. Then fluxes were calculated with these 'b' values, totaling 1000 fluxes for each bin. Their resulting standard deviations divided by the mean are here reported as % uncertainty. The average percent uncertainty after 10,000 simulations is 21.4%.

40

Ratio of parameters to optimized Z_{eu} base case
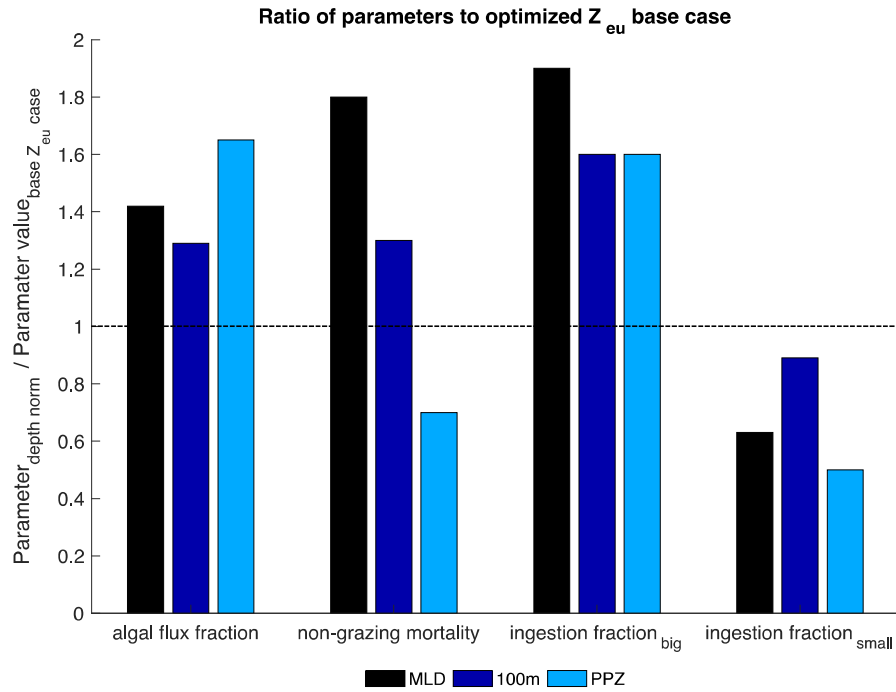
Figure S5. Departures from the baseline model (y-axis) are shown for the four model parameters (x-axis) across three depths (MLD in black, 100m in navy, PPZ in light blue). A value of 1 indicates that there is no change between the optimized value and the baseline case.

Dataset	NPP model	data mean values				optimized model mean values											
		Chl mg m^{-3}	EZ m	MLD m	POC flux mg m^{-2} d^{-1}	POC flux mg m^{-2} d^{-1}	r	f_{alg}	$\pm\sigma$	m_{ph}	$\pm\sigma$	fec_m	$\pm\sigma$	fec_s	$\pm\sigma$	log likelihood	global POC flux value
All Data	CBPM					68	0.22	0.21	0.11	0.05	0.02	0.13	0.08	0.09	0.02	-25.2	5.5
	VGPM	0.26	69	45	113	76	0.33	0.09	0.05	0.03	< 0.01	0.1	0.06	0.15	0.02	-23.3	6.2
	CAFE					64.5	0.32	0.13	0.08	0.08	0.02	0.16	0.08	0.09	0.02	-20	5.3
_BB09+_	CBPM					62	0.79	0.24	0.17	0.07	0.02	0.21	0.13	0.08	0.03	-4.8	5.4
	VGPM	0.21	74	38	67	73	0.91	0.2	0.12	0.03	0.01	0.23	0.16	0.12	0.03	-5.6	6.9
	CAFE					55	0.85	0.24	0.15	0.08	0.03	0.23	0.15	0.07	0.03	-4.2	5.6
234Th transects	CBPM					47	0.08	0.07	0.04	0.06	0.02	0.07	0.05	0.08	0.02	-15.1	3.8
	VGPM	0.3	69	44	48	63	0.17	0.06	0.03	0.02	< 0.01	0.07	0.03	0.12	0.02	-19.2	4.9
	CAFE					47	0.1	0.07	0.05	0.07	0.03	0.08	0.05	0.08	0.02	-12.2	4.2
HOT	CBPM					44	0.2	0.11	0.07	0.11	0.06	0.1	0.06	0.11	0.05	-0.09	4.5
	VGPM	0.07	86	50	41	32	0.31	0.12	0.06	0.07	0.03	0.13	0.07	0.18	0.06	-0.2	6.6
	CAFE					47	0.9	0.11	0.07	0.11	0.07	0.11	0.08	0.12	0.06	-0.05	5.6
BATS	CBPM					30	-0.44	0.12	0.1	0.1	0.07	0.13	0.12	0.11	0.16	-0.4	4.9
	VGPM	0.1	82	52	47	47	0.83	0.14	0.1	0.07	0.05	0.14	0.13	0.17	0.18	-0.7	6.5
	CAFE					39	0.82	0.13	0.07	0.08	0.03	0.12	0.07	0.14	0.07	-0.1	7

Table S1: Table of optimized parameters and performance metrics for each dataset and NPP product considered in the primary analyses. Sigma represents standard deviation. The integrated global POC flux value is in peta-grams. We include estimates from HOT and BATS for sake of comparison rather than to imply results from HOT and BATS should be extended to the global scale.

Depth	POC flux mg m^{-2} d^{-1}	POC flux mg m^{-2} d^{-1}	r	f_{alg}	m_{ph}	fec_m	fec_s	log likelihood	Integrated global flux Pg C yr^{-1}
MLD	35	35	0.86	0.1	0.11	0.13	0.05	-1.95	3.2
100	33	36	0.68	0.09	0.08	0.11	0.07	-1.7	3.5
1% Light	48	47	0.08	0.07	0.06	0.07	0.08	-15.1	3.8
PPZ	36	30	0.82	0.1153	0.04	0.11	0.04	-4.5	2.8

Table S2: Table of optimized parameters and performance metrics for each depth considered

Information on supplementary data set

We compiled POC flux (mg C m^{-2} d^{-1}) by first reviewing the compilation from Lutz et al., 2007. We added fluxes from more recent work (including studies also compiled by Henson et al., 2011 and Mouw et al., 2016) as well as historical studies not presented in Lutz et al., 2007. When direct data was not available for download (as in references 22 and 23) we digitized fluxes from visuals in publications, using the 'Web Plot Digitizer,' freely available online at https://automeris.io/WebPlotDigitizer/.

References

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, *391*(6667), 577.

Alldredge, A. L., & Silver, M. W. (1988). Characteristics, dynamics and significance of marine snow. *Progress in oceanography*, *20*(1), 41-82.

Antia, A. N., Koeve, W., Fischer, G., Blanz, T., Schulz☐Bull, D., Schölten, J., ... & Hebbeln, D. (2001). Basin☐wide particulate carbon flux in the Atlantic Ocean: Regional export patterns and potential for atmospheric CO2 sequestration. *Global Biogeochemical Cycles*, *15*(4), 845-862.

Antia, A. N. (2005). Solubilization of particles in sediment traps: revising the stoichiometry of mixed layer export. *Biogeosciences*, *2*(2), 189-204.

Armstrong, R. A., C. Lee, J.I. Hedges, S. Honjo, and S.G. Wakeham (2001), A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep Sea Research Part II Topical Studies in Oceanography*, **49**(1), 219-236.

Aumont, O., Maier☐Reimer, E., Blain, S., & Monfray, P. (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, *17*(2).

Aumont, O., & Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, *20*(2).

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite☐based chlorophyll concentration. *Limnology and oceanography*, *42*(1), 1-20.

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon☐based ocean productivity and phytoplankton physiology from space. *Global biogeochemical cycles*, *19*(1).

Behrenfeld, M. J., T O'Malley, R., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., ... & Boss, E. S. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, *444*(7120), 752.

Black, E. E., Buesseler, K. O., Pike, S. M., & Lam, P. J. (2018). 234Th as a tracer of particulate export and remineralization in the southeastern tropical Pacific. *Marine Chemistry* (201), 35-50.

Boyd, P. W., & Stevens, C. L. (2002). Modelling particle transformations and the downward organic carbon flux in the NE Atlantic Ocean. *Progress in Oceanography*, *52*(1), 1-29.

Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 7, 434-455.

Buesseler, K.O. (1998), The decoupling of production and particular export in the surface ocean. *Global Biogeochemical Cycles*, **12**(2), 297-310.

Buesseler, K. O., Antia, A. N., Chen, M., Fowler, S. W., Gardner, W. D., gafsson, O., ... & Steinberg, D. K. (2007). An assessment of the use of sediment traps for estimating upper ocean particle fluxes. *Journal of Marine Research*, *65*(3), 345-416.

Buesseler, K. O., Trull, T. W., Steinberg, D. K., Silver, M. W., Siegel, D. A., Saitoh, S. I., ... & Honda, M. C. (2008). VERTIGO (VERtical Transport In the Global Ocean): a study of particle sources and flux attenuation in the North Pacific. *Deep Sea Research Part II: Topical Studies in Oceanography*, *55*(14-15), 1522-1539.

Burd, A.B. and G.A. Jackson (2009), Particle aggregation. *Annual Review of Marine Science*, **1**, 65-90.

Carr, M. E., Friedrichs, M. A., Schmeltz, M., Aita, M. N., Antoine, D., Arrigo, K. R., ... & Bidigare, R. (2006). A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5), 741-770.

Cram, J. A., Weber, T., Leung, S. W., McDonnell, A. M. P., 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, 858–876. https://doi.org/10.1029/2017GB005710 □

DeVries, T., F. Primeau, and C. Deutsch (2012), The sequestration efficiency of the biological pump, *Geophysical Research Letters*, **39**, L13601, doi:10.1029/2012GL051963.

DeVries, T., and T. Weber (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.

Dunne, J. P., Armstrong, R. A., Gnanadesikan, A., & Sarmiento, J. L. (2005). Empirical and mechanistic models for the particle export ratio. *Global Biogeochemical Cycles*, *19*(4).

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).

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., ... & Takahashi, T. (2009). Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems*, *76*(1), 95-112.

Eppley, R.W., and B.J. Peterson, Particulate organic matter flux and planktonic new production in the deep ocean, Nature, 282, 677- 680, 1979

Fasham, M. J. R., Ducklow, H. W., & McKelvie, S. M. (1990). A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, *48*(3), 591-639.

Gardner, W. D., Hinga, K. R., & Marra, J. (1983). Observations on the degradation of biogenic material in the deep ocean with implications on accuracy of sediment trap fluxes. *Journal of Marine Research*, *41*(2), 195-214.

Gardner, W. D. (1985). The effect of tilt on sediment trap efficiency. *Deep Sea Research Part A. Oceanographic Research Papers*, *32*(3), 349-361.

Giering, S. L. C., Sanders, R., Martin, A. P., Lindemann, C., Möller, K. O., Daniels, C. J., ... & St John, M. A. (2016). High export via small particles before the onset of the North Atlantic spring bloom. *Journal of Geophysical Research: Oceans*, *121*(9), 6929-6945.

Giering, S. L. C., R. Sanders, A. P. Martin, S. A. Henson, J. S. Riley, C. M. Marsay, and D. G. Johns (2017). Particle flux in the oceans: Challenging the steady state assumption, *Global Biogeochemical Cycles*, 31, 159–171, doi: 10.1002/2016GB005424.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., van Dongen-Vogels, V., ... & Behrenfeld, M. J. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, *102*, 16-25.

Gust, G., Michaels, A. F., Johnson, R., Deuser, W. G., & Bowles, W. (1994). Mooring line motions and sediment trap hydromechanics: in situ intercomparison of three common deployment designs. *Deep Sea Research Part I: Oceanographic Research Papers*, *41*(5-6), 831-857.

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).

Henson, S. A., Sanders, R., & Madsen, E. (2012). Global patterns in efficiency of particulate organic carbon export and transfer to the deep ocean. *Global Biogeochemical Cycles*, *26*(1).

Henson, S. A., Yool, A., & Sanders, R. (2015). Variability in efficiency of particulate organic carbon export: A model study. *Global Biogeochemical Cycles*, *29*(1), 33-45.

Karl, D. M., Laws, E. A., Morris, P., & Emerson, S. (2003). Global carbon cycle (communication arising): metabolic balance of the open sea. *Nature*, *426*(6962), 32-32.

Karl, D. M., & Lukas, R. (1996). The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2-3), 129-156.

Kahru, M. (2017). Ocean productivity from space: Commentary. *Global Biogeochemical Cycles*, *31*(1), 214-216.

Knauer, G. A., Martin, J. H., & Bruland, K. W. (1979). Fluxes of particulate carbon, nitrogen, and phosphorus in the upper water column of the northeast Pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, *26*(1), 97-108.

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*(C9).

Lee, Z., Marra, J., Perry, M. J., & Kahru, M. (2015). Estimating oceanic primary productivity from ocean color remote sensing: A strategic assessment. *Journal of Marine Systems*, *149*, 50-59.

Li, Z., & Cassar, N. (2016). Satellite estimates of net community production based on O2/Ar observations and comparison to other estimates. *Global Biogeochemical Cycles*, *30*(5), 735-752.

Li, Z., & Cassar, N. (2017). A mechanistic model of an upper bound on oceanic carbon export as a function of mixed layer depth and temperature. *Biogeosciences*, *14*(22), 5015.

Mahadevan, A., E. D'Asaro, M.J. Perry and C. Lee (2012),Eddy-driven stratification initiates North Atlantic Spring phytoplankton blooms. *Science*, **337**, 54-58.

Marra, J., C. Langdon, and C. A. Knudson (1995), Primary production, water column changes, and the

demise of a Phaeocystis bloom at the Marine Light-Mixed Layers site (59N, 21W) in the northeast Atlantic Ocean. *Journal of Geophysical Research* **100**, 6633-6643.

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

Mawji, E., et al., The GEOTRACES Intermediate Data Product 2014, Mar. Chem. (2015), http://dx.doi.org/10.1016/j.marchem.2015.04.005.

Michaels, A. F., & Silver, M. W. (1988). Primary production, sinking fluxes and the microbial food web. *Deep Sea Research Part A. Oceanographic Research Papers*, *35*(4), 473-490.

Michaels, A. F., Silver, M. W., Gowing, M. M., & Knauer, G. A. (1990). Cryptic zooplankton "swimmers" in upper ocean sediment traps. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(8), 1285-1296.

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

Neal, R. M. (2003). Slice sampling. *Annals of statistics*, 705-741.

Owens, S. A., Pike, S., & Buesseler, K. O. (2015). Thorium-234 as a tracer of particle dynamics and upper ocean export in the Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, *116*, 42-59.

49

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*.

Primeau, F. (2006). On the variability of the exponent in the power law depth dependence of POC flux estimated from sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers*, *53*(8), 1335-1343.

Quere, C. L., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., ... & Klaas, C. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, *11*(11), 2016-2040.

Richardson, T. L., & Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science*, *315*(5813), 838-840.

Siegel, D. A., Fields, E., & Buesseler, K. O. (2008). A bottom-up view of the biological pump: Modeling source funnels above ocean sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers*, *55*(1), 108-127.

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.

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., ... & Perry, M. J. (2016). Prediction of the export and fate of global ocean net primary production: the EXPORTS science plan. *Frontiers in Marine Science*, *3*, 22.

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.

Steinberg, D.K., C.A. Carlson, N.R. Bates, S.A. Goldthwait, L.P. Madin, and A.F. Michaels (2000), Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep-Sea Research I*, **47**, 137-158.

Steinberg, D. K., Carlson, C. A., Bates, N. R., Johnson, R. J., Michaels, A. F., & Knap, A. H. (2001). Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry. *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8-9), 1405-1447.

Valdes, J. R., & Price, J. F. (2000). A neutrally buoyant, upper ocean sediment trap. *Journal of Atmospheric and Oceanic Technology*, *17*(1), 62-68.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean□driven atmospheric CO2 changes. *The Carbon Cycle and Atmospheric CO: Natural Variations Archean to Present*, 99-110.

Zehr, J. P., Weitz, J. S., & Joint, I. (2017). How microbes survive in the open ocean. *Science*, *357*(6352), 646-647.

II. Assessing mechanisms of the biological pump by integrating data from satellites, experiments and in situ observations

Introduction

Food web processes in the surface ocean undeniably control the flux of carbon into the deep sea through activities that convert suspended phytoplankton carbon into sinking particulate organic carbon (POC) flux (Alldredge and Silver, 1988), in what is commonly referred to as 'the biological pump' (Ducklow et al., 2001; Falkowski et al, 2000; Volk and Hoffert, 1985). Yet, the food web processes affecting POC flux are complex and difficult to model globally. Consequently, many studies focus on examining a particular process (e.g., Hansen et al., 1997; Burd et al., 2000; Suttle et al., 1990), and/or on the dynamics of a particular place (e.g., Mahadevan et al., 2012; Lancelot et al., 2000; Schlitzer, 2002; Richardson and Jackson, 2007) to better quantify the pathways of the biological pump. Recently, Siegel et al., (2014) embedded global satellite observations in a two-size class food web model (hereafter S14) to evaluate the processes controlling POC flux worldwide. This model has been widely used for a variety of applications, and its simplistic parameterizations have not been reassessed since its first use.

Here we introduce an adaptable food web modeling framework that is data driven, informed by experimental results, and calibrated with POC flux observations recently shown to be the most suitable for model development (Bisson et al., 2018) to demonstrate differences in the mechanisms that control POC flux in the surface. We adapt the Siegel et al. (2014) food-web model (hereafter S14) to include more realistic parameterizations that

represent processes not resolved in the original model (aggregation of small and large particles, differential phytoplankton mortality based on either size or temperature). The idea is that the addition and removal of parameterized processes alters model assumptions, and by evaluating model behavior, the importance of these processes can be quantified on regional to global scales.

It is not a new idea to use food webs to represent and interpret biological pump function (Pomeroy, 1974; Azam et al., 1983; Michaels and Silver, 1988; Legendre and Le Fèvre, 1995; Boyd and Newton, 1995; Franks, 2002 and refs therein). The mechanics of the biological pump involve biological activities that are well represented through an ecological perspective: zooplankton consume phytoplankton and excrete organic matter throughout the water column, bacteria metabolize dissolved organic matter and in some cases form endosymbiotic relationships with phytoplankton, phytoplankton form aggregates and sink, and phytoplankton die from viral infection, releasing dissolved organic matter into the water column. These processes can be examined through in situ observations, experimental work, theoretical perspectives, and modeling frameworks that can be used to test ecological hypotheses, but it is uncommon to do so. Our understanding of the ecological processes affecting POC flux is therefore limited to method specific perspectives.

In this study we optimize and examine five models ('No Size,' 'S14,' 'Aggregation,' 'Aggregation + Size Dependent Mortality,' 'Aggregation + Temperature Dependent Mortality') that require a different combination of processes, in order to compare what their differences imply about the functioning of the biological pump. We evaluate performance using POC flux observations and zooplankton grazing experimental results for all models. Our aim is that the model results will be useful for future model development and sampling

strategies, particularly to learn which processes are most influential and uncertain in the biological pump.

2. Modeling methods, optimizations, and sensitivities

2.1 Model Overview

The overall modeling goal is to build on S14 to optimize and examine four new plausible mechanistic models that predict POC export from satellite observations. We choose a nested modeling framework because it allows for parameters to be added and removed, so that their influence on the overall model product can be examined simply. The baseline model (S14) routes net primary production (NPP) through a two-size food-web model informed from satellite observations (full details are available in Siegel et al., 2014). The final model output is POC flux (TOT_{EZ}, mg C m^{-2} d^{-1}), which is determined at the base of the euphotic zone (Z_{eu}) through the sum of modeled fecal export (FEC_{EZ}) and sinking algal export (ALG_{EZ}).

Fecal fluxes are calculated from unknown herbivory rates for each size class, and are determined through mass balance using satellite observations of phytoplankton biomass ($P, dP/dt$), modeled net primary production (NPP, mg C m^{-2} d^{-1}) modeled mortality which includes death by viruses, cell lysis, and non-grazing death (collectively mph), modeled particle aggregation (ALG_{SEZ}, ALG_{lEZ}), and detrainment below the mixed layer ($\frac{P_s}{Z_{ml}} \frac{dZ_{ml}}{dt} H$, equations 3,4).

55

$$\frac{dP_s}{dt} = \frac{NPP_s}{Z_{eu}} - G_s - m_{phs} * P_s - \frac{AGG_{sEZ}}{Z_{eu}} - \frac{P_s}{Z_{ml}} \frac{dZ_{ml}}{dt} H\left(\frac{dZ_{ml}}{dt}\right). \tag{3}$$

$$\frac{dP_l}{dt} = \frac{NPP_l}{Z_{eu}} - G_l - m_{phl} * P_l - \frac{AGG_{lEZ}}{Z_{eu}} - \frac{P_l}{Z_{ml}} \frac{dZ_{ml}}{dt} H\left(\frac{dZ_{ml}}{dt}\right). \tag{4}$$

As in S14, the last terms in equations 3 and 4 represent the reduction in phytoplankton carbon due to mixed layer deepening where Z_{ml} is the mixed layer and H(x) = 1 if x > 0 and 0 otherwise.

The POC fecal flux (FEC_{EZ}) is calculated by multiplying herbivory rate for each size class (s = small, 0.2 – 20μm; l = large, 20-50 μm) by size-specific fecal flux fractions that quantify the fraction of ingested carbon that becomes fecal flux.

$$FEC_{EZ} = [(fec_l * G_l) + (fec_s * G_s)] * Z_{eu}. \tag{5}$$

In this study we consider 3 new processes (aggregation, and differential mortality based on either size or temperature,) that add 6 new possible parameters ($W_{sink}, agg_l, agg_s, m_{phs}, m_{phl}, m_{phr}$) listed and described in table 1. Differential mortality is modeled based on size (m_{phl}, m_{phs}) and temperature (m_{phr}). An expected mortality rate

of 0.1 d^{-1} is held constant across all models, consistent with previous work (Doney et al., 2009) in order to investigate how the different mechanistic models represent mortality processes. In the case of differential mortality based on size, we expect mph_s will be higher than mph_l because the high concentration of small phytoplankton makes them more vulnerable to infection through increased probability of viral contact, relative to large phytoplankton. An alternative hypothesis is that viral rates will be higher for large phytoplankton, because viruses act on blooms and/or areas of upwelling, where large cells dominate (Furhman, 1999) and are also susceptible to infection.

Temperature dependent mortality is modeled using a Q_{10} of 2, which is the factor whereby a rate doubles with an increase of 10 degrees Celsius, relative to a reference temperature T_0, (10 degrees here) and a free parameter to be optimized, m_{PAr}, following work in Chan et al., 2004, Chen and Liu, 2010, and Coles and Jones, 2000.

$$m_{PA} = m_{PAr} \times Q_{10}^{\frac{(T - T_0)}{10}}. \tag{5}$$

This increases mortality rates for warmer waters (e.g., the subtropics) compared to cooler waters in the high latitudes or along areas of upwelling.

In the original model direct algal sinking flux (ALG_{LZ}) is formulated as a fraction of net primary production performed from the micro sized phytoplankton, but it is adapted here to explicitly designate aggregation between large and small particles. The aggregation flux of small phytoplankton on other small phytoplankton is modeled as the square of the phytoplankton carbon contained in the small fraction.

$$ALG_{SEZ} = agg_s(P_s^2).$$ (6)

The explicit formulation for aggregation of small cells onto other small cells tests the importance of smaller sinking material within the biological pump (Richardson and Jackson, 2007; Giering et al., 2017; Durkin et al., 2015). The aggregation flux of large particles is modeled as the sum of a sinking fraction of large phytoplankton plus an additional term to account for aggregation of large phytoplankton with small phytoplankton

$$ALG_{LEZ} = W_{sink} + P_l + agg_l(P_l^2 + P_l \times P_s).$$ (7)

where W_{sink} is the sinking speed for large cells (Smayda et al., 1979) and assumed to be 1 m d^{-1}. The two aggregation coefficients ($agg_{l,s}$) account for the probability that a large particle forms an aggregate with either a large or small particle, or that a small particle aggregates with another small particle, respectively. Both agg_l and agg_s represent, but do not explicitly formulate, mechanisms for aggregation, so a specific expected value is not assigned to them (table 1). However we expect agg_l to exceed agg_s because aggregation is

thought to primarily influence large particle settling, although small particles do form aggregates (Richardson and Jackson, 2007).

The total aggregation term is the sum of the large and small aggregates:

$$ALG_{EZ} = ALG_{sEZ} + ALG_{lEZ}. \tag{8}$$

Finally, fecal flux (FEC_{EZ}) is calculated by applying size-specific fecal fractions to grazing rates, as in S14. In S14, fec_s is assumed to be 10% and fec_l is 30% to account for less trophic steps involved in the transfer of larger phytoplankton to zooplankton, increasing the efficiency with which that food is processed. However, when prey is large compared to the predator, more biomass is thought to be lost through sloppy feeding (Lampert, 1978, Møller and Nielson, 2001), which would convert POC to DOC, and thus reduce the POC available for export. Therefore, we take as a null hypothesis that the fecal flux fractions are equivalent (20%) for both size classes, and so that we can consistently compare the parameter outcomes with a model that does not resolve size, and which assumes an expected value of 20% for all phytoplankton.

2.2 Specific model descriptions

1. The 'No Size' model uses three parameters: mph, fec, and agg. ALG_{EZ} is modeled as the square of total phytoplankton carbon concentration and an aggregation

parameter: $agg(P^2)$. The herbivory calculation uses one equation for ingestion of all phytoplankton sizes, that is:

$$\frac{dP}{dt} = \frac{NPP}{z_{eu}} - G - m_{ph} * P - \frac{ALG_{sEZ}}{z_{eu}} - \frac{P}{z_{ml}} \frac{dz_{ml}}{dt} H\left(\frac{dz_{ml}}{dt}\right).$$ (9)

2. The 'S14' model is the original satellite informed food-web model, which uses four parameters: m_{ph}, fec_s, fec_l, and f_{alg}. The algal parameter, f_{alg}, is the fraction of NPP performed by large phytoplankton that becomes sinking POC.

3. The 'Aggregation' model uses all the ALG_{EZ} free parameters but not the differential mortality terms, totaling five parameters. In the herbivory equations, $m_{phs} = m_{phl}$.

4. The 'Aggregation + Temperature Dependent Mortality' uses all the ALG_{EZ} free parameters plus one to adjust the temperature dependent mortality term, m_{phT}, totaling five parameters.

5. The 'Aggregation + Differential Mortality' model uses six parameters in total: all the ALG_{EZ} free parameters plus two separate parameters for the mortality rates on small and large phytoplankton, that is m_{phs}, m_{phl}.

2.3 Model optimizations

We optimize the free parameters ($agg_s, agg_m, m_{phs}, m_{phl}, m_{phT}, fec_s, fec_l$, and f_{alg} in all

models ('No Size,' 'S14,' 'Aggregation,' 'Aggregation + Temperature Dependent Mortality,' 'Aggregation + Differential Mortality,' listed in order of increasing degrees of

freedom) in order to identify parameter values that reduce the model/data misfit, while retaining reasonable fidelity to their expected values. The fitting metric is the log posterior function, which combines the model-data misfit for POC flux (given by the log likelihood, equation 10) with penalties for parameter deviations from what is expected (given by the log prior, equation 11).

The model-data mismatch is minimized when the log likelihood is maximized:

$$\log \text{likelihood} = - \sum_{i=1}^{n} [\log(model) - \log(observations)]^2. \tag{10}$$

The model values and observations are log transformed because POC flux follows a lognormal distribution, and to give equal weight to differences from small and large fluxes.

The log prior is added to the log likelihood, which quantifies the departure of an optimized parameter from its expected value (see Table 1). There is no prior value on the size specific aggregation terms, because the expected value for agg_s or agg_l is not known from ecological or particle aggregation theory.

$$\log \text{prior} = - \frac{\sum_{j=1}^{n_p} [\log(parameter_n) - \log(expected_n)]^2}{n_p}. \tag{11}$$

 The log prior is normalized by the total number of parameters in a model (n_p) so that there is no penalty for additional parameters. The parameters are log transformed to increase

61

their magnitudes within the bounds of the expected parameter values ([0, 0.5]), so that the retrieved parameter samples retain reasonable fidelity to their expected value. Then, the log posterior is calculated:

$$\log posterior = \log likelihood + \log prior. \tag{12}$$

All models are optimized with the "Climatological Data" dataset from Bisson et al. (2018), which was demonstrated to be the most appropriate dataset of shallow POC flux observations for calibrating global models. A slice sampling technique is used to generate parameter values, which is a Markov chain sampling method (Neal, 2003). For each model we draw initial parameter values ($x_{u,k}$) equal to the number of parameters (k) in each experiment (Table 1) from a uniform distribution within [0,1]. These values define the vertical 'slices' through the log probability density function, which is the log posterior. At the value of $x_{u,k}$ a log posterior value is uniformly sampled within this vertical slice. New parameters are drawn from where the sampled value of the log posterior bisects the log posterior curve. Samples are drawn in this way until 2000 new parameter values are generated. Four instances of slice sampling are run to get four 'chains,' and sampling is stopped if the chains converge to the same pdf (given by a 'Potential Scale Reduction Factor' near 1 (Brooks and Gelman, 1998)). The first 50% of the samples are discarded as part of the 'burn-in' phase and the remaining parameter values are saved. The reported uncertainties represent the interquartile range of the sampled parameters.

The models are scored using the Pearson correlation coefficient (r) and the Root Mean

Square Difference (RMSD). The correlation coefficient is used to quantify the extent to which the model explains the variance in the data, and the RMSD is used to measure the average difference between model output and observations.

2.4 Model parameter sensitivities

We also run global POC flux sensitivities to each parameter in every model to quantify the influence of a parameter value on integrated global POC fluxes. First, for a given parameter in a certain model, we randomly draw 1000 values from the optimized parameter results, without replacement. Then, to calculate how global POC fluxes changes e.g., as a function of the agg_s term for the 'Aggregation' model, we hold $agg_l, m_{Tn}, fcc_v, fcc_i$ constant to their mean value and we run the global 'Aggregation' model 1000 times with the variable agg_s values, and record the integrated POC fluxes. This does not account for the co-variance between the model parameters, although it does quantify how changes in one parameter value will lead to changes in integrated POC flux.

3.1 Input data

The models are driven by monthly mean observations from the Sea-Viewing Wide-Field-of-View Sensor (SeaWiFS) satellite ocean color mission from September 1997- December 2008, as is done in S14 (Figure 1). We use the CAFE NPP satellite data product (Silsbe et

al., 2016) because it showed the best performance across all of the datasets analyzed in Bisson et al. (2018), and because it showed best performance with in situ NPP measurements compared to other NPP models (Silsbe et al., 2016). The choice of NPP product will affect parameter values to an extent, although we note the variations in parameter values introduced from NPP product choice are small compared to the standard deviations in parameter values for the model development dataset chosen here (i.e., 'Climatological data' see supplementary table 1 from Bisson et al., 2018).

We apply the slope of a power law particle size spectrum to estimate fractional biovolumes of particles (Kostadinov et al, 2010) between 0.2 -20μm, and from 20- 50 μm, which is available at <http://wiki.icess.ucsb.edu/measures/index.php/PSD>. In particular we use NPP (mg C m^{-2} d^{-1}), fraction of mirco- and sum of pico- and nano-phytoplankton size particle biovolumes, Z_{eu} (m), and phytoplankton carbon (P, mg C m^{-3}). We use SeaWiFS surface chlorophyll concentrations (< http://oceancolor.gsfc.nasa.gov/cgi/l3?sen=S&per=MO&prod=CHL_chlor_a>) to estimate the euphotic layer depth, Z_{eu} from formulations in Morel et al, 2007. We use particulate backscatter (b_{bp}) from the Garver-Siegel-Maritorena product at 443nm to estimate phytoplankton carbon from the relationship described in Graff et al., (2015) (b_{bp} fields are available at http://wiki.eri.ucsb.edu/measures/index.php/GSM). We also use mixed layer depths from the MLD_DT02 climatology (de Boyer Montégut et al., 2004),

available at <http://www.ifremer.fr/cerweb/deboyer/mld/Surface_Mixed_Layer_Depth.php> to use in combination with the satellite products.

We organize the data as follows: all data inputs are gridded onto a 1-degree latitude/longitude grid and averaged for each climatological month (Figure 1). As in S14, we require a minimum of 8 months a year of observations, which effectively excludes the high latitudes from analysis. The global mean euphotic layer depth is 63 ($\pm$19) m, the mean mixed layer depth is 65 ($\pm$ 29) m, the mean NPP is 490 ($\pm$ 142) mg C m^{-2} d^{-1}, and the mean b_{bp} is 0.018 ($\pm$ 0.018) m^{-1}, which corresponds to a mean phytoplankton carbon concentration of 23 ($\pm$ 22) mg C m^{-2} using the relationship presented in Graff et al., (2015).

3.2 Model development dataset

We use a carbon export field data compilation shown to minimize the variability arising from sampling (Bisson et al. 2018), and at the same depth as the model output, Z_{eu}. This dataset, called 'Climatological Data' in Bisson et al., (2018) includes 166 total observations of POC flux from the [234]Th technique at 15 different sites, with an average flux of 69 mg C m^2 d^{-1}. This dataset is spatially biased towards the Northern Hemisphere (Supplementary Figure 1), but the multiple observations are averaged within a given grid cell as to best represent the 1-degree, monthly resolution outputs of the model. This yields 15 lat/lon/month bins for direct comparison with model outputs.

4. Results

4.1 Retrieved parameter values across all models

Parameter values for the suite of five models are shown in figure 2. Prior to performing the optimizations we hypothesized that some parameter values will vary as new parameters are added, because when additional degrees of freedom are added, it might affect how parameterized processes operate together. This can be examined by comparing the S14 model to the models that build on it because they are identically formulated apart from their aggregation or mortality calculations. We find that the optimized values of the S14 model are nearly invariant when explicit aggregation is added, but are greatly changed when size is removed (compare the first through third bars for $m_{phs}, m_{phl}, fec_s, fec_l$). When differential

mortality based on size is added to the aggregation model, the mortality rate for large phytoplankton increases by nearly 100%. Otherwise, the optimal parameter values do not vary much across the different models considered here.

There are departures in the parameter values from their expected value. In particular, the fec_s terms for all of the models are less than their expected value of 0.2. Every model

except for the 'No Size' model predicts a reduced rate of small phytoplankton mortality from what is expected, as well as enhanced, if only slightly so, values for fec_l. While there were

no expected values for the two aggregation terms, there are variations in both the magnitude and the range of agg_s and agg_l, and agg_s is always less than agg_l. This indicates that for a

given magnitude of phytoplankton particles, more large cells will sink as aggregates

compared to small ones.   The implications of these, and of model sensitivities to change in parameter value, are reviewed in the discussion.

4.2 Model skill, sensitivities, and global statistics

Results from the model optimizations are shown in table 2, and figures 3 and 4. Model skill is evaluated on the basis of the log posterior, the correlation coefficient, and the root mean square difference. The fitting metric for model-data fit, the log posterior, increases with increasing degrees of freedom as expected, but this increase in performance does not imply improvement for other metrics of fit. The correlation coefficient is reduced in the case of the 'Aggregation + Size Dependent Mortality' compared to the 'Aggregation' model,' and the RMSDs are higher for 'Aggregation + Size Dependent Mortality' and 'Aggregation + Temperature Mortality.' Overall there is not a significant change in performance across the models, as measured by the model-data misfit for POC flux.

The global average statistics across the five models are different as well (Figure 3). The model that produces the highest average flux is the 'Aggregation' model, followed by the 'Aggregation + Size Dependent Mortality' model and then the other three. The model results for the average model fluxes are also reflected in the EZ-Ratio, which is the fraction of NPP that becomes POC flux at the based of the euphotic zone. The algal ratio varies widely across all the models, constituting between ~20 to 55% of the total POC flux. The implications of this, and for interpreting the plausibility of different mechanistic models, are reviewed in the discussion.

The global model mean values for integrated POC flux are roughly similar for all the models (Table 2), with a range from 6.1 to nearly 8 Pg C yr^{-1} (from 'No Size' to 'S14,' respectively). They also exhibit consistent patterns in their global sensitivities to particular parameters (Figure 4, Supplementary Figures 4-6). The global model sensitivities are consistent with what is expected: the models are especially sensitive to agg_l, f_{alg}, and fec_l along coasts and areas of upwelling, where there is a higher fraction of large phytoplankton. The reverse is true for the fec_s term, where adjustments to fec_s yield heightened POC fluxes in the gyres compared to coasts and high latitudes. Increasing agg_l, f_{alg}, and fec_l yields higher total POC flux, and increasing m_{phs}, m_{phl}, and m_{phr} yields lower total POC flux. In all models apart from 'No Size,' the fecal flux perturbations yield the largest spread for annually integrated POC flux values. In 'No Size,' the fecal flux and mortality parameter perturbations result in similarly large ranges for global integrated POC (~10 Pg C yr^{-1}). In all models, adjusting the algal and fecal flux terms leads to net positive POC flux changes, whereas adjusting the mortality terms yields net negative POC flux changes.

4.3 Space/time patterns in export flux

Results from the nested model experiments can be used to assess where and when particular mechanisms especially contribute to overall POC flux (Figure 5, Figures S2-4). Regions with similar magnitudes of total POC flux (i.e., oligotrophic or high latitude

68

regions) do not exhibit the same magnitudes of flux pathways (Figure 5) throughout the annual cycle. For example, although Station ALOHA and the South Pacific Gyre are thought to have similar capacity for POC flux, and indeed their annually average flux contributions are very similar, the timing of mechanisms contributing to POC flux varies between them. Station ALOHA is considerably less seasonably variable in each of the three pathways (compare shape of time series in Figure 5), and exhibits nearly equivalent fractional contributions from large fecal flux and aggregation year round. The South Pacific Gyre has a reduced fraction of small phytoplankton in the late summer, which corresponds to higher fractions of large fecal flux and algal flux. The peaks of aggregation and large fecal flux are slightly decoupled in the South Pacific Gyre, which contrasts to a tight coupling at Station Aloha for of aggregation and large fecal flux.

Similarly, the North Pacific and North Atlantic sites have roughly similar annually averaged contributions for each of the four pathways, but these sites vary widely in the flux pathways over time. For example, in the boreal summer the North Pacific has decreased algal sinking contribution (a change to 20% of total flux from May to August) whereas the North Atlantic has enhanced contributions of algal sinking (from 40% to 50%) from May to August. Throughout the entirety of the annual cycle in the North Atlantic, large fecal flux exceeds that of small fecal flux, but in the North Pacific, large and small fecal fluxes are negatively correlated (R= -0.74).

The magnitude of changes throughout the seasonal cycle at a regional basis is influenced by model construction (compare Figure 5 with Figures S2-S4). While the relationships between the processes remain unchanged (i.e., the general shapes of the seasonal cycles are preserved), the magnitude of particular fractions is especially enhanced for the North Pacific

69

and the North Atlantic. Depending on the model used, the contribution of aggregation can vary by a factor of 2.

Within any one model the global fractional combinations of mechanisms to total POC flux are little influenced by the time varying magnitude of POC flux (Figure 5; Supplementary Figures 2-4). However, the contribution of mechanisms across models varies, even for similar POC flux magnitudes (Figure 5, Figures S2-S4). It therefore is the model type and construction, not the POC magnitude that sets the combination of mechanisms contributing to overall POC flux for the global seasonal cycle. The pathways most affected by model construction is the algal flux fraction, which is enhanced in the 'Aggregation + Temperature Morality' and 'Aggregation + Size Dependent Mortality' models (37% and 38%, respectively) compared to S14 (25%). Summarily, then, though all models are optimized with the same set of POC products, they predict different magnitudes of the processes contributing to POC flux.

5.0 Discussion

In this study we embellish S14 to include parameterized processes known to influence POC flux. Below, we discuss which model components influence the model's prognostic skill, what is learned from adding and removing processes to our baseline food-web model, and what type of data are required for improving ocean carbon cycling models.

In this study, the modeled POC fluxes are relatively invariant across all nested models because the models are all optimized with the same dataset. This means that, first and

foremost, the addition of any mechanism (e.g., aggregation, differential mortality based on size or temperature) will not change the absolute magnitude of the POC export to a large degree, but rather the fractional contribution of the other mechanisms contributing to POC export. Because we are interested in understanding how well the modeled ecological mechanisms represent reality, it is futile to evaluate these models solely on the basis of their predictions for POC flux. The models must also be evaluated based on how realistic their mechanistic predictions are. Below, we consider how the different models vary 1) in the timing of their seasonal predictions for flux across sites, 2) in the their assumptions, and 3) in the distribution of their predicted grazing rates.

5.1 Model differences in the predicted seasonal cycles of mechanisms contributing to flux

It is well documented that increases in absolute POC flux are often a result of a shift in the mechanisms contributing to enhanced POC transit, as in Guidi et al, (2016) (where a significant fraction of the variability in carbon export can be predicted from viral genes in the oligotrophic ocean), and Busseler and Boyd (2009), where there is increased export of large particles following bloom periods. Given these findings, the most realistic models should produce algal flux fractions that are not constant throughout the year. In this study we refined the S14 parameterization that assumed a constant fraction of direct algal sinking relative to the primary production rate of large phytoplankton. When explicit aggregation for small and large particles is added (as in the 'Aggregation' suite of models) as a function of phytoplankton concentration, the timing and peaks of the aggregation flux change dramatically for the North Pacific and North Atlantic sites (compare Figure S2 with Figures

71

S3 and S4), compared to the S14 case. For all sites, enhanced POC flux coincides with an increase in larger cells, as evidenced by positive slopes for both the large fecal flux and aggregation fluxes. This means that when POC flux increases it is because large fecal flux and aggregation is also increasing. It is more plausible that these high latitude regions will experience changes in the fraction of aggregation throughout the seasonal cycle compared to a constant fraction of sinking cells year round, which is expected (and predicted by all models) at Station Aloha. These changes imply that the 'Aggregation,' 'Aggregation + Size Dependent Mortality,' and 'Aggregation + Temperature Dependent Mortality' models are most realistic.

5. 2 Consequence of removed model complexity

We tested a model ('No Size') that assumes all phytoplankton are the same size, and it is well known that plankton size varies. It is therefore unsurprising that the model produces values that vary considerably from the other models. The predicted global fraction of algal sinking to total flux is 60% for 'No Size' compared to 20%-40% for the other models. 'No Size' is highly sensitive to parameter changes relative to the other models, with a range of annually averaged POC export values from ~0 to 12 Pg C yr^{-1} depending on which parameter value is used (Supplementary Figure 4). These findings, together with the poor model metrics for 'No Size,' underscore the importance of distinguishing plankton types on the basis of size for improved model fidelity and performance.

5.3 Comparing modeled grazing rates with previous work

Another way to assess model performance is to compare the intermediate data products that the model produce to syntheses of field data. For example, all models evaluated predict herbivory rates on phytoplankton. Figure 6 shows model-predicted herbivory rates on small and large phytoplankton, which is analogous to grazing by microzooplankton and mesozooplankton, assuming the two-size food web model presented here. Previous work identified global mean values for the fraction of production consumed by microzooplankton to be 62% (Schmocker et al., 2013) and 67% (Calbet and Landry, 2004). The mean fraction of herbivory on small phytoplankton relative to total NPP is 70% for 'S14,' 88% for 'Aggregation,' 65% for 'Temp Mortality,' and 61% for 'Differential Mortality.' The fractional values are relatively large in the 'Aggregation' model because its agg_s parameter

is the smaller than agg_s for all other models, so nearly all of the small phytoplankton NPP is

available to be consumed because it has not become sinking aggregates.

Model predictions of grazing rates on large phytoplankton can be explored using the data synthesis of experimental mesozooplankton grazing data from Calbet 2001 that quantify the fraction of production grazed by mesozooplankton in order to compare the data with the models herein (Figure 6, second panel). The grazing rates were determined by gut pigment content and clearance rates from incubations. Because no direct matchups are available, we evaluate them on the basis of their frequency distribution functions. For consistency with the models we eliminate the few instances in the data set that exceed 1, because those are a likely consequence of uncertainties and bias within the grazing measurements for the two different techniques. The idea is that, although these grazing experiments were not

73

performed explicitly for model development, the distribution of their values can be exploited to compare global variability in estimates of grazing.

The experiments show a higher fraction of production grazed by mesozooplankton compared to any of the models: the mean for the Calbet (2001) data is 12%, compared to the mean for the 'S14' model (8%), the 'Aggregation + Temperature Dependent Mortality' model (9%), and the 'Aggregation + Size Dependent Mortality' model (5%). The higher values suggest food limitation for mesozooplankton (Calbet (2001), but may also imply that mesozooplankton (pelagic tunicates and cladocerans) are feeding on picophytoplankton (this idea is supported by Gorsky and Fenaux, 1998; Fortier et al., 1994). The mesozooplankton in the experiments exclusively fed on phytoplankton, whereas in the ocean they supplement their diet with smaller zooplankton; this could also account for the higher values reported in Calbet (2001). The models assume mesozooplankton only feed on large phytoplankton, and thus will underestimate the fraction of production consumed by mesozooplankton. With these considerations, it is likely the apparent discrepancy between the models and data is much less.

5.2 Biological implications for departures in parameter values from what is expected

The relative importance of mechanisms contributing to the biological pump is implied by the magnitude of change in a parameter's value from its expected value. Compared to the original S14 model, the algal fraction term (f_{alg}) quantifying the fraction of NPP going into direct algal sinking is nearly triple from what is formulated in the original

74

model, implying that direct algal sinking has a substantially more dominant role than was hypothesized. For all models, the fec_s term is similar to what is formulated in the original S14 model (10%) compared to the expected value of 20%. The fec_l term is slightly enhanced from what is expected, with every two-size class model predicted fec_l fractions between 20% and 35%. This supports the idea that carbon flux is more efficient in areas where there are less trophic steps between producer and predator(s).

The parameter optimizations are also used to test hypotheses arising from observations and ecological theory, and to address model plausibility. 'Aggregation + Differential Mortality,' and 'Aggregation + Temperature Mortality,' are considered to be equally plausible in their construction, so they must be evaluated based on other metrics. The 'Aggregation + Size Dependent Mortality' model tests the assumption that the most abundant organisms will be most susceptible to viral activity, and 'Aggregation + Temperature Mortality' assumes that temperature is the dominant control on non-grazing phytoplankton death. Globally, these different mortality paradigms account for a change of 0.5 Pg C yr^{-1} in TotEZ between the two models. The two mortality models produce similar performance metrics and global average statistics (Figure 3), but they vary in their predicted fractions for algal flux and large fecal flux (Figures 6, S3 and S4).

5.3 Data requirements for building improved models

Model performance can be diagnosed in order to identify requirements for advancing our mechanistic understanding of the biological pump from models driven by satellite observations. One strength of mechanistic modeling is that performance metrics are not limited to the model-data misfit for POC flux; other model components can be evaluated with data products (e.g., grazing rates) to assess the models across a range of situations.

We compare model outputs for the proportion of NPP that is grazed by small and large zooplankton with experimental work (Figure 6). This is a simple demonstration for how to overcome model-data scale issues when additional data is available but there are no direct space/time matchups. Grazing data is often reported as the number of prey consumed per day, which is not a possible model output. However, when the grazing rate data is normalized by the primary production in the experiments, the data product becomes directly comparable to a model prediction: the fraction of NPP that is grazed. In this case, the model and data distributions can be compared as long as the resolution of the data represents the resolution of the model. Future data products from the EXPORTS field campaign (Siegel et al., 2016) will provide improved matchups between zooplankton activity and NPP that may be useful for model calibration as described here. Additionally, future estimates of aggregate flux using optical techniques (e.g., the Underwater Video Profiler, Picheral et al., 2010), potentially deployed on autonomously profiling floats, may be useful to constrain ALG_{EZ}, as well as the proportion of aggregate flux from different size classes.

POC flux alone cannot explicitly quantify the magnitude of covarying surface ocean processes that produce it. Still needed are observations coupled to models in the surface ocean that link ecosystem structure (both the standing stocks of organisms from bacteria to

zooplankton, as well as their associated rates of production, herbivory, and mortality) to POC flux at depth across a range of environmental gradients, especially at dynamic sites with strong seasonal transitions, and in the Southern Hemisphere where there is especially sparse data. Improvements in the parameterizations of processes as well as quantitative descriptions how these parameterizations may shift across oceanographic state will especially be helpful to inform flexible model development. The analyses herein assume that the two-size class food-web model is appropriate to apply at every location and time around the world, but realistically a more flexible food-web structure that adapts to the environmental conditions across places and times will be a better representation of carbon cycling in the ocean. The ongoing EXPORTS field campaign (Siegel et al., 2016) will collect abundant observations within one region that can be used to develop a comprehensive understanding of the processes controlling the fate of photosynthetically fixed carbon, though on a limited scale. Until we have data that sufficiently characterizes the mechanistic links of surface ocean characteristics to POC flux at depth, we must rely on applying more primitive, yet still useful, models to the highest quality datasets available (e.g., the work presented herein), or by using model frameworks that capitalize on low uncertainty, high abundance environmental data (e.g., DeVries and Weber, 2017) that implicitly solve for the unknown values of biological pump components.

6. Summary

In this study we compared several plausible models built from S14 to evaluate ecological processes and carbon flux.

77

In particular,

- Our model outcomes underscore the complexity of the biological pump: depending on the time and place there can either be a strong direct, inverse, or no relationship between the balance of physical aggregation, small and large fecal flux.

- POC flux is not a strong constraint on the models; ecosystem measurements, such as the fraction of NPP grazed by zooplankton, are stronger constraints. Future studies will benefit from including other surface ecosystem measurements (including rates of zooplankton grazing and aggregation) in model evaluations.

Parameter Description			
Abbreviation	*Description*	*Units*	*Expected value*
W_{sink}	sinking speed of large cells	$m\,d^{-1}$	1
agg_s	term for aggregation by small particles on small particles	$m^4\,mg^{-1}\,d^{-1}$	none
agg_l	term for aggregation by any particle on large particles	$m^4\,mg^{-1}\,d^{-1}$	none
m_{ph}	mortality rate for small (mph_s) and large (mph_l) phytoplankton	d^{-1}	0.1
m_{phT}	mortality rate for phytoplankton based on temperature	d^{-1}	0.1
fec	fraction of grazed energy that becomes fecal flux for small (fec_s) and large (fec_l) prey	none	0.2

**Table 1.** This identifies the parameter names, symbols, and prior values for easy reference throughout the manuscript.
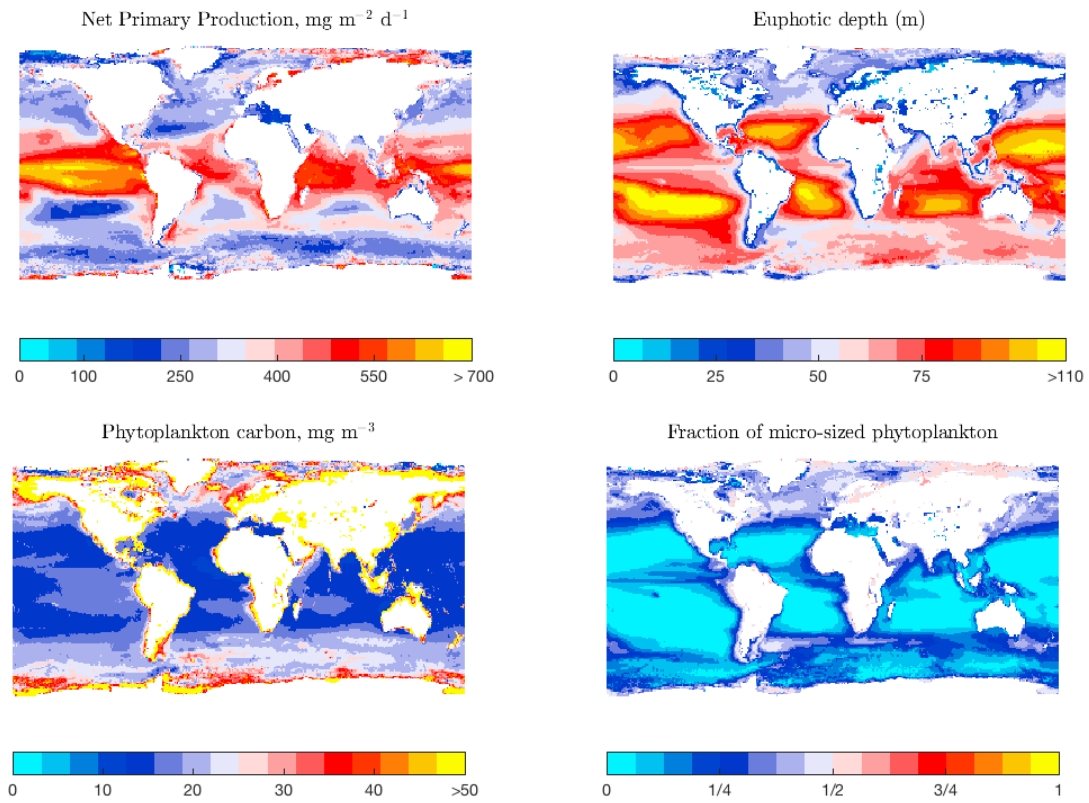
Figure 1. Annual climatologies for selected satellite data inputs into the model are shown. Phytoplankton carbon is in mg m^{-3} and approximated using particulate backscatter at 443nm.
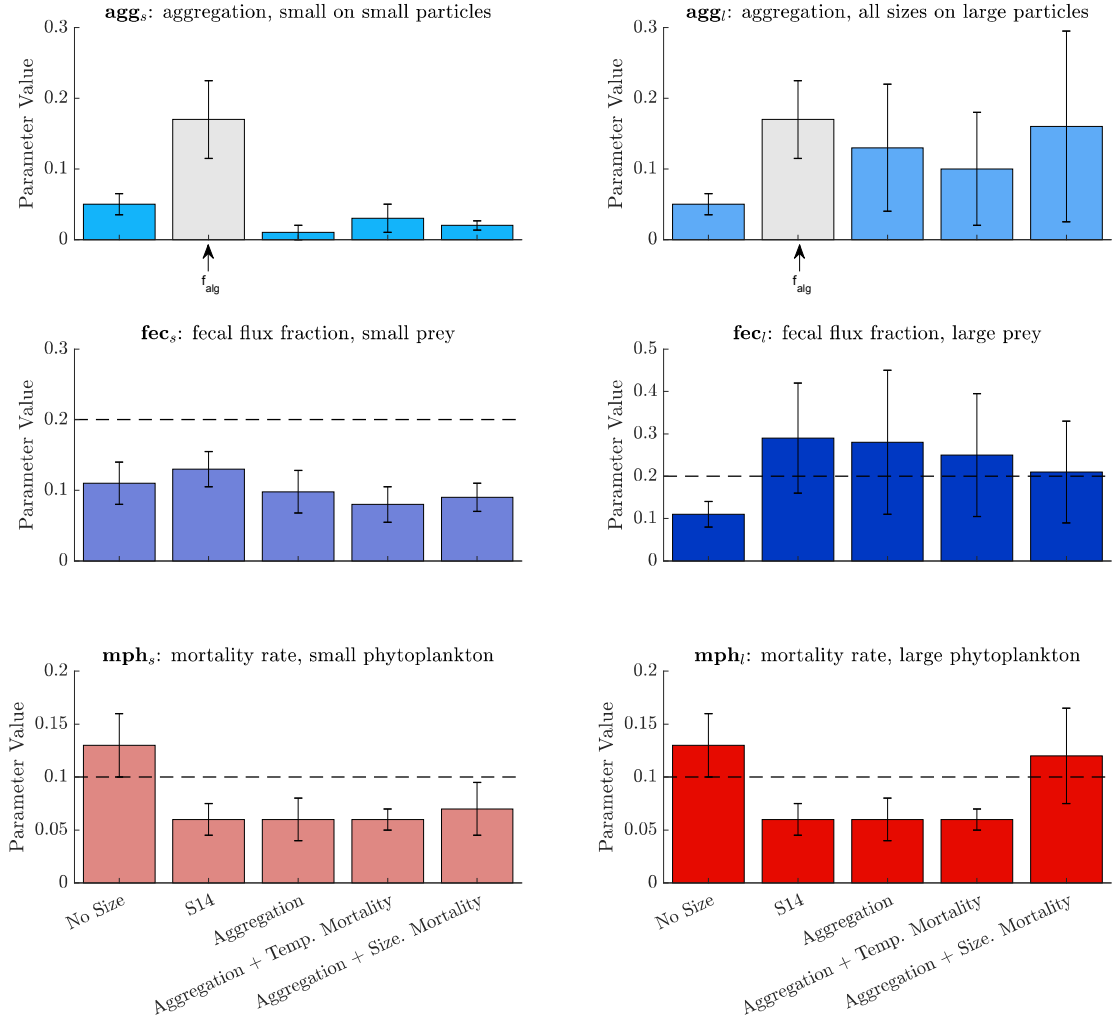
Figure 2. Parameter values and their interquartile ranges for the five models are shown. Expected values for parameters are shown with a dashed line when appropriate. Note the different y-axis limits for the fecal flux terms. The f_{alg} term is from the S14 model, and is a different parameterization than agg_s or agg_l.
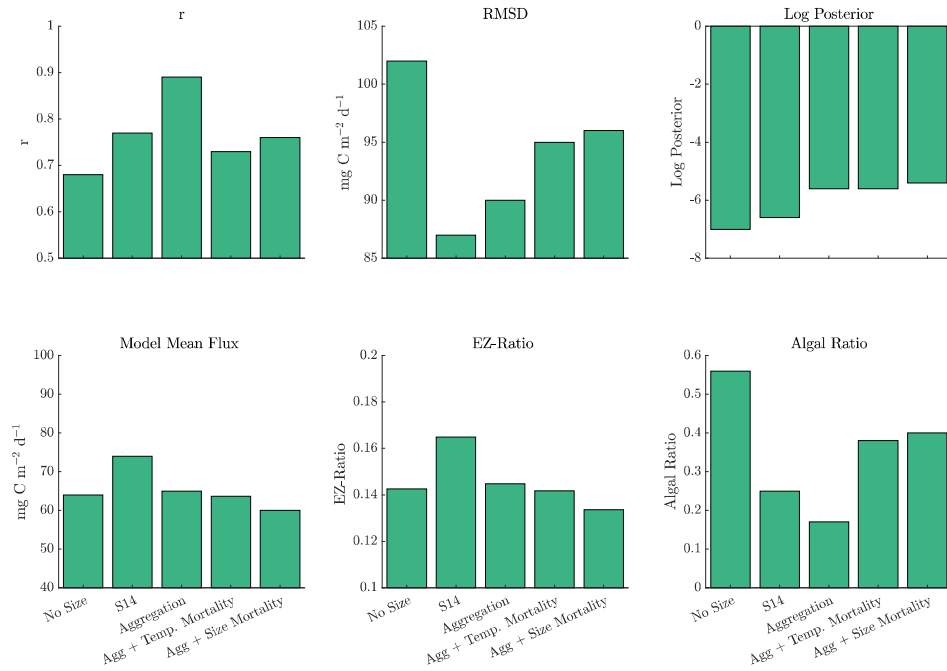
Figure 3. Basic statistics for the five models are shown, in order of increasing number of parameters (from 3 in the case of 'No Size,' to 6 in the case of 'Aggregation + Size Dependent Mortality'). Model mean fluxes are in mg C m^{-2} d^{-1}, and the EZ-ratio is the flux at the base of the euphotic zone divided by NPP. The algal ratio is the fraction of flux contributed by direct algal sinking and aggregation.

| model	r	l	Global POC	agg$_s$	agg$_l$	mph$_s$	mph$_l$	fec$_s$	fec$_l$
No Size	0.68	-7.0	6.1 ± 1.3	0.05 ± .03	0.05 ± .03	0.13 ± 0.10	0.13 ± 0.10	0.11 ± 0.05	0.11 ± 0.05
S14	0.77	-6.6	7.9 ± 0.7	0.29 ± 0.18	0.29 ± 0.18	0.04 ± 0.02	0.04 ± 0.02	0.10 ± 0.02	0.33 ± 0.19
Aggregation	0.89	-5.6	7.3 ± 0.5	0.01 ± .01	0.13 ± 0.09	0.06 ± .02	0.06 ± .02	0.10 ± .03	0.28 ± 0.17
Aggregation + Temp Mortality	0.73	-5.6	6.8 ± 0.6	0.03 ± 0.025	0.11 ± 0.08	0.024 ± .01	0.024 ± .01	0.08 ± .03	0.25 ± 0.15
Aggregation + Size Mortality	0.76	-5.4	6.3 ± 0.4	0.02 ± 0.05	0.16 ± 0.13	0.07 +0 .03	0.12 ± 0.05	0.09 ± 0.02	0.21 ± 0.12

Table 2. The model performance metrics and parameter mean values plus/minus half of the interquartile range are shown. In the case of 'Aggregation + Temp Mortality,' the mph values reported are for the coefficient in equation 5.
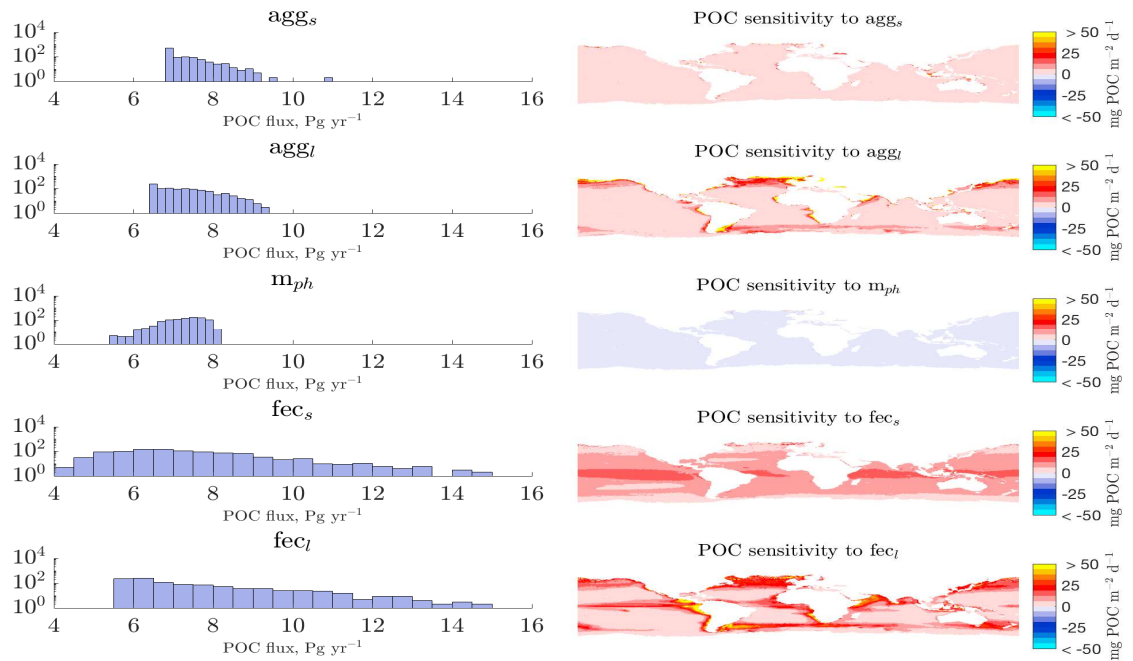
Figure 4. a) Histograms of annually averaged POC flux (Pg C yr^{-1}) for each of the parameter sensitivity tests are shown. b) Global maps showing model sensitivity in POC flux ('Aggregation' model) to changes in parameter value (dp) from the optimize parameter values (p) are shown. The values are given by a centered difference, or .5*(POC$_{p+dp}$- POC$_p$) + .5*(POC$_p$- POC$_{p-dp}$).
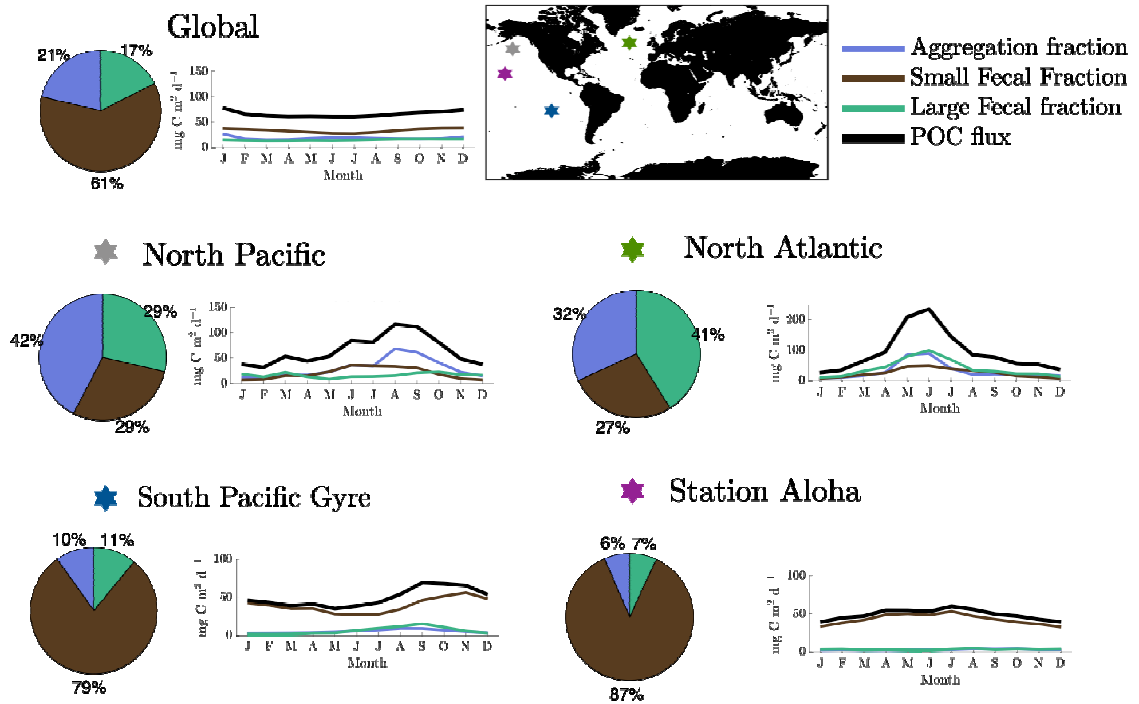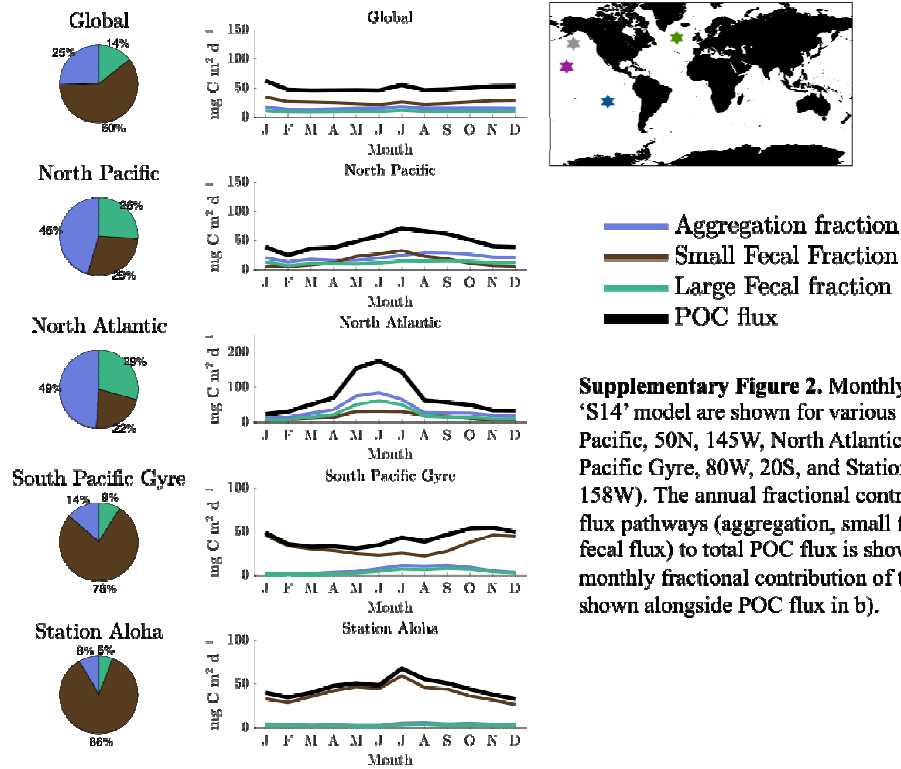
Figure 5. Monthly outputs from the 'Aggregation' model are shown for the global average as well as various locations (North Pacific, 50N, 145W, North Atlantic, 47N, 20W, South Pacific Gyre, 80W, 20S, and Station Aloha, 22N, 158W). The annual fractional contribution of three flux pathways (aggregation, small fecal flux, large fecal flux) to total POC flux is shown in the pie charts, and the monthly fractional contribution of these pathways is shown alongside them for each location.
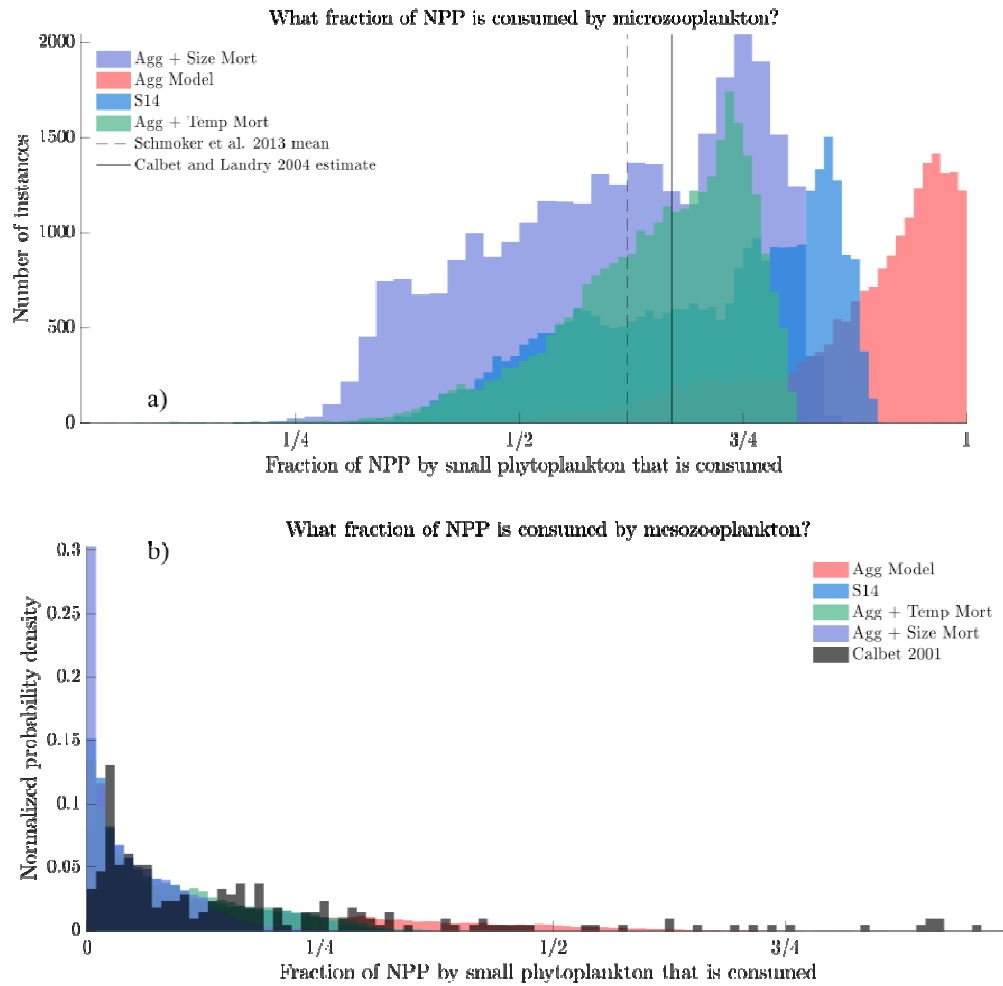
Figure 6. a) Annually averaged outputs for the grazed energy on small phytoplankton (Grz$_s$) divided by total NPP are shown for the four models that use size in their construction. Vertical lines indicate global averages from previous studies for comparison. b) Annually averaged outputs for the grazed energy on large phytoplankton (Grz$_l$) divided by total NPP are shown for the four models that use size in their construction for comparison with digitized data from Calbet, 2001, Figure 1A.
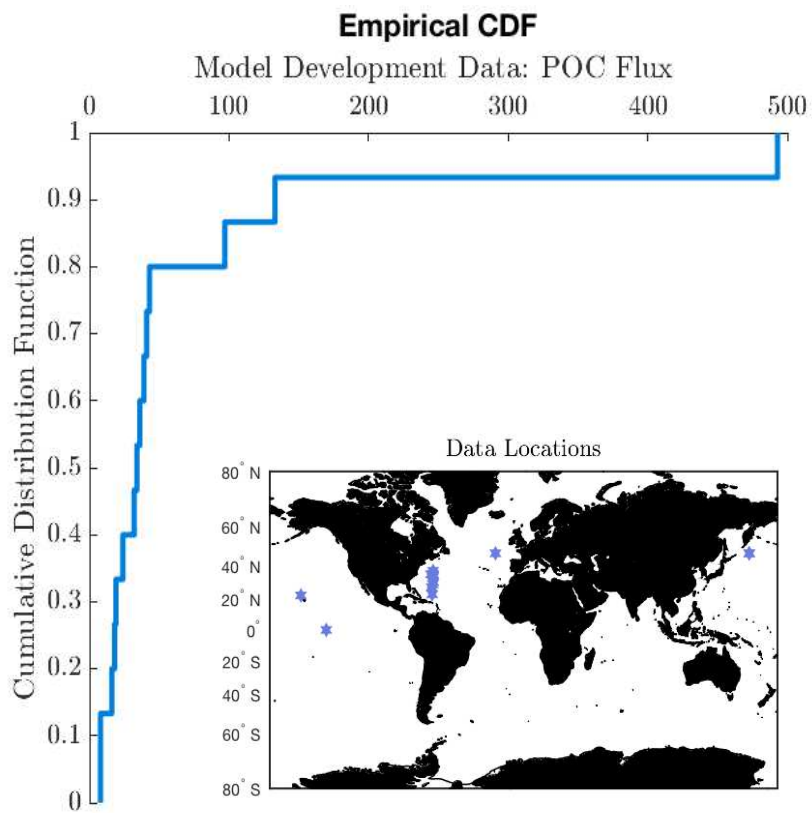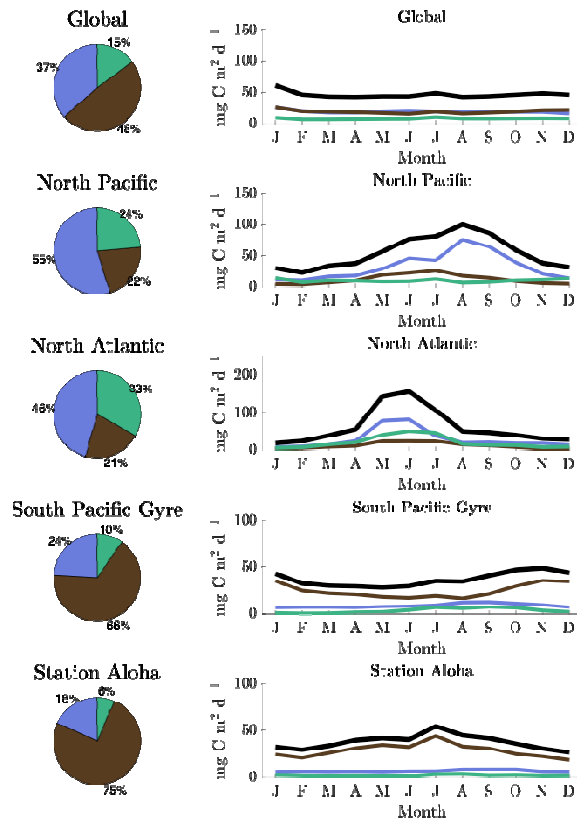
Supplementary Figure 1. The distribution of the data values is shown along with the data locations.

Supplementary Figure 2. Monthly outputs from the 'S14' model are shown for various locations (North Pacific, 50N, 145W, North Atlantic, 47N, 20W, South Pacific Gyre, 80W, 20S, and Station Aloha, 22N, 158W). The annual fractional contribution of three flux pathways (aggregation, small fecal flux, large fecal flux) to total POC flux is shown in a), and the monthly fractional contribution of these pathways is shown alongside POC flux in b).
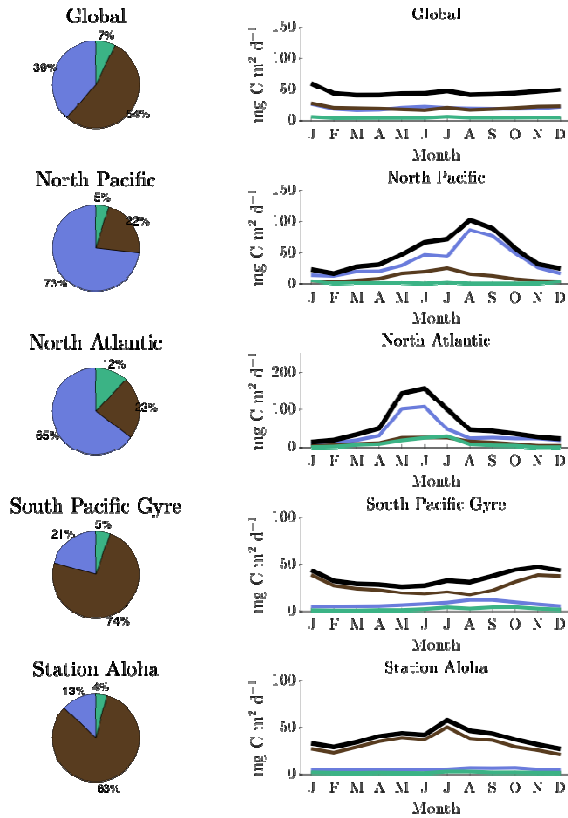
Supplementary Figure 3. Monthly outputs from the 'Size Differential' Mortality model are shown for various locations (North Pacific, 50N, 145W, North Atlantic, 47N, 20W, South Pacific Gyre, 80W, 20S, and Station Aloha, 22N, 158W). The annual fractional contribution of three flux pathways (aggregation, small fecal flux, large fecal flux) to total POC flux is shown in a), and the monthly fractional contribution of these pathways is shown alongside POC flux in b).

Supplementary Figure 4 . Monthly outputs from the 'Aggregation + Temperature Differential Mortality' model are shown for various locations (North Pacific, 50N, 145W, North Atlantic, 47N, 20W, South Pacific Gyre, 80W, 20S, and Station Aloha, 22N, 158W). The annual fractional contribution of three flux pathways (aggregation, small fecal flux, large fecal flux) to total POC flux is shown in a), and the monthly fractional contribution of these pathways is shown alongside POC flux in b).

— Aggregation fraction
— Small Fecal Fraction
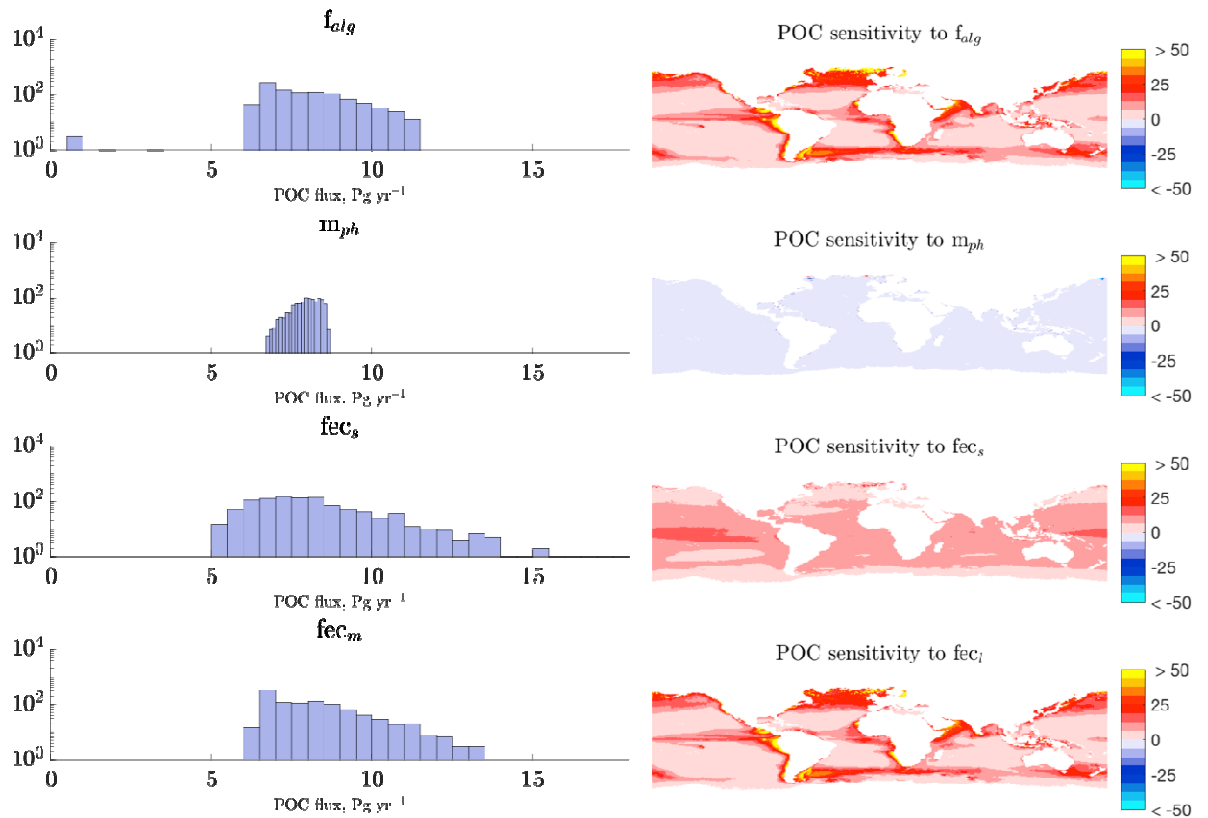— Large Fecal fraction
— POC flux
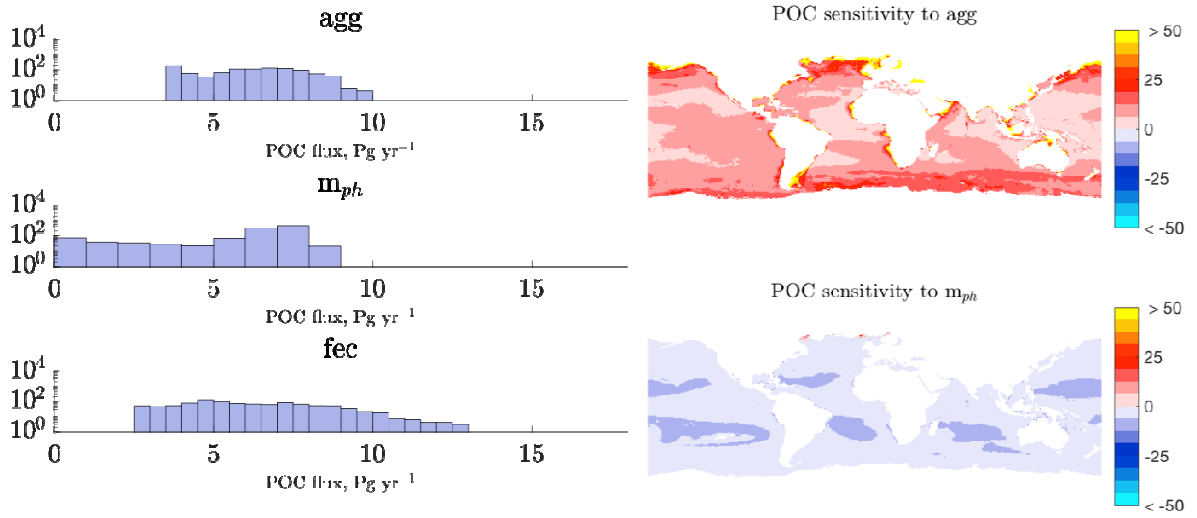
Supplementary Figure 5. a) Histograms of annually averaged POC flux (Pg C yr[-1]) for each of the parameter sensitivity tests are shown. **b)** Global maps showing model sensitivity in POC flux ('No Size' model) to changes in parameter value (dp) from the optimize parameter values (p) are shown. The values are given by a centered difference, or $.5*(POC_{p+dp} - POC_p) + .5*(POC_p - POC_{p-dp})$.
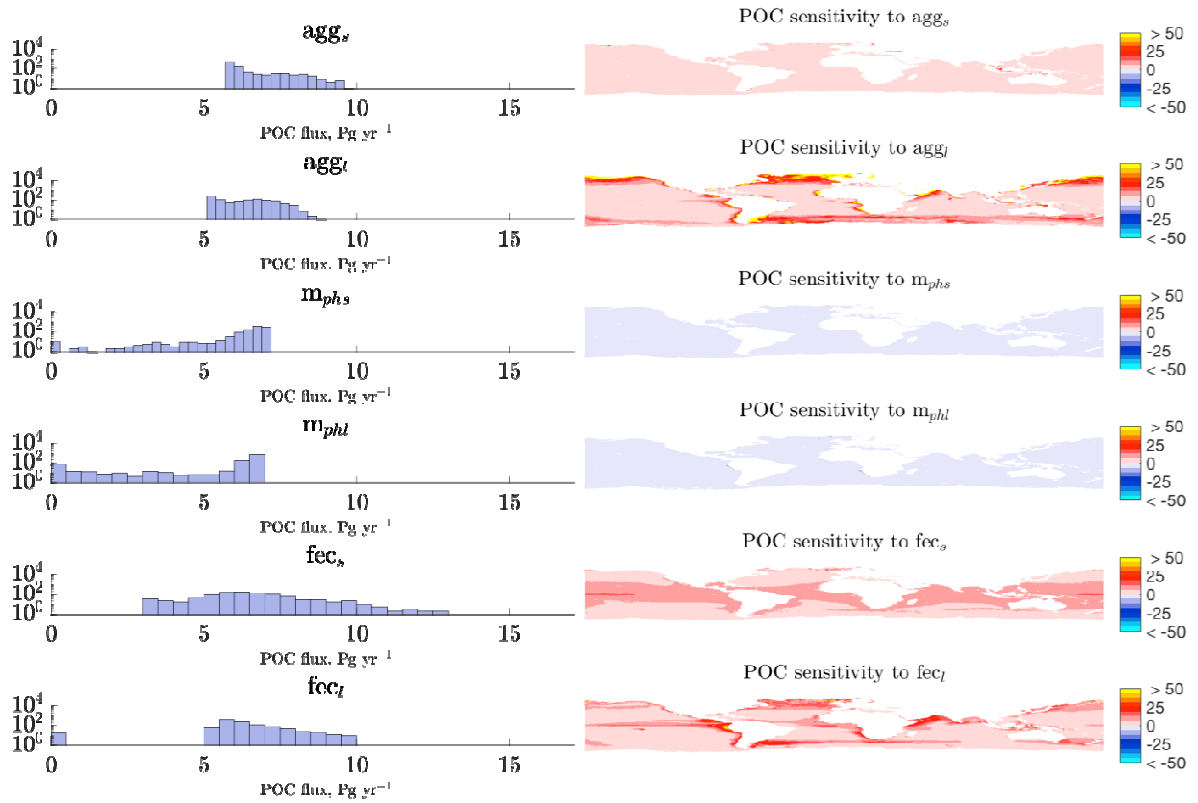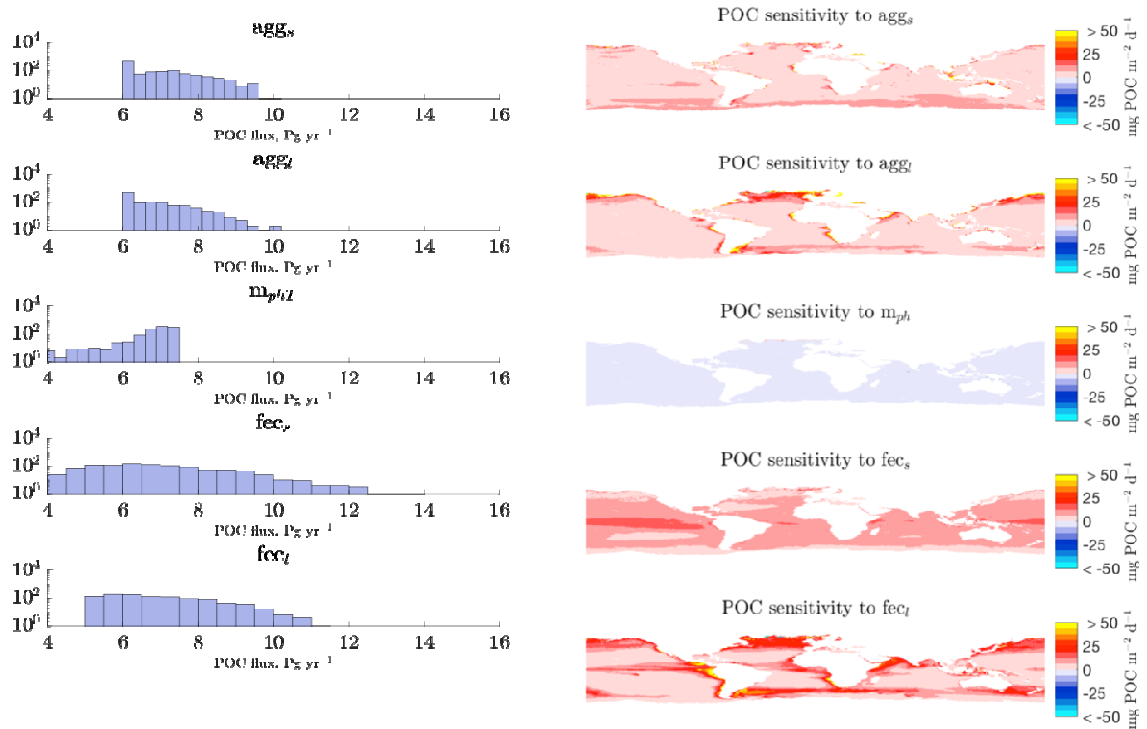
Supplementary Figure 6. a) Histograms of annually averaged POC flux (Pg C yr^{-1}) for each of the parameter sensitivity tests are shown. b) Global maps showing model sensitivity in POC flux ('Size Dependent Mortality' model) to changes in parameter value (dp) from the optimize parameter values (p) are shown. The values are given by a centered difference, or .5*(POC$_{p+dp}$- POC$_p$) + .5*(POC$_p$- POC$_{p-dp}$).

Supplementary Figure 7. a) Histograms of annually averaged POC flux (Pg C yr⁻¹) for each of the parameter sensitivity tests are shown. b) Global maps showing model sensitivity in POC flux ('S14' model) to changes in parameter value (dp) from the optimize parameter values (p) are shown. The values are given by a centered difference, or .5*(POC$_{p+dp}$- POC$_p$) + .5*(POC$_p$- POC$_{p-dp}$).

Supplementary Figure 8. a) Histograms of annually averaged POC flux (Pg C yr[-1]) for each of the parameter sensitivity tests are shown. b) Global maps showing model sensitivity in POC flux ('Temperature Dependent Mortality' model) to changes in parameter value (dp) from the optimize parameter values (p) are shown. The values are given by a centered difference, or $.5*(POC_{p+dp} - POC_p) + .5*(POC_p - POC_{p-dp})$.

References

Alldredge, A. L., & Silver, M. W. (1988). Characteristics, dynamics and significance of

marine snow. *Progress in oceanography*, *20*(1), 41-82.

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., & Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine ecology progress series*, 257-263.

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., ... & Boss, E. S. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, *444*(7120), 752.

Boyd, P., & Newton, P. (1995). Evidence of the potential influence of planktonic community structure on the interannual variability of particulate organic carbon flux. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(5), 619-639.

Buesseler, K. O. (1998). The decoupling of production and particulate export in the surface ocean. *Global Biogeochemical Cycles*, *12*(2), 297-310.

Burd, A. B., Moran, S. B., & Jackson, G. A. (2000). A coupled adsorption–aggregation model of the POC/234Th ratio of marine particles. *Deep Sea Research Part I: Oceanographic Research Papers*, *47*(1), 103-120.

Brooks, S.P. and Gelman, A. (1998). General methods for monitoring convergence of

iterative simulations. *Journal of Computational and Graphical Statistics*. 7, 434-455.

Calbet, A. (2001). Mesozooplankton grazing effect on primary production: a global comparative analysis in marine ecosystems. *Limnology and Oceanography*, *46*(7), 1824-1830.

Calbet A, Landry MR. 2004. Phytoplankton growth, microzooplankton grazing, and carbon cycling in marine systems. *Limnol. Oceanogr.* 49:51–57.

Chan, F., Pace, M. L., Howarth, R. W., & Marino, R. M. (2004). Bloom formation in heterocystic nitrogen□fixing cyanobacteria: The dependence on colony size and zooplankton grazing. *Limnology and oceanography*, *49*(6), 2171-2178.

Chen, B., & Liu, H. (2010). Relationships between phytoplankton growth and cell size in surface oceans: Interactive effects of temperature, nutrients, and grazing. *Limnology and Oceanography*, *55*(3), 965-972.

Coles, J. F., & Jones, R. C. (2000). Effect of temperature on photosynthesis□light response and growth of four phytoplankton species isolated from a tidal freshwater river. *Journal of Phycology*, *36*(1), 7-16.

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., ... & Takahashi, T. (2009). Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. *Journal of Marine Systems*, *76*(1-2), 95-112.

Ducklow, H. W., Steinberg, D. K., & Buesseler, K. O. (2001). Upper ocean carbon export and the biological pump. *OCEANOGRAPHY-WASHINGTON DC-OCEANOGRAPHY SOCIETY-*, *14*(4), 50-58.

Durkin, C. A., Estapa, M. L., & Buesseler, K. O. (2015). Observations of carbon export by small sinking particles in the upper mesopelagic. *Marine Chemistry*, *175*, 72-81.

Eppley, R. W., & Peterson, B. J. (1979). Particulate organic matter flux and planktonic new production in the deep ocean. *Nature*, *282*(5740), 677.

Falkowski, P., Scholes, R. J., Boyle, E. E. A., Canadell, J., Canfield, D., Elser, J., ... & Mackenzie, F. T. (2000). The global carbon cycle: a test of our knowledge of earth as a system. *science*, *290*(5490), 291-296.

Franks, P. J. (2002). NPZ models of plankton dynamics: their construction, coupling to physics, and application. *Journal of Oceanography*, *58*(2), 379-387.

Fortier, L. J. Le Fevre, and L. Legendre. 1994. Export of bio-genic carbon to fish and to

the deep ocean: The role of large plankton microphages. J. Plankton Res. **16:** 809–839.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, *399*(6736), 541.

Giering, S. L., Sanders, R., Martin, A. P., Henson, S. A., Riley, J. S., Marsay, C. M., & Johns, D. G. (2017). Particle flux in the oceans: Challenging the steady state assumption. *Global Biogeochemical Cycles*, *31*(1), 159-171.

Gorsky, G., and R. Fenaux. 1998. The role of *Appendicularia* in marine food webs, p. 161–170. *In* Q. Bone [ed.], The biology of pelagic tunicates. Oxford University Press.

Hansen, P. J., Bjørnsen, P. K., & Hansen, B. W. (1997). Zooplankton grazing and growth: Scaling within the 2□2,□µm body size range. *Limnology and oceanography*, *42*(4), 687-704.

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*(C9).

Lancelot, C., Hannon, E., Becquevort, S., Veth, C., & De Baar, H. J. (2000). Modeling phytoplankton blooms and carbon export production in the Southern Ocean: dominant controls by light and iron in the Atlantic sector in Austral spring 1992. *Deep Sea Research*

Part I: Oceanographic Research Papers, *47*(9), 1621-1662.

Landry, M. R., & Hassett, R. (1982). Estimating the grazing impact of marine micro-
zooplankton. *Marine biology*, *67*(3), 283-288.

Lee, Z., Marra, J., Perry, M. J., & Kahru, M. (2015). Estimating oceanic primary
productivity from ocean color remote sensing: A strategic assessment. *Journal of Marine
Systems*, *149*, 50-59.

Legendre, L., & Le Fèvre, J. (1995). Microbial food webs and the export of biogenic
carbon in oceans. *Aquatic Microbial Ecology*, *9*(1), 69-77.

Mahadevan, A., D'asaro, E., Lee, C., & Perry, M. J. (2012). Eddy-driven stratification
initiates North Atlantic spring phytoplankton blooms. *Science*, *337*(6090), 54-58.

Michaels, A. F., & Silver, M. W. (1988). Primary production, sinking fluxes and the
microbial food web. *Deep Sea Research Part A. Oceanographic Research Papers*, *35*(4),
473-490.

Pomeroy, L. R. (1974). The ocean's food web, a changing paradigm. *Bioscience*,
24(9), 499-504.

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*(9), 462-473.

Richardson, T. L., & Jackson, G. A. (2007). Small phytoplankton and carbon export from the surface ocean. *Science*, *315*(5813), 838-840.

Schlitzer, R. (2002). Carbon export fluxes in the Southern Ocean: results from inverse modeling and comparison with satellite-based estimates. *Deep Sea Research Part II: Topical Studies in Oceanography*, *49*(9-10), 1623-1644.

Schmoker, C., Hernández-León, S., & Calbet, A. (2013). Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. *Journal of Plankton Research*, *35*(4), 691-706.

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.

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., ... & Perry, M. J. (2016). Prediction of the export and fate of global ocean net primary production: the EXPORTS science plan. *Frontiers in Marine Science*, *3*, 22.

Sullivan, M. B., Weitz, J. S., & Wilhelm, S. (2017). Viral ecology comes of age. *Environmental microbiology reports*, *9*(1), 33-35.

Suttle, C. A., Chan, A. M., & Cottrell, M. T. (1990). Infection of phytoplankton by viruses and reduction of primary productivity. *Nature*, *347*(6292), 467.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean□driven atmospheric CO2 changes. *The carbon cycle and atmospheric CO2: natural variations Archean to present*, *32*, 99-110.

Wilson, S. E., & Steinberg, D. K. (2010). Autotrophic picoplankton in mesozooplankton guts: evidence of aggregate feeding in the mesopelagic zone and export of small phytoplankton. *Marine Ecology Progress Series*, *412*, 11-27.

Yoo, A. B., Jette, M. A., & Grondona, M. (2003, June). Slurm: Simple linux utility for resource management. In *Workshop on Job Scheduling Strategies for Parallel Processing* (pp. 44-60). Springer, Berlin, Heidelberg.

III. **Testing relationships between the particle size distribution and flux in the global ocean**

Introduction

The transit of particulate organic carbon (POC) from the surface ocean to ocean depths fuels deep-sea metabolism and exerts a control over atmospheric CO_2 levels (Archer et al., 2000; Kwon et al., 2009), making POC flux observations central to many theoretical, experimental, and modeling studies from cellular to global scales (e.g., Karl et al., 1988; Steinberg et al., 2008; Burd and Jackson, 2009; Mari et al, 2017; Siegel et al., 2016). Ideally, observations of POC flux are numerous and well calibrated, so that they can be attributed to variables of interest throughout appropriate places, depths, and time. However, the only direct method for determining POC flux is sediment traps, which are riddled with uncertainties (Michaels et al., 1990; Buesseler et al., 2007) and have a sparse worldwide distribution (Bisson et al., 2018), making it difficult to compare fluxes from traps to other measurements of interest. Thus, new methods are needed to address the need for higher resolution POC flux estimates.

One such method uses theory to link the standing stocks of the aggregate size distribution to the flux of POC throughout the water column. POC flux can be calculated if

the particle size distribution is known (given by n, where n is the number concentration of particles within a given size range (in theory, from 0 to infinity; in practice, from 250um to 1.5mm), according to

$$Flux\ (D) = \int_0^\infty n(D)m(D)w(D)dD \tag{1}$$

where m is the mass of an individual particle (mg C), D is the diameter of the particle (mm) and w is its sinking velocity (m d^{-1}), as first introduced in Guidi et al. (2008). The mass of a particle is given by $m(D) = \frac{\pi \rho}{6} D^3$ where ρ is average density. Both the mass (m) and the sinking speed (w) of a particle (in the Guidi et al., 2008 study, Stoke's law is used) can be modeled as a power law, so their product can also be represented as a power law, as in Guidi et al, (2008, 2009, 2015), with unknown parameters A and b.

$$wm = AD^b \tag{2}$$

POC flux is the product of the number of particles in each size (diameter) class, the mass, and the sinking speed across all diameters (D). The exponent b gives information about a particle's fractal dimension, which quantifies particle shape and can be used to calculate particle sinking speeds, as in Guidi et al., (2008). The fractal dimensions of marine aggregates are commonly between 1.125 and 1.3 (Alldredge, 1998) but can be as high as 3.75 (Jackson et al., 1997). Particle shape changes with particle type, size, and from both abiotic and biotic interactions throughout the water column (Naselli-Flores et al., 2007; Stanca et al., 2013; Alldredge et al., 1993; Murray and Jackson, 1992). The prefactor on the

power relationship (A) does not mechanistically indicate anything; rather, changes in the magnitude of A will adjust the magnitude of POC flux for a given number size distribution.

This method has been applied to observations from the Underwater Vision Profiler (UVP), which is mounted on a CTD rosette and collects data throughout the water column. The UVP has an order of magnitude greater amount of observations compared to traps, and unprecedented resolution within the water column for a range of particle sizes (90μm- 2cm), showing great promise to contribute estimates of POC flux for diverse applications (Picheral et al., 2010; Guidi et al., 2008, 2016; Biard et al., 2016; Jouandet et al., 2014, and others). Is it also easy to deploy, generates data rapidly, enumerates and sizes particles without destroying them or altering their relevant characteristics, and is constantly being improved upon (Gorsky et al., 1992; 2000). Yet, the required parameters to calculate POC flux from UVP particle size distribution data have been assumed to be constant since their first use (Guidi et al., 2008), although it is probable they will vary around the world's oceans and with depth (Jouandet et al., 2011).

In the original study of Guidi et al., (2008), 118 UVP profiles from the Eastern Atlantic and South Pacific were optimized to retrieve a single magnitude for A (12.5±3.4) and b (3.81± 0.7) that best fit fluxes from sediment traps deployed from 100m to 1000m. In this study we use 3246 UVP profiles from the global ocean with estimates of POC flux across depth to ask: how might A and b vary for different depths and/or different oceanographic conditions?

Given that (1) POC flux can be calculated from the particle size distribution and the two parameters A and b, and (2) b quantifies the aggregate fractal dimension and particle sinking

speeds throughout the water column, we expect that A and *b* will differ as a function of environmental characteristics and depth. To test this, we optimize the largest amount of UVP profiles available using observed and modeled POC flux in order to identify controls on A and *b*. The results will be useful for future modeling studies requiring abundant flux measurements worldwide under a variety of conditions.

2.0 Data and Methods

The overall goal of this study is to leverage existing, flux estimates with sufficient spatial resolution to enable POC flux estimates from UVPs. POC fluxes from sediment traps or the [234]Th technique will not work here because there aren't enough matchups over the relevant depth (below surface to depth) to rigorously assess the parameters needed for POC flux. Instead we use POC fluxes obtained using a tracer inversion approach from DeVries and Weber (2017) because they are highly depth resolved and are available across the entire globe; thus, virtually every UVP measurement will have a matchup product in space. Given the matchups in space, there are abundant UVP profiles representing instantaneous snapshots in time that can be averaged over the course of a year to match the annual average outputs of the DeVries and Weber (2017) model. The DeVries and Weber (2017) model produces annually averaged POC fluxes derived within an ocean circulation model to more realistically retrieve what POC flux has to be given observed oxygen, DOC, and sediment-trap POC flux data. More details are in section 2.1.

2.1 Tracer based calculations of POC flux

DeVries and Weber (2017) present methods to implicitly solve for POC flux at 24 discrete depths (from 50 to 5500 meters) throughout the ocean (2 degree grid, 24 vertical levels) using satellite data and in situ observations across depth. The model uses information on the upper ocean processes contributing to flux (i.e., from satellite observations) as well as the interior processes promoting flux decay (i.e., from concentrations of dissolved organic carbon and dissolved O_2, and POC flux from Lutz et al., (2007)). The advantage of this approach is that it resolves changes in flux with depth (hereafter referred to as 'DW17' products) within the water column, so it can be compared with fluxes derived from UVP data. DW17 fluxes are annually averaged and gridded to 2-degree resolution, which necessarily requires that the comparison with UVP products should be also annually averaged and binned to 2-degrees. We only use depths below the euphotic zone, defined as the 1% light level and determined using surface chlorophyll measurements from satellite data as described in Morel et al., (2007). The depths of interest are centered bins at 137 m, 190 m, 250 m, 500 m, and 2500 m. We choose these depths because the model is not seasonal, so we want to avoid comparisons where there are especially seasonal changes (i.e., in the surface ocean) in UVP data. Of course, there are seasonal changes with depth as well, and we account for this by suitably averaging over a sufficient number of profiles to minimize impacts from seasonal variations (details in 2.2). Finally, the globally integrated annual average for DW17 fluxes is 9.1 ± 0.2 Pg C yr^{-1}, which has considerably low uncertainty compared to other models for POC flux. Implications for using this modeled POC product, rather than direct POC flux observations, are reviewed in the discussion.

2.2 UVP Data acquisition and data treatment

104

The UVP was designed to quantify large (>100 μm) aggregates and zooplankton throughout the water column at data acquisition rates comparable to CTD profilers (more details in Picheral et al., 2010). There are 5 UVP versions currently, all of which can capture images similar to those acquired from remotely operated vehicles or manned submersibles, but with considerably less effort and expense. Different versions of UVPs will capture varying sizes of particles. The lower size class threshold is fixed by optical resolution whereas the sampling volume determines the upper size limit. The newer UVP generations (3,4,5) can image particles as small as 90μm and as large as 2cm to 1000m depth (the UVP5 is rated to 3000m) whereas the older UVPs (1,2) are limited to a range of 250 μm to 1.5mm to 1000m depth.

For this study, 5234 UVP profiles were downloaded from ECO-TAXA (<http://ecotaxa.obs-vlfr.fr>, accessed 20 July 2018) particle module for use in this study (Picheral and Irisson, 2017). The particle profiles are binned to 5m depth intervals, and all particle sizes between 250μm and 1.5 mm are considered in order to use the sizes coherent across all UVP versions.

We organize the data to match the space/time/depth model outputs as closely as possible. In particular, we:

1) Remove data in shallow waters (<500 m) because the satellite products used to drive the SIMPLE-TRIM model are influenced by terrestrial input. This reduces 5234 profiles down to 3885.

105

2) Use data equatorward of 65 degrees latitude because the high latitudes have known issues related to cloud cover and poor coverage. This reduces the profiles to 3553.

3) Grid UVP profiles to the 2-degree resolution of the modeled POC fluxes. This bins the profiles into 475 unique bins based on location.

4) Remove bins with less than 6 profiles in a given bin, which reduces the total bins to 259 (Figure 1) with 3246 profiles. This step is done so that the highly variable particle size distribution is sufficiently averaged.

For each bin, the average size spectrum is calculated according to $n = \Delta C / \Delta s$ where ΔC is the change in particle biovolume with increasing size class and Δs is the given size range for a particular bin, in millimeters. The particle size spectra with two UVP bins closest to a chosen depth are averaged for comparisons with model output.

2.3 Honjo deep trap data set

We also compare UVP flux estimates to deep sediment trap data from the Honjo et al., (2008) compilation. We bin UVP and trap locations at one-degree resolution, which reduces the matches for the 2-degree bins from 456 to 61. We keep matches that have depth differences with the UVP profiles of less than 50 m and with at least 3 observations, which reduces the matchups to 18 sites. Importantly, the Honjo dataset does not provide vertically resolved fluxes. However, it does provide direct measurements of deep flux (500-3000m), which can still be used to test parameter values.

106

2.4 Flux model optimizations

We optimize the two parameters (A, *b*) in the flux model (equation 1) using a minimization procedure that reduces the misfit between the UVP model and the POC flux products. This is quantified by a cost function (equation 3) which also takes into account how the optimized parameters deviate from the original Guidi et al., (2008) parameters (A=12.5 ± 3.4; *b*= 3.81 ± 0.7).

$$\text{cost} = \sum_{i}^{N_{bin}} \left[\log \left(\frac{Model}{POC\ flux\ products} \right)^2 \right] + \frac{|A-12.5|^2}{3.4} + \frac{|b-3.81|^2}{0.7}. \tag{3}$$

The deviations from the original parameter values are normalized by their standard deviation reported in Guidi et al., (2008).

Surface ocean characteristics influence particle type and particle magnitude so we also optimize the parameters for grid cells based on surface ocean ecosystem characteristics. We subsample the bins based on 'high' and 'low' values for net primary production (NPP) and sea surface temperature (SST). We use the CAFE satellite data product (Silsbe et al., 2016) because it showed the best performance across all of the datasets analyzed in Bisson et al. (2018), and because it showed best performance with in situ NPP measurements compared to other NPP models (Silsbe et al., 2016). CAFE climatologies of NPP are available from 1997-2008. We use NOAA SST products (Reynolds et al., 2007) binned to 1-degree and averaged over 1997-2008 for consistency with the NPP product. The idea is that changes in NPP covary with changes in particle composition and ecosystem behavior, which determines

the quantity and shape of particles available for export. Sea surface temperature also differentially affects phytoplankton growth and heterotrophic metabolism (Huntley and Lopez, 1992). At higher temperatures, heterotrophic metabolism slightly exceeds the rate of phytoplankton growth; this effect has been hypothesized to affect the amount of NPP available for export (Cael et al., 2017).   For each NPP and SST optimization, we define 'high' as bins with data values that exceed the 75% quantile, and 'low' as bins with data values less than the 25% quantile.

Model optimizations are compared on the basis of their cost function value, the Pearson correlation coefficient (R), and Root Mean Squared Differences (RMSD).

3. Results

Results from the model optimizations are shown in figures 2-8, and model performance metrics are shown in table 1. Figure 2 shows particle size spectra, the POC flux transfer function, and POC flux for a range of depths. Figure 3 and 4 show how these attributes vary depending on the surface ocean environmental conditions. Figures 5-8 and table 1 show model performances for the various optimization experiments conducted in this study.

The particle number size spectra (Figures 2-4, top panels) show a higher abundance of smaller size particles compared to large particles, as expected across the world's oceans.

POC fluxes are obtained from the particle number size spectra and transfer function (AD^b, 2nd panel). Higher fluxes are predicted for smaller sized particles (Figures 2-4 bottom panels) across all depths. Given a particular particle size, the optimized fluxes are lower at depth and higher toward the surface. The transfer function that converts the particle size spectra to flux systematically varies with size and depth. Implicit in the power law relationship is the assumption that larger diameters require a higher conversion factor for POC flux. If A and b were the same with every optimization experiment conducted there would be no visible difference in the relationships shown in the second panel for Figure 2-4. There are deviations in this transfer function across the five depths considered with increasing particle size.

Depending on the surface ocean NPP, the predicted shape and magnitude of resulting fluxes and of the transfer function varies (Figures 3,4). For the two depths shown (centered at 137 m and 500 m) across two conditions of NPP there are differences in the transfer functions comparable in magnitude to the differences observed across all depths in Figure 2. For depths of 500 m, the predicted fluxes are the same independent of surface NPP value (Figure 3, lower panel). The largest changes in predicted POC fluxes are for different depths rather than difference surface NPP values. For high NPP bins, there is slightly larger flux across all size classes, although this is likely not statistically significant.

Surface temperature differences also influence the parameter optimizations (Figure 4). In contrast to the NPP case, the particle number spectra do show separation for high and low temperature areas. The greatest concentration of particles occurs in low temperature surface

waters, and the lowest concentration of particles is in deep waters with high sea surface temperatures. The transfer function is highest for the high temperature case at 137 m, and is apparently dissimilar to the other situations. The resulting fluxes are bounded by end members for low temperature bins at 137 m (high flux), and by high temperature bins at 500 m (low flux).

The retrieved parameters and model performances reveal systematic differences with depth and systematic departures from the original Guidi et al., (2008) parameters. The value of A decreases with increasing depth (Table 1), which coincides with decreasing POC fluxes and decreasing RMSDs (which range from 3-135 mg C m^{-2} d^{-1}). The changes in POC flux are due to the changes in A moreso than the changes in b, which are minimal. Correlation coefficients range from -0.2 to 0.8 (for low temperature surface water optimizations and 2500m optimizations, respectively). The model performance is equivalent on the basis of the cost function (with values from ~ 0.3-0.5) for all optimizations except shallow low temperature waters and deep sediment traps (with values of 1.8 and 2.8, respectively).

The original parameters overestimate flux at depth compared to fluxes generated from both the DV17 products and the Honjo dataset (Figure 5, 6). This is a consequence of higher A values from the original parameters compared to lower A values predicted from this study. The optimized values of b are generally within 8% of the constant value of 3.8 used by Guidi et al. (2008) resulting in the parallel lines apparent in Figure 5. The flux differences are constrained between 0 and 100 mg C m^{-2} d^{-1}, and the highest discrepancy is observed at the 500m depth optimization (Figure 7). Future work will need to assess error and uncertainties for the entire analysis presented herein in order to statistically discriminate differences

110

between the transfer functions and parameter values retrieved for depth or surface environmental conditions. The implications of this are reviewed in the discussion.

4. 0 Discussion

In this study we have examined how the required parameters A and b to produce POC flux from particle size spectra vary across depth and surface ocean environment. This study is a first step proof of concept approach to identity differences in A and b. Error estimates on the parameter retrievals are not examined and discussed here; future work will need to rigorously account for the differences between the transfer functions. Below, we discuss implications for parameter variation as well as ideas for future work to improve the analysis herein.

4.1 Sources of parameter variation

Why do parameters vary? The original study of Guidi et al., (2008) used 'b' values to estimate fractal dimension of particles, and showed that differences in the value of 'b' meant that the particle shapes and sinking speeds were changing. We identified parameter values that are optimum for flux that are different depending on the depth used. The source of the variation in flux and the shape of the transfer function is driven by changes in A and not b. The magnitude of the parameter values is determined by the formula for the cost function, and more work is needed to address parameter uncertainties and sources of error. A constant b implies that particle shape is not changing with depth, which is unlikely. If only the

number of particles were changing with depth, a similar transfer function might be expected because the particle size spectra itself would be attenuating. The different transfer functions imply the particle fluxes are differentially changing as depth increases with respect to size. A particle could change shape from both biotic (remineralization) and abiotic (physical shear and disaggregation) processes. We find that A decreases with depth using the DW17 products. When using the available deep sediment trap observations we also find diminished values of A and b from the original Guidi et al (2008) case, which supports the idea that A and b will vary with depth.

The formula for POC flux presented herein is itself problematic because it imposes a monotonic relationship between diameter and flux for a specific size range (i.e., AD^b increases with increasing rate as particle diameter increases). The shape of the resulting flux profile therefore requires accurate counts of the particle number spectra for a wide range of particle sizes so that N*AD^b yields reasonable estimates of POC flux. In this study, the POC fluxes decrease for larger diameters because there are few particles of large size. However, because there are no observations of particles less than 250 μm, the apparent maximum in particle flux is bounded by this end member. In the real ocean there are abundant particles less then 250 μm in size and greater than 1.5mm. This effectively truncates the theoretical curve of POC flux as a function of size, which should have minimums at both the low and high particle sizes. In this study, only a subset of that curve is sampled (from 250μm to 1.5mm), so we observe a nearly constant negative slope with increasing particle size, rather than a curve with a clearly defined maximum value. The results herein imply no particle flux minimum for smaller particles sizes, which is unrealistic given that, at some threshold,

small particles will be buoyant and not constitute sinking flux. Future generations of UVPs will resolve lower size classes, which will give a greater fidelity of the POC flux –particle size relationship, and will undoubtedly change the optimized values of A and *b*.

4.2 Choice and consequence of POC flux product

There are limited POC flux products available to use for assessing the particle size – POC flux relationship. In this study we chose a calibrated model output to optimize a model for POC flux. When should one model be used to optimize a different model? The required assumption is that a model product faithfully represents the average particle flux environment. This is the same assumption required of any given observation used to develop a steady state model, and it has been shown that not all observations represent the average conditions of the ocean (Bisson et al., 2018). Ideally there would be sufficient number of sediment trap observations to use as direct matchups in the work presented herein, with minimum uncertainties associated from spatial temporal variability of the system (Bisson et al., 2018) or from collection (Buesseler et al., 2007). As currently available sediment trap observations for POC flux are not abundant enough to use as direct matchups in the work presented herein, we chose a model development product that is consistent with the paradigms for POC flux in the ocean (higher POC flux in the surface, lower POC flux at depth) and built from abundant, low uncertainty observations. We do so to investigate the hypothesis that A and *b* will vary, rather than to produce a statistical relationship that is globally applicable now. This approach invites future analyses to combine abundant matchups across different methods for flux (including the UVP) so that the systematic relationships between A and *b* identified herein can be evaluated more completely. The

ongoing EXPORTS field campaign (Siegel et al., 2016) will produce these flux products, which may enable improved estimates of A and *b* as well as their covariations with environmental parameters.

4.3 External sources of validation

Although direct matchups with existing POC fluxes are limited, we can compare the flux distribution of the work presented herein with the distribution of an existing POC profile dataset (Cael and Bisson, 2018). In particular, we consider how the profiles of the fluxes vary using the optimized and original parameters in order to assess differences in the POC flux attenuation rates. POC flux attenuations rates are commonly assessed by fitting a power law relationship ($Flux(b) = C \times z^{-b}$) to the POC flux profile, in order to determine the

parameter value quantifying the loss rate (Martin et al., 1987). We use a profile data set (Cael and Bisson, 2018) that consists of 722 POC flux profiles from either sediment traps or the ^{234}Th technique at depths up to 4600 meters collected over the last four decades. The profiles have between 3 and 14 observations and were deployed between 1 and 425 days. We fit a power law ($Flux(b) = C \times z^{-b}$) to each profile (following Martin et al., 1987) if there

are at least 3 observations with depth. We then compare how the power law parameter varies across for the optimized and un-optimized case compared to the POC flux observations (Figure 8). We find that the original parameters from Guidi et al., (2008) predict a lower attenuation rate for POC flux than the optimized case, which is due to the overestimation of POC fluxes with depth using the original Guidi et al., (2008) parameters. The distribution of the optimized case most closely represents the POC flux observations, which provides an

external source of validation to the work presented herein. Moreover it demonstrates that accounting for depth-variable A and *b* parameters improves the fidelity of the UVP flux model.

Future work will benefit from including more rigorous estimates for parameter uncertainties and optimization routines that take into account the particle size spectra variation within each bin. The results herein imply differences in the shapes of settling particles and on the activities that might especially influence them, but cannot say how and why the particles are changing shape. Improvements in optical signatures of flux with depth will allow this analysis to move beyond the assumption homogeneous POC flux.
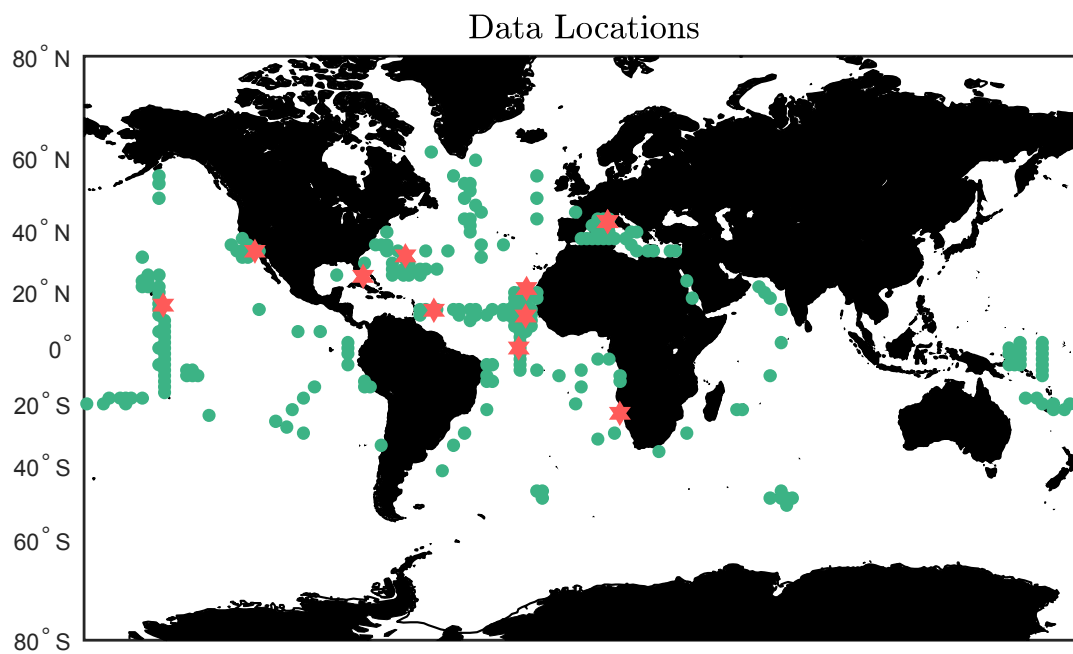
Figure 1. Data locations for the UVP profiles (green) and deep sediment traps (red) used in this study.
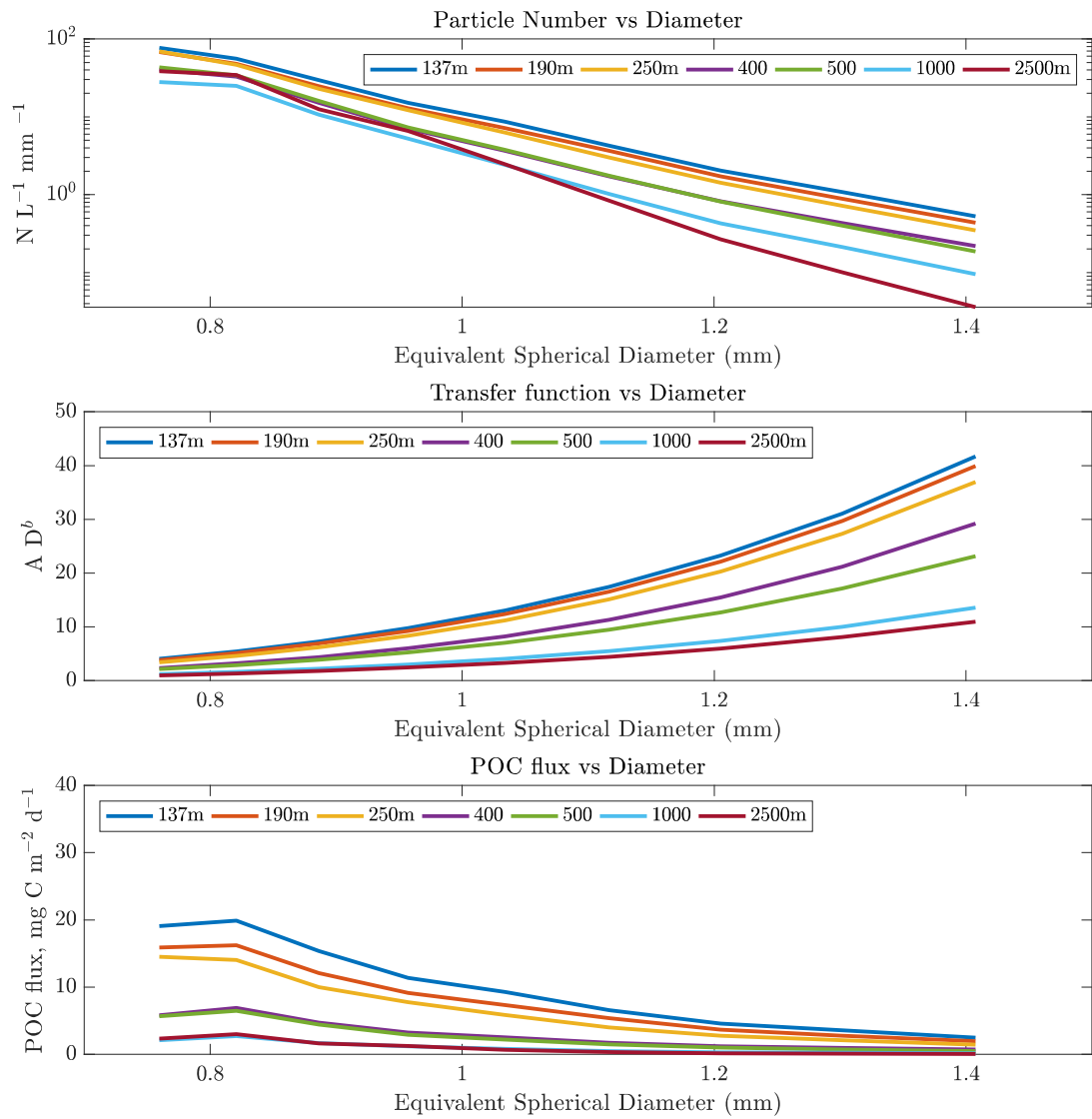
Figure 2. Particle size spectra (top panel), POC transfer functions, and POC fluxes are shown for different depths and equivalent spherical diameters.
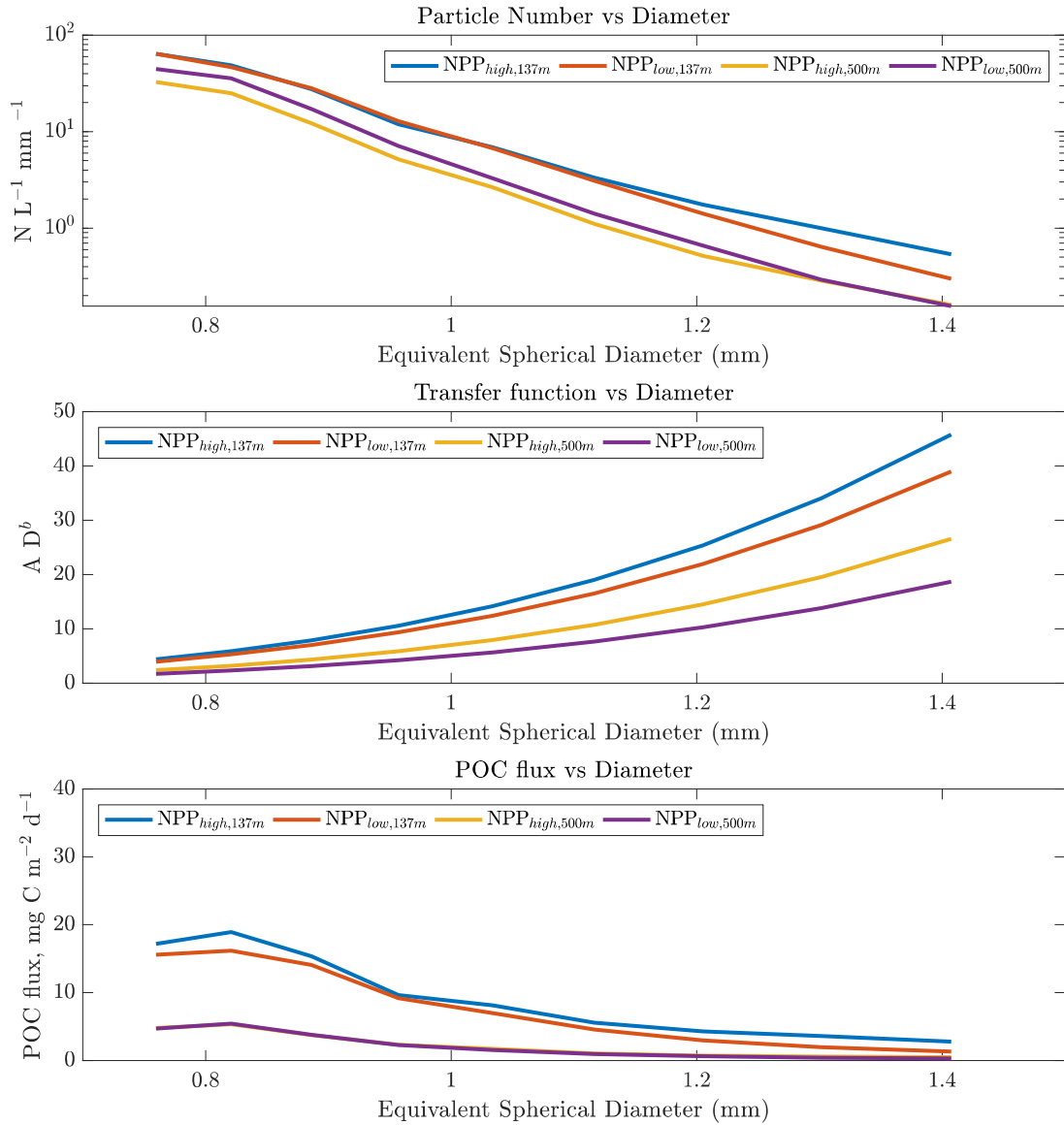
Figure 3. Particle size spectra (top panel), POC transfer functions, and POC fluxes are shown for different surface NPP values and equivalent spherical diameters. Note that the purple and yellow lines in the bottom panel are equivalent.
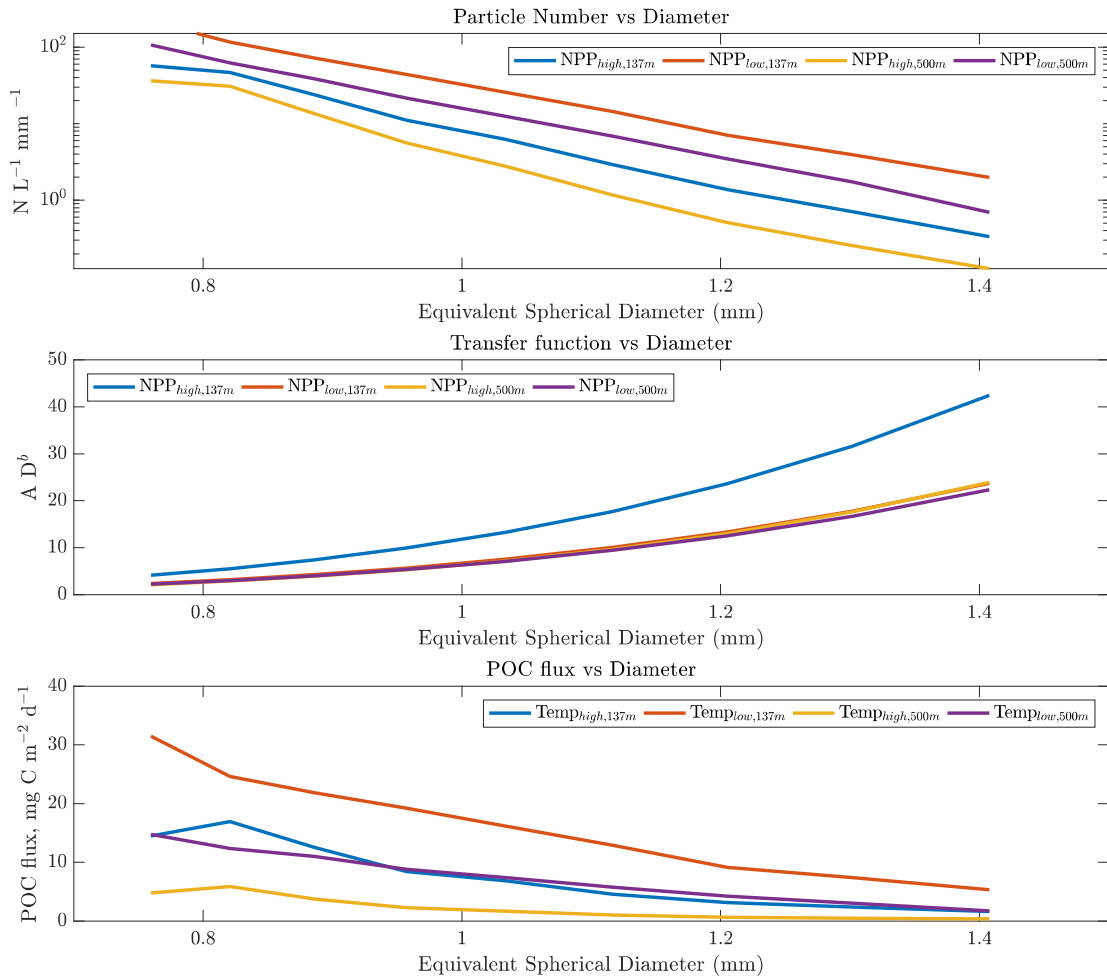
Figure 4. Particle size spectra (top panel), POC transfer functions, and POC fluxes are shown for different depths and equivalent spherical diameters.

Depth (m)	a	b	Model Performance			POC fluxes, mg C per m^{-2} d^{-1}			
			Cost	RMSD	R	UVP μ	UVP σ	DW17 μ	DW17 σ
137	11.5	3.8	0.5	35	0.5	68	56	60	38
190	10.9	3.8	0.5	30	0.4	56	53	48	32
250	9.9	3.9	0.5	22	0.3	42	30	38	25
400	7.2	4.1	0.5	15	0.3	28	20	26	19
500	6.2	3.9	0.5	14	0.2	24	19	22	16
1000	3.6	3.9	0.5	5	0.0	10	5	9	6
2500	2.9	3.9	0.3	3	0.8	6	3	3	1
depth =137; npp=high	12.5	3.8	0.4	35	0.7	70	69	65	56
depth =137; npp=low	11.0	3.7	0.3	28	0.3	59	41	49	21
depth = 500; npp=high	7.0	3.9	0.6	15	0.2	24	10	27	23
depth = 500; npp=low	5.0	3.9	0.4	9	0.3	18	11	16	11
depth =137; temp=high	11.7	3.8	0.4	32	0.6	65	57	56	37
depth = 137; temp = low	6.7	3.7	1.8	123	-0.2	148	180	87	46
depth =500; temp=high	6.3	3.9	0.5	11	0.3	21	12	20	15
depth =500; temp =low	6.3	3.7	0.9	32	-0.4	45	38	35	15
Honjo Traps	1.76	4.7	2.8	5	0.17	8	6	6	6

Table 1. Model performance metrics for the suite of optimizations are shown. Mean (μ) and standard deviation (σ) POC fluxes are reported in mg C m^{-2} d^{-1}.
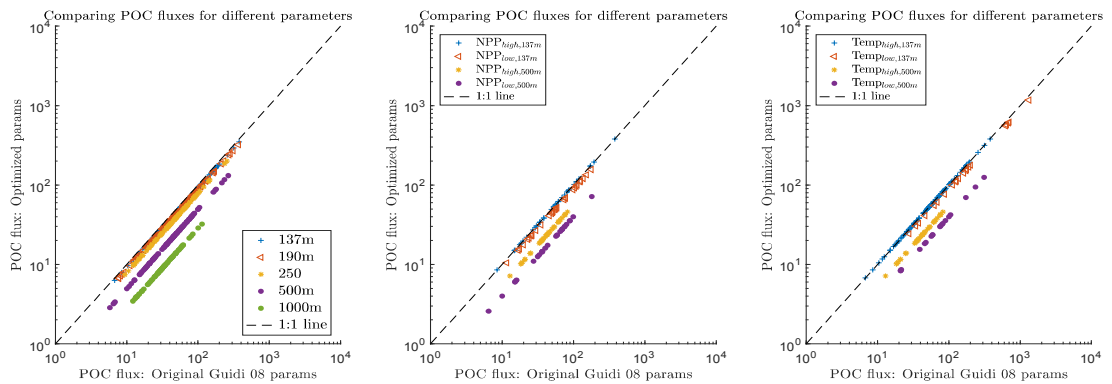

Figure 5. Comparisons of POC fluxes (in mg C m^{-2} d^{-1}) between the optimized model in this study and the original model from Guidi et al., (2008) are shown for different depths and surface ocean conditions.
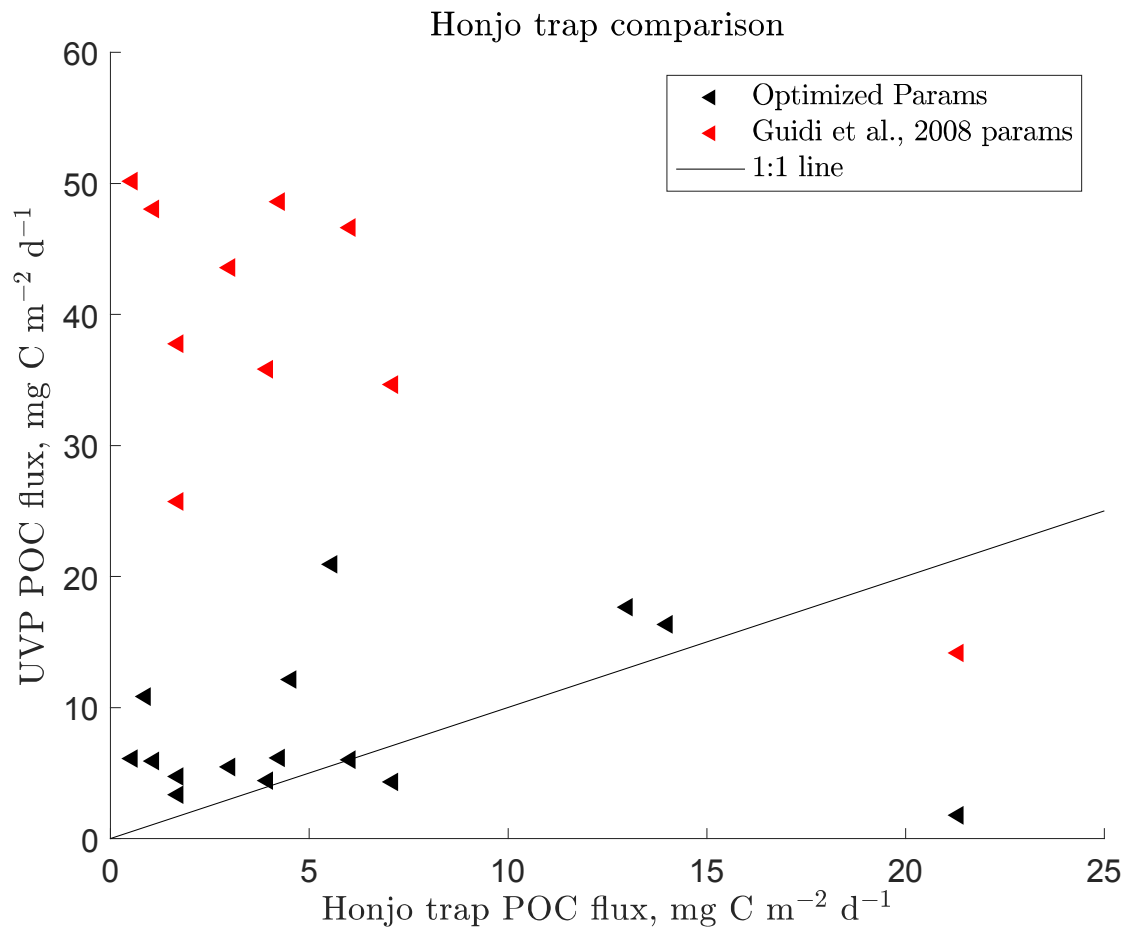
Figure 6. Comparisons of POC fluxes (in mg C m⁻² d⁻¹) between the UVP model and the

Honjo trap compilation shown using either optimized parameters from this study or the

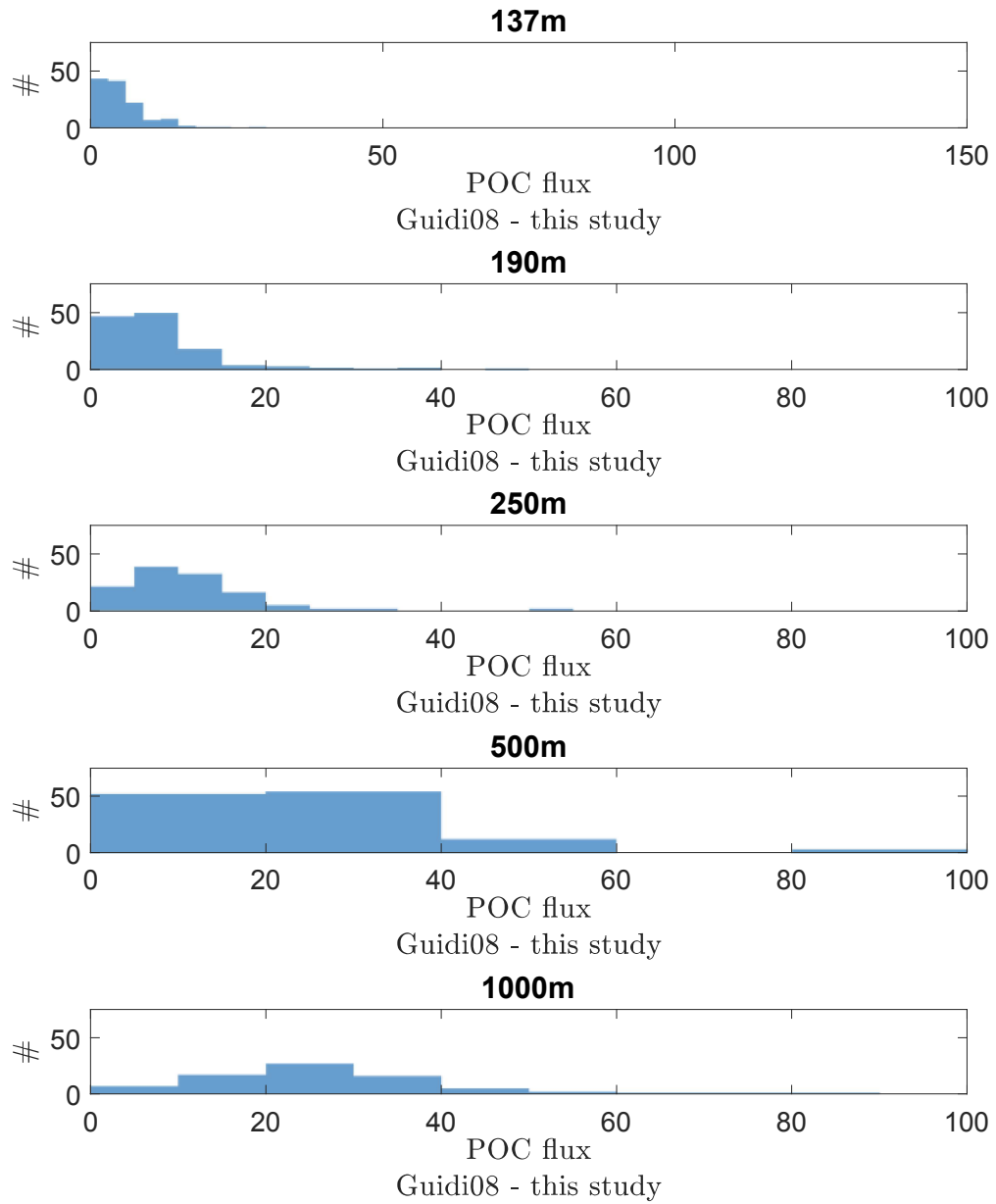original parameters of Guidi et al., 2008.

Figure 7. POC flux differences(in mg C m^{-2} d^{-1}) derived from the Guidi et al (2008) parameters – the parameters from this study are shown for different depth horizons.
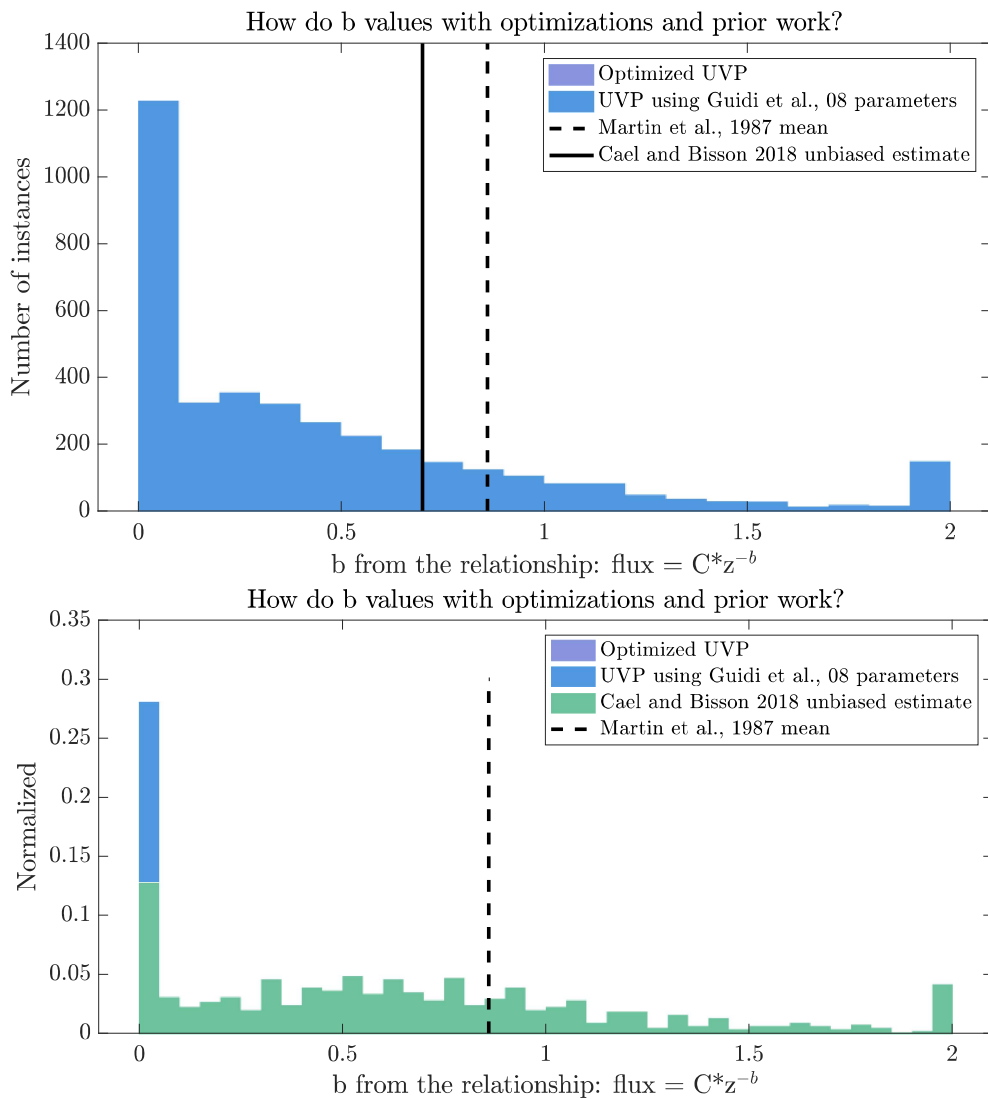
Figure 8. Top panel: The distribution of b values from the power law relationship that quantifies flux attenuation are shown for UVP-derived POC fluxes using optimized parameters from this study as compared to the Guidi et al., (2008) study. The solid line is the unbiased 'b' value from a observational profile compilation, and the dashed line is the estimate from the Martin et al., 1987 study. Bottom panel: The distribution of 'b' values

using optimized or Guidi et al., 2008 parameters is shown alongside the distribution of 'b'

values from the profile compilation of Cael and Bisson (2018) in green.

References

Alldredge, A. (1998). The carbon, nitrogen and mass content of marine snow as a function

of aggregate size. *Deep Sea Research Part I: Oceanographic Research Papers*, *45*(4-5),

529-541.

Alldredge, A. L., Passow, U., & Logan, B. E. (1993). The abundance and significance of

a class of large, transparent organic particles in the ocean. *Deep Sea Research Part I:*

Oceanographic Research Papers, *40*(6), 1131-1140.

Archer, D. E., Eshel, G., Winguth, A., Broecker, W., Pierrehumbert, R., Tobis, M., &

Jacob, R. (2000). Atmospheric pCO2 sensitivity to the biological pump in the ocean. *Global*

Biogeochemical Cycles, *14*(4), 1219-1230.

Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., ... & Not,

F. (2016). In situ imaging reveals the biomass of giant protists in the global

ocean. *Nature, 532*(7600), 504.

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.

Buesseler, K. O., Antia, A. N., Chen, M., Fowler, S. W., Gardner, W. D., Gustafsson, O., ... & Steinberg, D. K. (2007). An assessment of the use of sediment traps for estimating upper ocean particle fluxes. *Journal of Marine Research*, *65*(3), 345-416.

Burd, A. B., & Jackson, G. A. (2009). Particle aggregation. *Annual review of marine science*, *1*, 65-90.

Cael, B. B., Bisson, K., & Follows, M. J. (2017). How have recent temperature changes affected the efficiency of ocean biological carbon export?. *Limnology and Oceanography Letters*, *2*(4), 113-118.

Huntley, M.E., and M.D. Lopez, 1992. Temperature-dependent production of marine copepods: A global synthesis. *Am. Nat* 140: 201-242. doi: 10.1086/285410.

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.

Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., ... & Gorskya, 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.

125

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.

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., ... & Coelho, L. P. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465.

Jouandet, M.-P., T. W. Trull, L. Guidi, M. Picheral, F. Ebersbach, L. Stemmann, and S. Blain (2011), Optical imaging of mesopelagic particles indicates deep carbon flux beneath a natural iron-fertilized bloom in the Southern Ocean, Limnol. Oceanogr., 56(3), 1130–1140.

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.

Jackson, G.A., Maffione, R., Costello, D.K., Alldredge, A.L., Logan, B.E., Dam, H.G., 1997. Particle size spectra between 1mm and 1cm at Monterey Bay determined using multiple instruments. Deep-Sea Research I 44 (11), 1739–1767.

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

126

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.

Mari, X., Passow, U., Migon, C., Burd, A. B., & Legendre, L. (2017). Transparent exopolymer particles: Effects on carbon cycling in the ocean. *Progress in Oceanography*, *151*, 13-37.

Michaels, A. F., Silver, M. W., Gowing, M. M., & Knauer, G. A. (1990). Cryptic zooplankton "swimmers" in upper ocean sediment traps. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(8), 1285-1296.

Murray, A. G., & Jackson, G. A. (1992). Viral dynamics: a model of the effects of size, shape, motion and abundance of single-celled planktonic organisms and other particles. *Marine Ecology Progress Series*, 103-116.

Naselli-Flores, L., Padisák, J., & Albay, M. (2007). Shape and size in phytoplankton ecology: do they matter?. *Hydrobiologia*, *578*(1), 157-161.

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*(9), 462-473.

Picheral M, Colin S, Irisson J-O (2017). EcoTaxa, a tool for the taxonomic classification of images. http://ecotaxa.obs-vlfr.fr.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax. 2007. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim*. **20**: 5473–5496. doi:10.1175/2007JCLI1824.1

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., ... & Perry, M. J. (2016). Prediction of the export and fate of global ocean net primary production: the EXPORTS science plan. *Frontiers in Marine Science*, *3*, 22.

Smith, K. L., Ruhl, H. A., Huffard, C. L., Messié, M., & Kahru, M. (2018). Episodic organic carbon fluxes from surface ocean to abyssal depths during long-term monitoring in NE Pacific. *Proceedings of the National Academy of Sciences*, *115*(48), 12235-12240.

Stanca, E., Cellamare, M., & Basset, A. (2013). Geometric shape as a trait to study phytoplankton distributions in aquatic ecosystems. *Hydrobiologia*, *701*(1), 99-116.

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