Global Biogeochemical Cycles

RESEARCH ARTICLE
10.1029/2018GB005934

How Data Set Characteristics Influence Ocean Carbon Export Models

K. M. Bisson[1], D. A. Siegel[1], T. DeVries[1], B. B. Cael[2,3], and K. O. Buesseler[3]

[1]Earth Research Institute, University of California, Santa Barbara, CA, USA, [2]Massachusetts Institute of Technology, Cambridge, MA, USA, [3]Woods Hole Oceanographic Institution, Woods Hole, MA, USA

Key Points:
- Individual export flux observations do not represent climatological conditions and should not be used to optimize global export models
- The choice of net primary production (NPP) model used introduces sources of extrinsic variability that affect optimized parameter values by up to a factor of 2
- Model performance is greatly improved by choosing a suitably averaged data set to calibrate it

Supporting Information:
- Supporting Information S1
- Data Set S1
- Data Set S2
- Table S1

Correspondence to:
K. M. Bisson,
bisson@eri.ucsb.edu

Citation:
Bisson, K. M., Siegel, D. A., DeVries, T., Cael, B. B., & Buesseler, K. O. (2018). How data set characteristics influence ocean carbon export models. *Global Biogeochemical Cycles*, *32*, 1312–1328. https://doi.org/10.1029/2018GB005934

Received 26 MAR 2018
Accepted 8 AUG 2018
Accepted article online 24 AUG 2018
Published online 13 SEP 2018

Abstract Ocean biological processes mediate the transport of roughly 10 petagrams of carbon from the surface to the deep ocean each year and thus play an important role in the global carbon cycle. Even so, the globally integrated rate of carbon export out of the surface ocean remains highly uncertain. Quantifying the processes underlying this biological carbon export requires a synthesis between model predictions and available observations of particulate organic carbon (POC) flux; yet the scale dissimilarities between models and observations make this synthesis difficult. Here we compare carbon export predictions from a mechanistic model with observations of POC fluxes from several data sets compiled from the literature spanning different space, time, and depth scales as well as using different observational methodologies. We optimize model parameters to provide the best match between model-predicted and observed POC fluxes, explicitly accounting for sources of error associated with each data set. Model-predicted globally integrated values of POC flux at the base of the euphotic layer range from 3.8 to 5.5 Pg C/year, depending on the data set used to optimize the model. Modeled carbon export pathways also vary depending on the data set used to optimize the model, as well as the satellite net primary production data product used to drive the model. These findings highlight the importance of collecting field data that average over the substantial natural temporal and spatial variability in carbon export fluxes, and advancing satellite algorithms for ocean net primary production, in order to improve predictions of biological carbon export.

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 & Hoffert, 1985), commonly called the "biological pump." The exported carbon feeds mesopelagic communities and is sequestered from the atmosphere on time scales 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 (Figure S1). The same is true for multiple, near-simultaneous flux measurements over large regions (>100 km wide; Black et al., 2018; Buesseler et al., 2007; Estapa et al., 2015), 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/year (Boyd & Trull, 2007; Dunne et al., 2005; 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 & Silver, 1988; Armstrong et al., 2001; Burd & Jackson, 2009; Cram et al., 2018; Mahadevan et al., 2012; Steinberg et al., 2000). 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 (<200 m) 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 & 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 (Antia, 2005; Gardner et al., 1983). These factors can result in both undersampling and oversampling biases, depending on the specific trap design and the environment in which it is deployed (see Buesseler et al., 2007, and references therein). The interpretation of sediment trap data is further confounded by advection, because material collected in the trap may have been transported laterally several tens to hundreds 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 & 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 (Figure 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 (Figure 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., Dunne et al., 2005; Henson et al., 2011, 2012; Laws et al., 2000; Li & Cassar, 2016; Siegel et al., 2014). 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 time scale of particle settling. Although 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 Figure 3).

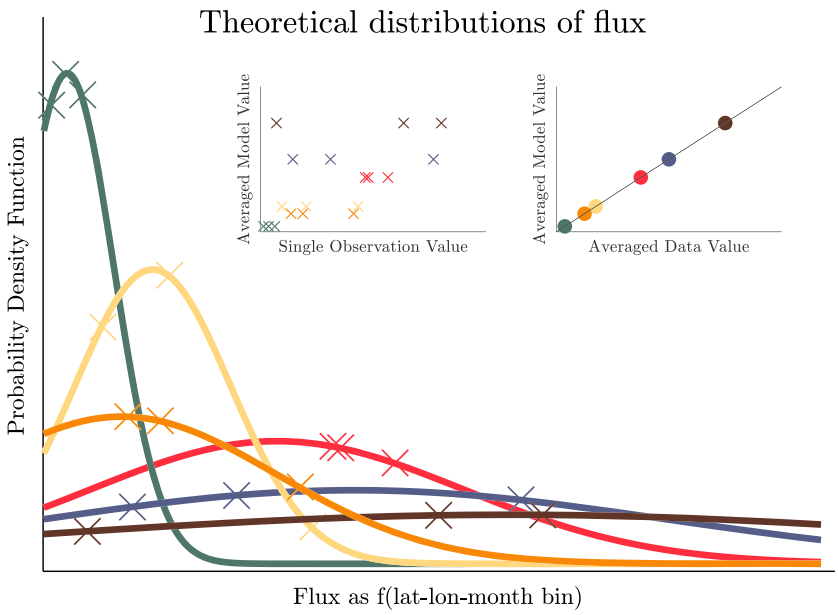

Figure 1. Theoretical distributions of biological carbon flux arising from complexities within latitude-longitude-month model bins are plotted. The inset shows how model–data comparisons look when data are either (left) randomly sampled or (right) sufficiently averaged. The distributions are randomly sampled 3 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.

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:

1. How do inconsistencies between the scale of POC flux observations and POC export models affect our ability to optimize these models using field observations?
2. 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 data sets 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 data set. 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} (Figure 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}):

$$TOT_{EZ} = ALG_{EZ} + FEC_{EZ}. \quad (1)$$

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

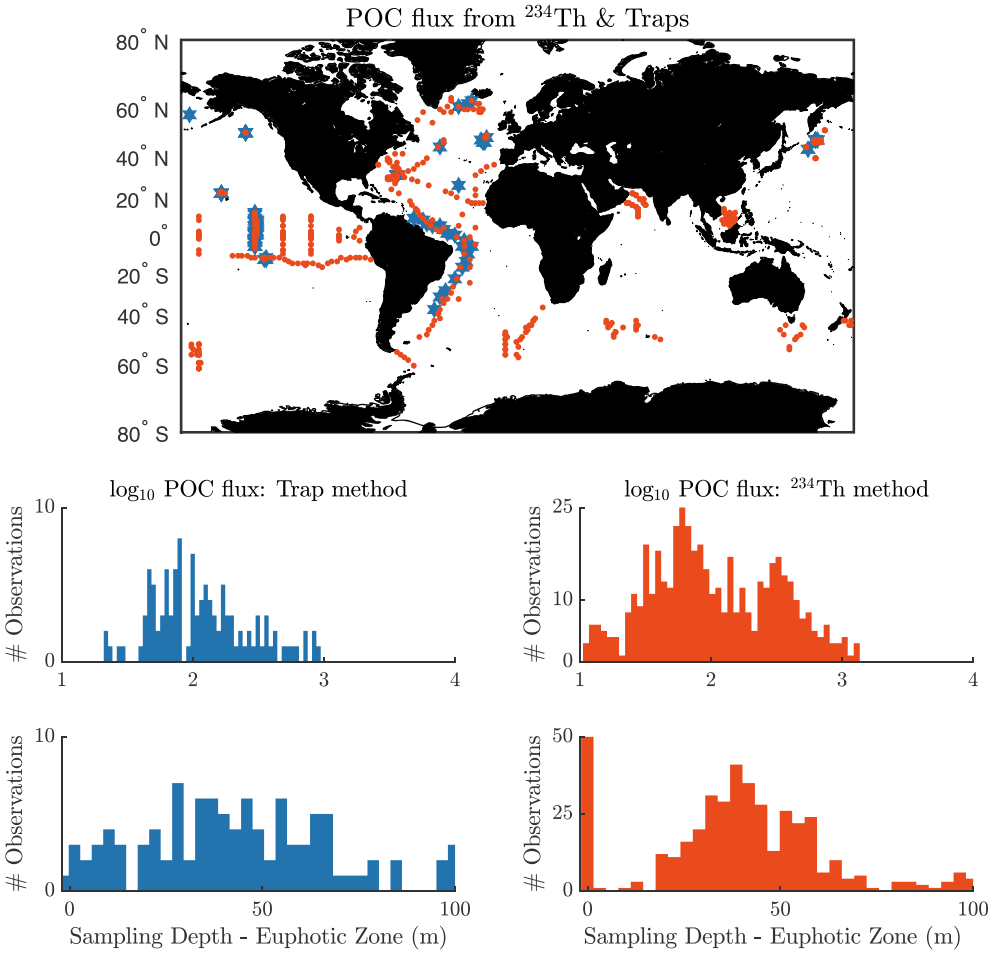

Figure 2. Map showing the sediment trap (blue) and POC flux from ^{234}Th (red) locations of the bins (n = 591). The bins are unique latitude/longitude/months in the complete data set arising from the data treatment described in section 2. The bottom panels show the histograms of these data products 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.

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 Boyd and Stevens (2002) and Michaels and Silver (1988)):

$$ALG_{EZ} = f_{Alg} \times NPP_M. \quad (2)$$

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

Fecal fluxes (FEC_{EZ}) 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:

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

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

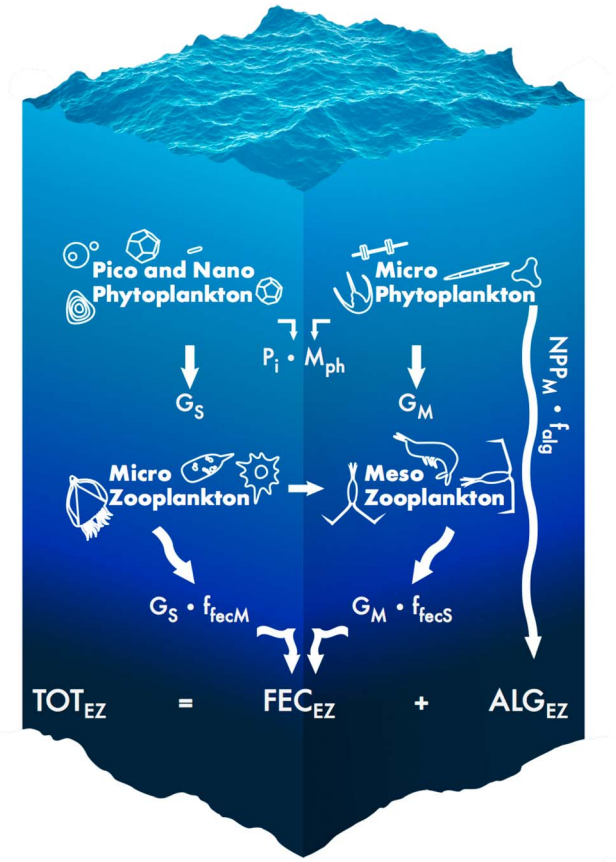

Figure 3. Diagram showing the conceptual elements of the model. NPP is routed into two phytoplankton size classes (picophytoplankton and nanophytoplankton (small), microphytoplankton (large)) which are ingested at a rate given by microzooplankton and mesozooplankton herbivory (Gs and gm, respectively). 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 microzooplankton and mesozooplankton. Fecal flux from both zooplankton size 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.

gains from net primary production (NPP_i) and losses to herbivory (G_i), nongrazing biological mortality (given a specific rate m_{ph} which is assumed by S14 to be 0.1 day), and reduction of biomass as the mixed layer (Z_{ml}) deepens (i.e., when $\frac{dZ_{ml}}{dt} > 0$).

$$\frac{dP_i}{dt} = \frac{NPP_i}{Z_{eu}} - G_i - m_{ph}P_i - \delta_{i,M}\frac{ALG_{EZ}}{Z_{eu}} - \frac{P_i}{Z_{ml}}\frac{dZ_{ml}}{dt}H\left(\frac{dZ_{ml}}{dt}\right). \quad (4)$$

$H(x) = 1$ 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_{fecM}, 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 section 4.

2.2. Model Input Data Sets

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 443 nm. 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 1° spatial grid.

We test three different NPP inputs (Carbon-based Productivity Model, CbPMv2 (Westberry et al., 2008); Vertically Generalized Productivity Model, VGPM (Behrenfeld & 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 Figure S2 in the supporting information. Although all three NPP models have similar globally integrated mean values (CbPM 52 Pg C/year, VGPM 51 Pg C/year, CAFE 52 Pg C/year), 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 3 (Figure S2). Implications of the choice of NPP product are discussed in section 4.1.

Table 1
Table Showing Sources of Error and Uncertainty for Each Data Set Used[a]

Sources of error	Natural variability		Instrumental/ sampling biases	Normalization to Z_{eu}
	Temporal	Spatial		
All data	✓	✓	✓	✓
Climatological data		✓	✓	
^{234}Th Profiles	✓	✓	✓	
HOT and BATS		✓	✓	✓
Uncertainty	40%*	60%*	≥30%	0–60%*

[a]The starred columns (temporal, spatial, and Z_{eu} normalization) are used in the cost function to account for data set-specific sources of error. Checked columns indicate that error is present within that data set, 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 time scale of interest is one month. A description of how the error values are calculated is in section 2.

2.3. Synthesis of Export Flux Observations

The model is fit to five different POC flux data sets 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 data set of export flux observations from sediment traps and the ^{234}Th method ("All Data"). This data set is then subsampled to create the four other data sets: an extension of the data presented in Buesseler and Boyd (2009) of well-sampled (multistation averages) ^{234}Th-estimated fluxes at Z_{eu} ("Climatological Data"), a data set consisting of ^{234}Th profiles at from transect cruises ("^{234}Th Profiles"), and data sets from sediment traps at the Hawaii Ocean Time Series (HOT) and Bermuda Atlantic Time Series (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 Data Set, All Data

Existing databases, papers, and data from personal communication are assembled into a global data set of shallow (<200 m) POC flux from sediment traps and from the ^{234}Th technique, totaling 1719 observations (available in the supporting information).

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 (>500 m) as to not be affected by terrestrial input;
3. bin all data into 1° 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) and 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 1,721 to 1,343 (Figure 2). Step 3 further organizes the 1,343 data points into 591 unique bins (Figure 2). This data set has the largest global coverage but also the most uncertainty associated with it (see Table 1).

The All Data data set 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 data set 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 data set also differs from that used in Henson et al. (2011) because the Henson et al. (2011) data set exclusively comprises ^{234}Th-estimated POC fluxes normalized to a nominal depth (100 m) 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 supporting information for compilation). We require at least four observations per latitude/longitude/month bin at the euphotic layer depth. This data set 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 data set expands on the data set presented in Buesseler and Boyd (2009, Table 1) (BB09) to include 165 total observations from 14 sites (Figure S3). We eliminated the two Ocean Station Papa observations from the original BB09 data set because each of those were single observations, as well as the Southern Ocean observations because there are no 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. The ^{234}Th Profiles

We also compile a data set consisting of ^{234}Th profiles with sufficient vertical sampling resolution in the upper 200 m in order to recalculate fluxes at specific depths. This ^{234}Th Profiles subset uses ^{234}Th-derived POC flux profiles that were collected along transects as part of the U.S. and Dutch GEOTRACES programs (Black et al., 2018; Owens et al., 2015) in addition to a regional study (Estapa et al., 2015), which estimated export flux values at Z_{eu} (Figure S3 in the supporting information). 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 data set (Table 1). There are 91 unique latitude/longitude/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 supporting information).

2.3.4. HOT and 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 third and fourth subdata sets from All Data for the climatological POC export at HOT and BATS. The HOT data set is from January 1998 to October 2010, totaling 115 measurements (Figure S1), and the BATS data set is from October 1997 to November 2010, totaling 135 measurements (Figure S1). All traps are cylindrical particle interceptor traps deployed at 150 m for 3 days (Karl & Lukas, 1996; Steinberg et al., 2001). These data are normalized to the depth of the euphotic layer, as is done in the global data set, and monthly climatologies of POC export are calculated spanning the time periods above. HOT and BATS are climatologically averaged fluxes of localities at 150 m, 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 data set 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 ($\overline{x_i}$) at HOT and BATS, where i is the number of observations and m is the month, or u ncertainty$_{time}$ = mean$\left(\frac{s_m}{\overline{x}_{i,m}}\right)$. The conservative estimate is reported (BATS had 43% uncertainty, HOTS had 30% uncertainty) (Figure 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) flux data worldwide.
2. The spatial error estimate of 63% was calculated from the Estapa et al. (2015) data set, 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 data set as well as in the time series data sets because the data were not collected at the base of Z_{eu}. For each data point, 1,000 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 1,000 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$_z$ = σ/μ). 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 Figure 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 section 1, reported flux values vary depending on the sampling and analysis methods used, even within a single sampling type (e.g., tethered traps versus 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 data set using a Bayesian approach that maximizes a log likelihood function (Neal, 2003),

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

In equation (5), i is the index of each latitude/longitude/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 large parameter deviations from what is expected. The log prior uses the four (indexed by j) baseline parameters (f_{alg} = 0.1, f_{fecM} = 0.3, f_{fecS} = 0.1, m_{ph} = 0.1) from S14 as the prior information.

$$\log \text{ prior} = \sum_1^4 \left(\log\left(parameter_j\right) - \log\left(expected_j\right)\right)^4. \tag{7}$$

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

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

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

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

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 latitude/longitude/month bin, the effective error from natural variability was reduced. From this, the standard deviation (sigma) was calculated as

$$\sigma_t^2 = \log\left(1 + {v_i}/{\log(data)^2}\right) \tag{10}$$

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. First, four initial parameter values (to represent f_{alg}, f_{fecM}, f_{fecS}, 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) and (6)) was calculated. This process was repeated iteratively to generate a "chain" of 2,000 samples of each parameter from the posterior probability density function. 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 & 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 five different data sets for the optimization, and using three 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 data set.

3. Results

The optimized parameters and their associated uncertainties using the different data sets are presented in Figures 4a–4d, alongside the correlation coefficients and log likelihoods for each optimization run (Figures 4e and 4f), model mean export fluxes (Figure 4g), and the average EZ ratio (Figure 4h; export flux at base of the euphotic layer divided by NPP) for the corresponding locations and times in each data set. The mean values discussed below are for outputs generated using the latitude/longitude/month bins specific to each data set. 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 underpredict 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 (Figure 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 nongrazing 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 day, 0.13 ± 0.08, and 0.09 ± 0.02, respectively (Figures 4a–4d). This contrasts with the baseline case with parameter values f_{Alg} = 0.1, m_{ph}=0.1 day, f_{fecM}=0.3, and f_{fecS}=0.1. Notably, the optimized algal fraction is double the baseline case, and the nongrazing mortality parameter is half the baseline case. Further, the relative contribution of small and large phytoplanktons 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 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 data set.

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 ± 0.17, m_{ph} = 0.07± 0.02 day, f_{fecM} = 0.21 ± 0.13, and f_{fecS}=0.08 ± 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

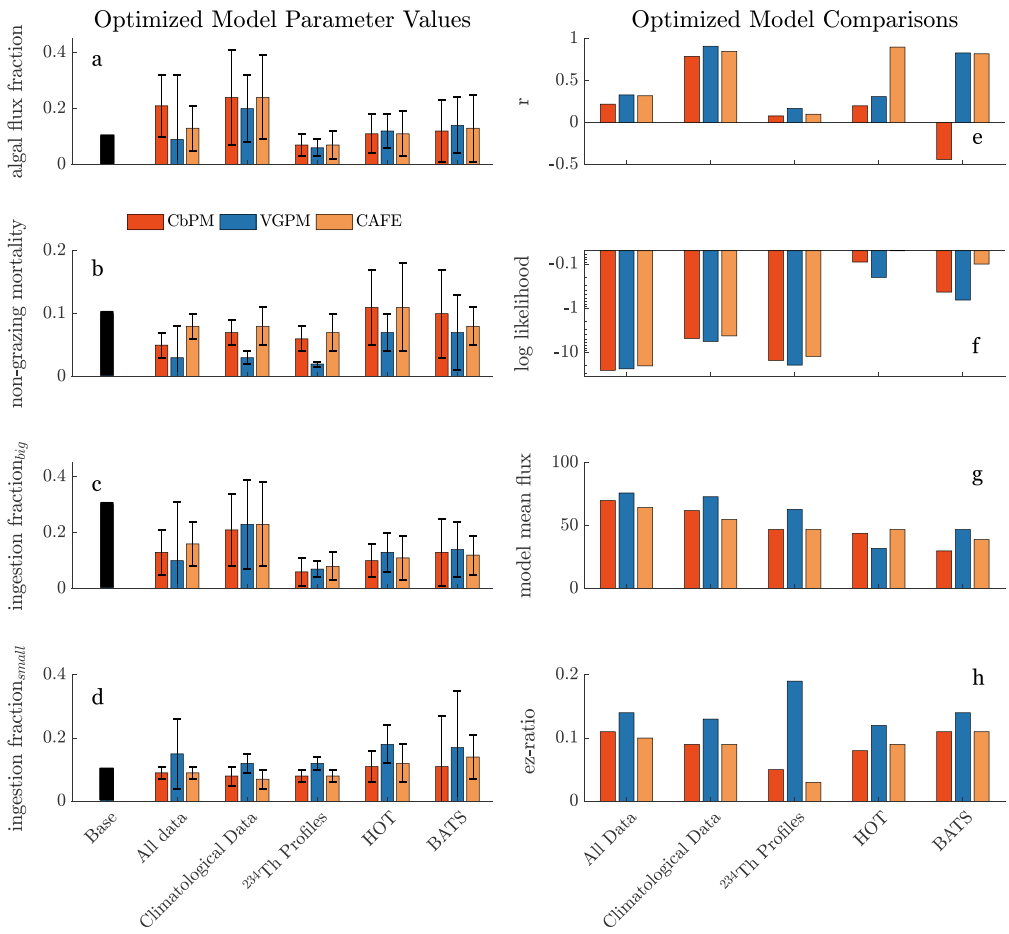

Figure 4. (left panel) Bar graphs for optimized parameter values for the various NPP input products and the model development data sets are shown. NPP product comparisons are shown in color (red is CbPM, blue is VGPM, and tan is CAFE). The a priori values for each parameter are shown in cream on the far left. The *x* axis shows the data set used. Error bars indicate the magnitude of the standard deviation for each product. Note the *y* scale differences among parameter type. (right panel) The first panel shows the Pearson correlation coefficient, *r*, which compares the specified data set against the optimized model. The second panel shows the 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 third panel shows the values for the mean model flux (mg C m^{-2} d^{-1}), and the fourth panel shows the values for the EZ ratio, the fraction of production that is exported.

is smaller than previously prescribed. The Climatological Data algal flux fraction is again roughly double the S14 value, and the value for nongrazing 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 Figures 5a and 5b, which is analogous to the conceptual diagram presented earlier (inset panel in Figure 1). A data set with high noise results in a poor correspondence with the model, but a data set 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 data set are roughly equal (Figure 4), with values f_{Alg} = 0.07 ± 0.04, m_{ph} = 0.06 ± 0.02 day, f_{fecM}= 0.07 ± 0.05, and f_{fecS}=0.08 ± 0.02. This result implies an equal contribution of small and large phytoplanktons to the fecal flux pathway, and also a diminished role of aggregation compared to the baseline case and to the other data sets. The optimized model represents improvement over the baseline case, which predicts a mean of 47 compared to 69 mg C m^{-2} d^{-1} when confronted

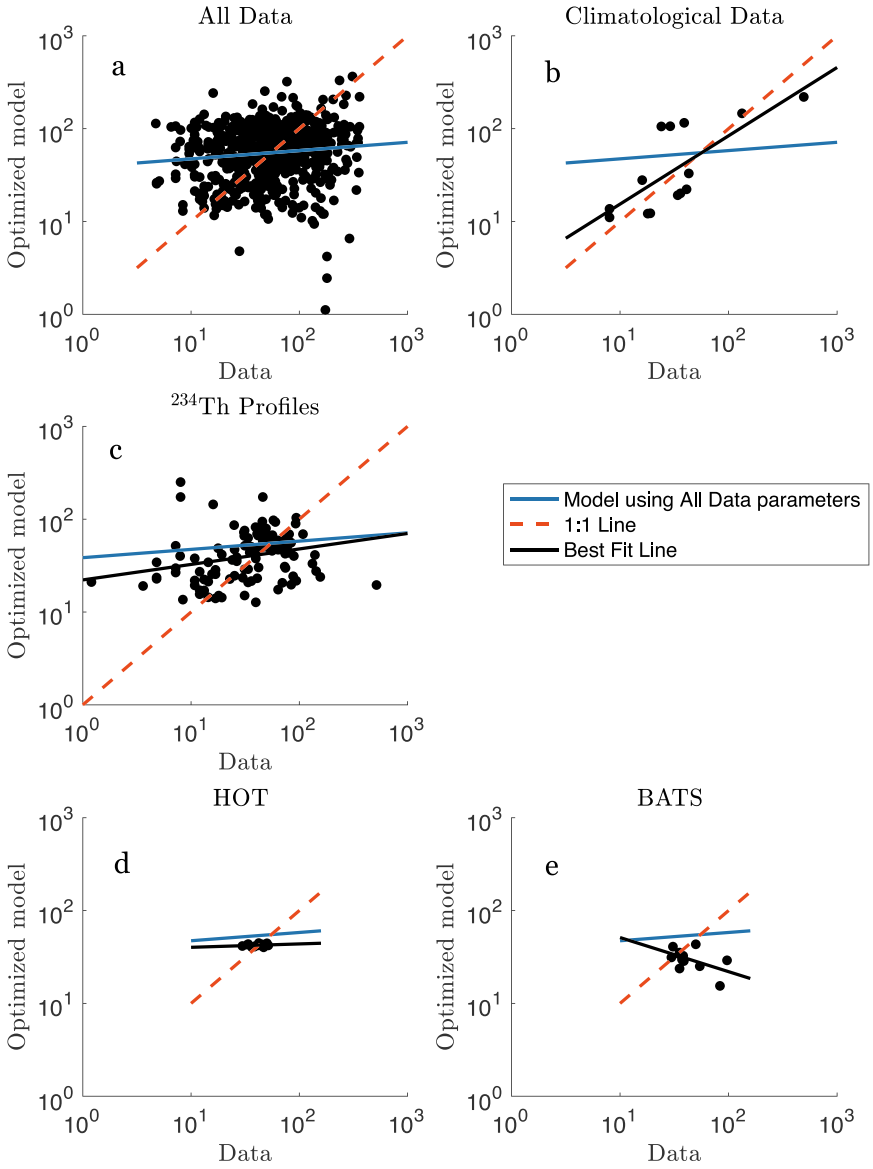

Figure 5. (a–e) Model versus data scattergraphs are compared among the different data set types using the CbPM model as the NPP input. The navy line is the best fit line for "All Data" case, which is shown in the other graphs to illustrate how the model would look if the All Data parameters were used in those different scenarios. The red line is the 1:1 line, and the black line is the best fit line for (b)–(d).

with a data mean of 53 mg C m^{-2} d^{-1}. Despite this result, the correlation is very weak (r = 0.08; Figure 5c). For low-flux environments, the ^{234}Th Profiles optimized model overpredicts 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 Figures 4, 5d, and 5e. 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 ± 0.07, m_{ph} = 0.11 ± 0.06 day, f_{fecM} = 0.10 ± 0.06, and f_{fecS} = 0.11 ± 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 ± 0.11, m_{ph} = 0.10 ± 0.07 day, f_{fecM} = 0.13 ± 0.12, and f_{fecS}=0.11 ± 0.16, which has substantially more uncertainty with each parameter compared to those estimated from the HOT data set.

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), although 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 (Figure 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. 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 data sets 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 data sets 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 (Doney et al., 2009; Kahru, 2017; Lee et al., 2015), especially in high-nutrient low-chlorophyll conditions, and at extreme temperatures or chlorophyll concentrations (Carr et al., 2006, and references 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 (Figures 4a–4d). For example, with All Data, optimized parameters can vary by 100%, especially when comparing the VGPM generated parameters to those from CbPM and CAFE (Figure 4). At HOT, the fraction of ingested small phytoplankton that drives export varies from 9 to 18% (Figure 4d) depending on NPP product used. Lower NPP values from VGPM (Figure 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 affects not only the optimized parameter values but also model performance. Figure 4 shows how model performance is linked to NPP product. The most 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 (Figure S1). The various NPP products also differ in their predictions of the variability of NPP over a seasonal cycle at HOT and BATS (Figure 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 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., DeVries & Weber, 2017; Siegel et al., 2014).

4.2. Implications for the Magnitude and Mechanisms of POC Export

The optimized parameters found using the three global data sets 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 data sets used, Climatological Data has the minimum amount of

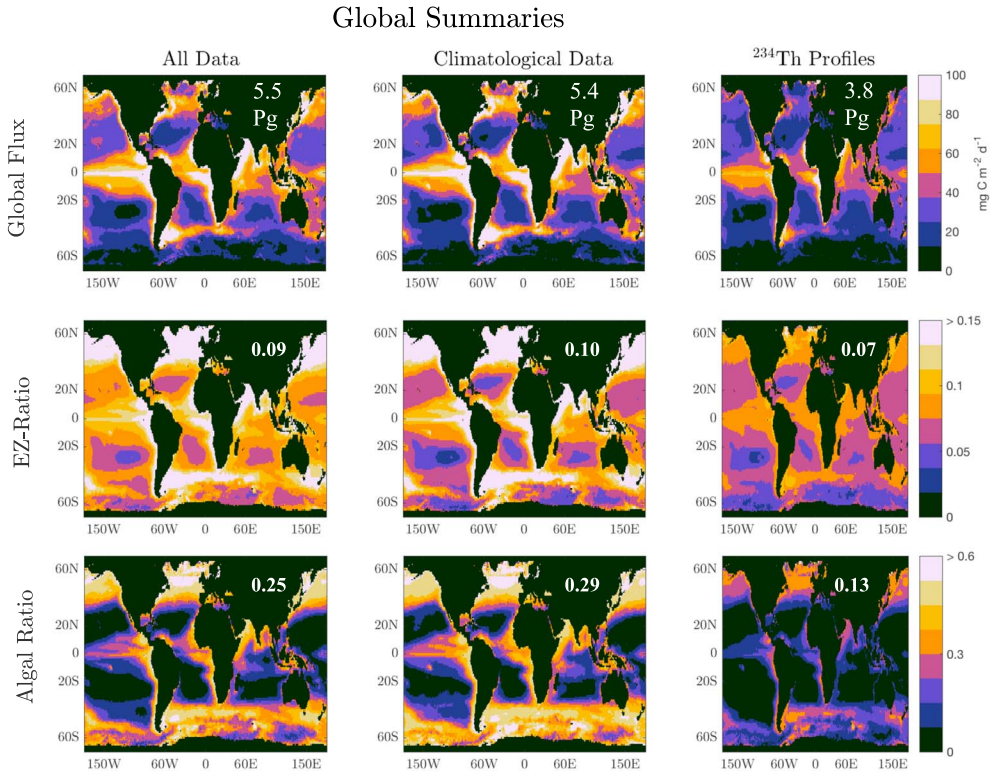

Figure 6. Global summary maps and statistics are shown for the three data sets that span several ocean basins ("All Data," "Climatological Data," "^{234}Th Profiles"). The annually integrated POC flux values are shown in white on the top panel, in petagrams. The EZ ratio (the proportion of production that is exported beneath the euphotic layer) is shown in the second row, with global averages in white. The algal ratio (the fraction of flux from aggregates) is shown in the third row with global averages in white. All of the reported statistics and maps are done with CbPM as an input production product.

uncertainty associated with the observations, across depth, time, space, and method (Table 1). It is also the data set 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 (Figure 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 data set used to optimize the model. Using the CbPM, the annually integrated global POC export is 5.5, 5.4, and 3.8 Pg C/year when optimizing the model with All Data, Climatological Data, and ^{234}Th Profiles, respectively (Figure 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 Pg C/year; Siegel et al., 2014). However, the model that is optimized to match the ^{234}Th Profiles data set predicts a substantially smaller globally integrated flux value of 3.8 Pg C/year. 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 data set and the others is that much of the data come from oligotrophic regions (e.g., South Pacific and the Sargasso Sea), and are therefore skewed toward

low-export environments (the mean flux of the ^{234}Th Profiles data set is 48 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 data set.

The pathways of POC export are also sensitive to the data set 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 data set 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 data sets is the smaller value of nongrazed mortality parameter, which is similar for all data sets (0.05, 0.07, 0.06 day for All Data, Climatological Data, and ^{234}Th Profiles, respectively), compared with the baseline case of 0.1 day. These values approach the range of values typically used in global ecosystem models (e.g., Aumont & Bopp, 2006, where phytoplankton mortality = 0.01 day; Aumont et al., 2003, where mortality ranges from 0.008 up to a maximum of 0.2 day depending on the type of phytoplankton; Le Quéré et al., 2005, where maximum mortality is 0.05 day; Fasham et al., 1990, where mortality varies from 0.045 to 0.06 day). These optimized nongrazed 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 nongrazing 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 (Boyd & Stevens, 2002; Michaels & Silver, 1988).

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 100 m, and the depth of winter mixing (e.g., Antia et al., 2001; Buesseler & Boyd, 2009; Henson et al., 2011; Li & Cassar, 2017). Here we use the ^{234}Th Profiles data set to recalculate model fluxes at several different reference depth: the primary production zone (the depth where fluorescence reaches 10% of its maximum; Owens et al., 2015), the monthly mean mixed layer depth, and 100 m (see supporting information for methodology). The choice of these in any given study will have different rationales (mechanistic or nominal) depending upon their usage (Palevsky & Doney, 2018).

Depending on the depth normalization used, the optimized global flux varies by ~40%, from 2.8 to 4.0 Pg C/year (Table S2). These export fluxes are small compared to other global estimates because the data set 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 primary production zone, 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/year (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 nongrazing 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 do not represent climatological conditions, as most export data do not.

The data sets 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 data set compiled here reveals that spatial variability within a 1° 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 three to seven 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 & 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 data sets 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 are not subject to the same intrinsic and extrinsic sources of variability that affect carbon flux estimations, such as ocean tracer data (e.g., DeVries & 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
The authors gratefully acknowledge support from the NASA Ocean Biology and Biogeochemistry Program. D.A.S., K. M.B., and K.O.B. are supported by NNX16AR49G. K.O.B. is also supported by OBB16_2-0031. B.B.C. is supported by NNX16AR47G and T.D. is supported by NNXA122G. K.M.B. and B.B.C. also acknowledge support from the National Science Foundation Graduate Research Fellowship program. All satellite data are freely available via https://oceancolor.gsfc.nasa.gov/cgi/l3 and https://www.science.oregonstate.edu/ocean.productivity/, and the flux data products are available in the supporting information. We thank Erin Black, Stephanie Owens, and Meg Estapa for their assistance in evaluating thorium profiles, and we thank Toby Westberry for generating the CAFE NPP product during the SeaWiFS era. We especially acknowledge two anonymous reviewers for their thoughtful contributions to this manuscript.

References

Abraham, E. R. (1998). The generation of plankton patchiness by turbulent stirring. *Nature*, *391*(6667), 577–580. https://doi.org/10.1038/35361

Alldredge, A. L., & Silver, M. W. (1988). Characteristics, dynamics and significance of marine snow. *Progress in Oceanography*, *20*(1), 41–82. https://doi.org/10.1016/0079-6611(88)90053-5

Antia, A. N. (2005). Solubilization of particles in sediment traps: Revising the stoichiometry of mixed layer export. *Biogeosciences*, *2*(2), 189–204. https://doi.org/10.5194/bg-2-189-2005

Antia, A. N., Koeve, W., Fischer, G., Blanz, T., Schulz-Bull, D., Schölten, J., et al. (2001). Basin-wide particulate carbon flux in the Atlantic Ocean: Regional export patterns and potential for atmospheric CO_2 sequestration. *Global Biogeochemical Cycles*, *15*(4), 845–862. https://doi.org/10.1029/2000GB001376

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., & Wakeham, S. G. (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*(1-3), 219–236. https://doi.org/10.1016/S0967-0645(01)00101-1

Aumont, O., & Bopp, L. (2006). Globalizing results from ocean in situ iron fertilization studies. *Global Biogeochemical Cycles*, *20*, GB2017. https://doi.org/10.1029/2005GB002591

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), 1060. https://doi.org/10.1029/2001GB001745

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*(1), 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Black, E. E., Buesseler, K. O., Pike, S. M., & Lam, P. J. (2018). ^{234}Th as a tracer of particulate export and remineralization in the southeastern tropical Pacific. *Marine Chemistry*, *201*, 35–50. https://doi.org/10.1016/j.marchem.2017.06.009

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. https://doi.org/10.1016/S0079-6611(02)00020-4

Boyd, P. W., & Trull, T. W. (2007). Understanding the export of marine biogenic particles: Is there consensus? *Progress in Oceanography*, *72*, 276–312.

Brooks, S. P., & 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. https://doi.org/10.1029/97GB03366

Buesseler, K. O., Antia, A. N., Chen, M., Fowler, S. W., Gardner, W. D., Gafsson, O., et al. (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., & Boyd, P. W. (2009). Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnology and Oceanography*, *54*(4), 1210–1232.

Buesseler, K. O., Trull, T. W., Steinberg, D. K., Silver, M. W., Siegel, D. A., Saitoh, S. I., et al. (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. https://doi.org/10.1016/j.dsr2.2008.04.024

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

Carr, M. E., Friedrichs, M. A., Schmeltz, M., Aita, M. N., Antoine, D., Arrigo, K. R., et al. (2006). A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5-7), 741–770. https://doi.org/10.1016/j.dsr2.2006.01.028

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

Dall'Olmo, G., & Mork, K. A. (2014). Carbon export by small particles in the Norwegian Sea. *Geophysical Research Letters*, *41*, 2921–2927. https://doi.org/10.1002/2014GL059244

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

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

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., et al. (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. https://doi.org/10.1016/j.jmarsys.2008.05.015

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*, GB4027. https://doi.org/10.1029/2004GB002390

Estapa, M., Durkin, C., Buesseler, K., Johnson, R., & Feen, M. (2017). Carbon flux from bio-optical profiling floats: Calibrating transmissometers for use as optical sediment traps. *Deep Sea Research Part I: Oceanographic Research Papers*, *120*, 100–111.

Estapa, M. L., Siegel, D. A., Buesseler, K. O., Stanley, R. H., Lomas, M. W., & Nelson, N. B. (2015). Decoupling of net community and export production on submesoscales in the Sargasso Sea. *Global Biogeochemical Cycles*, *29*, 1266–1282. https://doi.org/10.1002/2014GB004913

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. https://doi.org/10.1357/002224090784984678

Gardner, W. D. (1985). The effect of tilt on sediment trap efficiency. *Deep Sea Research Part A. Oceanographic Research Papers*, *32*(3), 349–361. https://doi.org/10.1016/0198-0149(85)90083-4

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. https://doi.org/10.1357/002224083788520180

Giering, S. L. C., 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*, 159–171. https://doi.org/10.1002/2016GB005424

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., van Dongen-Vogels, V., et al. (2015). Analytical phytoplankton carbon measurements spanning diverse ecosystems. *Deep Sea Research Part I: Oceanographic Research Papers*, *102*, 16–25. https://doi.org/10.1016/j.dsr.2015.04.006

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. https://doi.org/10.1016/0967-0637(94)90079-5

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*, GB1028. https://doi.org/10.1029/2011GB004099

Henson, S. A., Sanders, R., Madsen, E., Morris, P. J., Le Moigne, F., & Quartly, G. D. (2011). A reduced estimate of the strength of the ocean's biological carbon pump. *Geophysical Research Letters*, *38*, L04606. https://doi.org/10.1029/2011GL046735

Henson, S. A., Yool, A., & Sanders, R. (2015). Variability in efficiency of particulate organic carbon export: A model study. *Global Biogeochemical Cycles*, *29*, 33–45. https://doi.org/10.1002/2014GB004965

Kahru, M. (2017). Ocean productivity from space: Commentary. *Global Biogeochemical Cycles*, *31*, 214–216. https://doi.org/10.1002/2016GB005582

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. https://doi.org/10.1038/426032a

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. https://doi.org/10.1016/0967-0645(96)00005-7

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. https://doi.org/10.1016/0198-0149(79)90089-X

Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. *Journal of Geophysical Research*, *114*, C09015. https://doi.org/10.1029/2009JC005303

Laws, E. A., Falkowski, P. G., Smith, W. O., Ducklow, H., & McCarthy, J. J. (2000). Temperature effects on export production in the open ocean. *Global Biogeochemical Cycles*, *14*, 1231–1246.

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. https://doi.org/10.1016/j.jmarsys.2014.11.015

Li, Z., & Cassar, N. (2016). Satellite estimates of net community production based on O_2/Ar observations and comparison to other estimates. *Global Biogeochemical Cycles*, *30*, 735–752. https://doi.org/10.1002/2015GB005314

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), 5015–5027. https://doi.org/10.5194/bg-14-5015-2017

Mahadevan, A., D'Asaro, E., Perry, M. J., & Lee, C. (2012). Eddy-driven stratification initiates North Atlantic spring phytoplankton blooms. *Science*, *337*(6090), 54–58. https://doi.org/10.1126/science.1218740

Marsay, C. M., Sanders, R. J., Henson, S. A., Pabortsava, K., Achterberg, E. P., & Lampitt, R. S. (2015). Attenuation of sinking particulate organic carbon flux through the mesopelagic ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(4), 1089–1094.

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. https://doi.org/10.1016/0198-0149(87)90086-0

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. https://doi.org/10.1016/0198-0149(88)90126-4

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. https://doi.org/10.1016/0198-0149(90)90043-U

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

Mouw, C. B., Barnett, A., McKinley, G. A., Gloege, L., & Pilcher, D. (2016). Global Ocean particulate organic carbon flux merged with satellite parameters. *Earth System Science Data*, *8*(2), 531–541. https://doi.org/10.5194/essd-8-531-2016

Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, *31*(3), 705–767. https://doi.org/10.1214/aos/1056562461

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. https://doi.org/10.1016/j.dsr2.2014.11.010

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

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. https://doi.org/10.1016/j.dsr.2006.06.003

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

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

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*, 181–196. https://doi.org/10.1002/2013GB004743

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. https://doi.org/10.1016/j.dsr.2007.10.006

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*, 1756–1777. https://doi.org/10.1002/2016GB005521

Steinberg, D. K., Carlson, C. A., Bates, N. R., Goldthwait, S. A., Madin, L. P., & Michaels, A. F. (2000). Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Research, Part I*, *47*(1), 137–158. https://doi.org/10.1016/S0967-0637(99)00052-7

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. https://doi.org/10.1016/S0967-0645(00)00148-X

Valdes, J. R., & Price, J. F. (2000). A neutrally buoyant, upper ocean sediment trap. *Journal of Atmospheric and Oceanic Technology*, *17*(1), 62–68. https://doi.org/10.1175/1520-0426(2000)017<0062:ANBUOS>2.0.CO;2

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In E. T. Sundquist & W. S. Broecker (Eds.), *The carbon cycle and atmospheric CO_2: Natural variations archean to present, Geophysical Monograph Series* (Vol. 32, pp. 99–110). Washington, DC: American Geophysical Union. https://doi.org/10.1029/GM032p0099

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*, GB2024. https://doi.org/10.1029/2007GB003078

Zehr, J. P., Weitz, J. S., & Joint, I. (2017). How microbes survive in the open ocean. *Science*, *357*(6352), 646–647. https://doi.org/10.1126/science.aan5764