Supplement for "Particle flux parameterizations: quantitative and mechanistic similarities and differences"

26 September 2018

B. B. Cael & Kelsey Bisson

i. Derivations

The full three-dimensional time-varying conservation equation for POC [p, g m^{-3}] can be written:

$$\partial_t p + \vec{u} \cdot \nabla p - \kappa \nabla^2 p = \mathcal{S} \tag{1}$$

where $\vec{u} = (u_x, u_y, u_z)$ is the incompressible velocity field, κ is the molecular diffusivity, and $\mathcal{S}$ represents the net effect of sources and sinks. We are interested in large horizontal and temporal scales, so we consider the system as integrated over large enough spatial and temporal scales for time-variation and lateral advection and diffusion to be negligible, or equivalently we assume that the system is in steady state and one-dimensional. The equation then becomes

$$u_z \partial_z p - \kappa \partial_{zz} p = \mathcal{S} \tag{2}$$

We then assume that the two terms on the left hand side are small, because both molecular diffusivity and ocean vertical velocities on large horizontal and temporal scales are small relative to POC fluxes. We are left with the expression that sources must balance sinks (at least to leading order), i.e. $\mathcal{S} \approx 0$; note we have now assumed away all of the effects of tracer advection and diffusion. We then assume that the divergence[1] of POC flux dominates the source component of $\mathcal{S}$ and is balanced by the sink component of $\mathcal{S}$, which is commonly referred to as remineralization but includes all processes that remove material from the POC pool (e.g. fragmentation).

Splitting $\mathcal{S}$ into these parts, then, we have the basic equation from which these models arise, where w is apparent settling velocity, and k is the apparent first-order remineralization rate:

$$-\partial_z(wp) - kp = 0 \tag{3}$$

If we then assume $k \approx$ const. and $w \approx$ const., the solution to this first-order linear differential equation for p is

$$p(z) = Ae^{-\frac{k}{w}z} \tag{4}$$

where A is a constant dependent on initial conditions, i.e. on p at a shallow reference depth. Multiplying both sides by w to get POC flux f yields

$$f(z) = wp(z) = wAe^{-\frac{k}{w}z} \tag{5}$$

so if we define $\ell := w/k$ and $C := wA$, we find

$$f_e(z) = Ce^{-z/\ell} \tag{6}$$

i.e. the exponential model. The ballast[2] and double exponential models can be trivially derived as extensions of this. If we consider two distinct pools of POC, i.e. $p = p_1 + p_2$, each with their own associated (w_1, w_2, k_1, k_2), then each has an exponential solution for f_1 and f_2 as above, and the total flux is $f = f_1 + f_2$ which is the sum of these exponentials, i.e. the double exponential model. If $k_2 = 0$, e.g. because p_2 is protected from remineralization by an associated mineral matrix, then $f_2 = c$, where $c =$ const., i.e. the ballast model.

If instead we assume that k is inversely related to time (evidence indicates this relationship to hold for the decomposition of organic carbon in

[1] i.e. the difference between the POC flux entering a infinitesimal depth-layer $z + \Delta z$ and the POC flux leaving that depth-layer, divided by the thickness of that layer.

[2] n.b. these models are discussed in a different order here than in the main text. The order in the main text we found to be most natural in the context of the literature on these models; the order here we found to be most natural in terms of deriving these models sequentially.

sediments across many orders of magnitude [Middleburg, 1989]), i.e. $k = \hat{k}/t$, which assuming $w \approx$ const. implies[3] $k = \hat{k}/t = \hat{k}/(z/w) = \hat{k}w/z$, Eq. (3) becomes

$$-\partial_z(wp) - \Big(\frac{\hat{k}w}{z}\Big)p = 0 \tag{7}$$

The solution to this first-order linear differential equation for p is

$$p(z) = Az^{-\hat{k}} \tag{8}$$

where again A is a constant dependent on initial conditions; multiplying both sides by w and defining $b := \hat{k}$ and $C := wA$ yields

$$f_p(z) = Cz^{-b} \tag{9}$$

i.e. the power-law model. Note this means that the interpretation of b is that it is the remineralization rate of p after one day[4] of decomposition and sinking. We can similarly derive f_p from assuming $w = \hat{w}z$; we then have instead of Eq. (7)

$$-\partial_z(\hat{w}zp) - kp = 0 \tag{10}$$

which similarly leads to a power-law solution for f_p but with a different $b := (k + \hat{w})/\hat{w}$. Though there is evidence for increasing settling velocities with depth [Berelson, 2001], for a sinking particle z is the time-integral of w, making this definition of w in terms of z somewhat ad hoc. Furthermore, $b > 1$ as defined above, unlike in most fits to $f(z)$ data.

If we assume instead of an inverse relationship between k and t a weaker scaling with time such that $k = \hat{k}t^{-s}$, where $s < 1$, we have $k = \hat{k}w^s z^{-s}$ for a depth-dependence of k. This is more consistent with the scaling law found by Middleburg [1989], who found $k \sim t^{-.95}$. Eq. (7) then becomes instead

$$-\partial_z(wp) - (\hat{k}w^s z^{-s})p = 0 \tag{11}$$

which is a first-order linear differential equation for p, the solution to which is

$$p(z) = Ae^{(\frac{\hat{k}w^{s-1}}{s-1})z^{1-s}} \tag{12}$$

where again A is a constant dependent on initial conditions, and multiplying both sides by w yields

$$f(z) = wAe^{(\frac{\hat{k}w^{s-1}}{s-1})z^{1-s}} \tag{13}$$

so if we define $\ell := (s-1)/(\hat{k}w^{s-1})$ and $C := wA$, we find

$$f_s(z) = Ce^{z^{(1-s)}/\ell} \tag{14}$$

i.e. the stretched exponential model. Interestingly, though s can take a range of values $s \in [0, 1)$, typically in the fits to profiles described below we find s close to the 0.95 value from Middleburg [1989]; for instance, in the figure from the main text, $s = 0.90$.

Rather than assuming a decreasing k with time, we can assume that remineralization of p occurs as a second-order kinetic process, i.e. $\frac{dp}{dt} = -\hat{k}p^2$, rather than a first-order process as above. We then have instead of Eq. (3)

$$-\partial_z(wp) - \hat{k}p^2 = 0 \tag{15}$$

which is a first-order nonlinear differential equation for p, the solution to which is

$$p(z) = \frac{w/\hat{k}}{z + Aw/\hat{k}} \tag{16}$$

where again A is a constant dependent on initial conditions, and multiplying both sides by w and defining $a := Aw/\hat{k}$ and $C := w^2/\hat{k}$ yields

$$f_r(z) = \frac{C}{z + a} \tag{17}$$

i.e. the rational model.

[3] n.b. $\hat{k}$ is dimensionless. Throughout, hats are given to parameters whose units are specified similarly, e.g. the units of $\hat{w}$ are 1/d.

[4] $t = 0$ can be thought of as the time at which particles sink through the reference depth horizon.

The last model from the main text, what we refer to as the gamma model, can be derived from the application of the model described by Rothman & Forney [2007] to particles with a constant settling velocity w. The full derivation of their model is involved, so we begin with their main finding for $p(t)$ (in our notation) and refer the reader to their paper for a derivation from first principles. They assume organic carbon is randomly distributed on the mineral surfaces of a porous medium populated by randomly distributed heterotrophic bacteria. A sinking particle certainly fits this description: a porous medium with organic carbon, minerals, and bacteria distributed throughout. They then assume that the decay of a particular parcel of organic carbon is limited by its accessibility to hydrolytic enzymes. Although all of the organic carbon in a given particle is assumed to be equally labile, bacterial remineralization occurs at an increasingly slower rate as bacteria must consume material that becomes increasingly more difficult to access within its heterogeneous microenvironment. They derive as a solution (their Eq. 2) for the POC concentration on a particle with time to be

$$p(t) \propto \int_{k_{min}}^{\infty} \frac{1}{k} e^{-kt}\, dk \tag{18}$$

which yields a flux if we assume a constant w by multiplying by w and replacing $t = z/w$:

$$f(z) \propto \int_{k_{min}}^{\infty} \frac{1}{k} e^{-\frac{k}{w}z}\, dk = \Gamma(0, z/\ell_{max}) \tag{19}$$

where $\ell_{max} := w/k_{min}$, and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function

$$\Gamma(x, y) := \int_{y}^{\infty} v^{x-1} e^{-v}\, dv \tag{20}$$

k_{min} itself is associated with the typical distance between microbes, which is a function of microbial density assuming microbes are randomly distribution within particles. If we subsume all of the prefactors (including w) into a constant C, we are left with

$$f(z) = C\Gamma(0, z/\ell_{max}) \tag{21}$$

i.e. the gamma model.

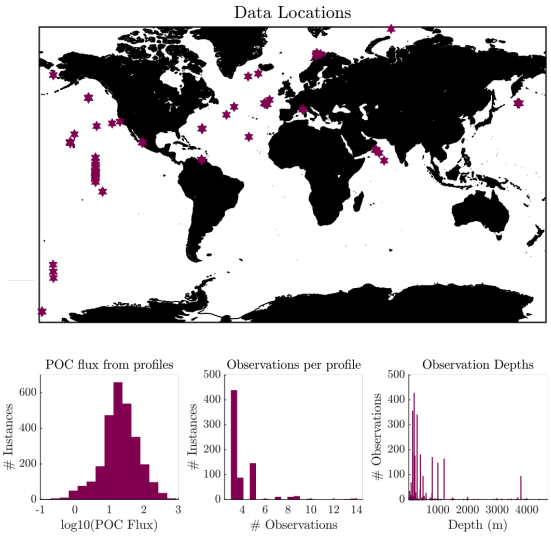


Figure S1. A map of data locations with histograms of data properties. POC flux is reported in mg C m^{-2} d^{-1}.

ii.a. Flux Profiles

We compiled 722 POC flux profiles measured with sediment traps or the ^{234}Th technique, including studies from 1978 to 2014 at depths up to 4680m. The spatial bias in the profile locations reflects that of available data; many of the profiles come from three major time series sites: Bermuda Atlantic Time Series (BATS), Hawaii Ocean Time Series (HOT),and Carbon Retention In A Colored Ocean (CARIACO). When direct download of data was not available (refs. 13, 15, and 17 in the supplementary data), fluxes were retrieved using the 'Web Plot Digitizer,' available online at <https://automeris.io/WebPlotDigitizer/>. Repeat digitizations indicated that digitization introduced ∼1% errors or smaller. Profiles have between 3 and 14 observations, with a mean of 3.9. The duration of deployment for the trap profiles ranges from 1 to 425 days; no corrections were made to account for potential POC remineralization within sample cups [Antia, 2005]. All fluxes are converted to mg C m^{-2} d^{-1} using a molar mass of 12.01 for carbon. Flux observations range from 0 to 1238 mg C m^{-2} d^{-1}, with a mean of 40 mg C m^{-2}. Data are available in an additional supplemental document.

ii.b. ELD and MLD

From each flux-depth model and measurement profile we estimated flux at the ELD and MLD by normalizing to a climatological depth. To generate an ELD climatology, we first downloaded level-3 Chl values from the NASA ocean color website <https://oceancolor.gsfc.nasa.gov/cgi/l3> and constructed monthly climatologies from 1997-2008 observations, the interval over which there were no major sampling issues in the SeaWiFS mission and during which most of the profiles were measured. From this we calculated ELD, defined as the depth of 1% photosynthetically available radiation, according to relationships between Chl and ELD described in Morel et al. [2007] at 9 km spatial resolution. The minimum ELD is 14m, the mean ELD is 62m, and the maximum ELD is 97m. For MLD we used the MLD-DT02 climatology [de Boyer Montegut et al. 2004] which defines MLD as the depth at which temperature varies greater than 0.2°C from its value at 10m and has a spatial resolution of 2°. The minimum MLD is 14m, the mean MLD is 43m, and the maximum MLD is 156m.

ii.c. Normalizations

We normalize the POC flux profiles to each depth climatology using the seven flux-depth models. For each profile we fit each of the seven functions using non-linear least squares regression and minimizing relative error, resulting in optimum parameter values specific to each profile and function. We then estimate export at both depth horizons for each profile at that profile's latitude, longitude, and month, and using these optimum parameter values. When fitting functions with $n > 2$ free parameters, e.g. the ballast model, we restricted fits to profiles with $> n$ measurements, to retain at least one degree of freedom in the fit. To choose unbiased parameter values for normalization by a power law or exponential, we compare the fits of profiles described above to normalizations of only the shallowest observation in each profile with a fixed b or ℓ, respectively. We systematically vary both b and ℓ (to two significant digits) and choose the parameter value as that results in an equal amount of over- and underestimations of export relative to the estimate using profile-specific parameters. In all cases we only normalize measurements that are below the depth of interest. Our results were not sensitive to restricting to subsets of the total set of profiles (e.g. only profiles with a measurement $>$ 1km deep, or with $n \geq 4$). As the results for ELD and MLD were qualitatively similar, we do not expect they are sensitive to the choice of depth horizon. We do not expect the general findings to be sensitive to the choice of depth horizon as the results for ELD and MLD are qualitatively similar, though certainly the magnitude of the disparity between models is magnified at shallower depths.

Power-law- and exponential-normalized export estimates at both the euphotic and mixed layer depth tended to bound estimates from other models, e.g. $f_e < f_r < f_p$. For any of the five models other than f_p and f_e, f_p yielded a larger export estimate for and f_e yielded a smaller export estimate for 76-100% of fits, for both ELD and MLD. Table S1 shows the fraction of profiles for which f_e significantly overestimates or f_p significantly underestimates export relative to the five other models for a nominal measurement uncertainy of 25%. Occasionally f_p sigificantly underestimates relative to f_b or f_g (11% and 14% of the time, respectively), but in general significant over/underestimations are infrequent. We view this as justification for the use of (f_e, f_p) as lower and upper bounds of an estimated export range when depth-normalizing particle flux measurements.

Table S1. Fraction of profile fits for for which f_e significantly overestimates or f_p significantly underestimates export relative to other models, using a nominal measurement uncertainty of 25%. For instance, f_e is less than or within 25% of f_r for 97% of profiles, so only significantly overestimates f_r 3% of the time. The value reported is the larger value between normalizations to ELD and to MLD, e.g. f_e(MLD) significantly overestimates f_r(MLD) 2% of the time while f_e(ELD) significantly overestimates f_r(ELD) 3% of the time, so the ELD value is reported. We use 4/5 rather than 3/4 in the third column because 4/5 is the same *relative* distance from 1 as 5/4; if 3/4 was used instead the values in the table would be the same.

model	# fits	$\% < \frac{4}{5} f_e$	$\% > \frac{5}{4} f_p$
ballast	277	1	11
rational	722	3	0
double	187	0	0
stretched	277	1	0
gamma	722	0	14

For unbiased parameter estimates, we find $b \approx 0.7$

and $\ell \approx 500$m (exact values: $b = 0.69$ for f_p(ELD), $b = 0.68$ for f_p(MLD), $\ell = 530$m for f_e(ELD), and $\ell = 500$m for f_e(MLD); the differences between these exact values and the approximate values reported in the main text are negligible) – see Figure S2. Across all profiles, $f_{0.7}$ (power-law-normalized export using the shallowest measurement in a profile and $b = 0.7$) always overestimates relative to f_{500m} (exponential-normalized export using the shallowest measurement in a profile and $\ell = 500$m).

Minimzing absolute error rather than relative error in the fitting procedure yielded similar quantitative results and did not affect our main conclusions. For instance, f_p(MLD):f_e(MLD) >1.25/1.5/2 for 71/57/39% of profiles, respectively; f_p(ELD):f_e(ELD) >1.25/1.5/2 for 62/35/14% of profiles, respectively. These values are similar but slightly smaller than those for minimizing relative error (see main text; 75/62/49% for MLD and 67/45/25% for ELD). Because larger variability is seen in shallower measurements [e.g. Buesseler, 1991] and uncertainty and variability are typically reported as percentages for these measurements, minimizing relative error is more justifiable in the absence of more comprehensive error information for these measurements.

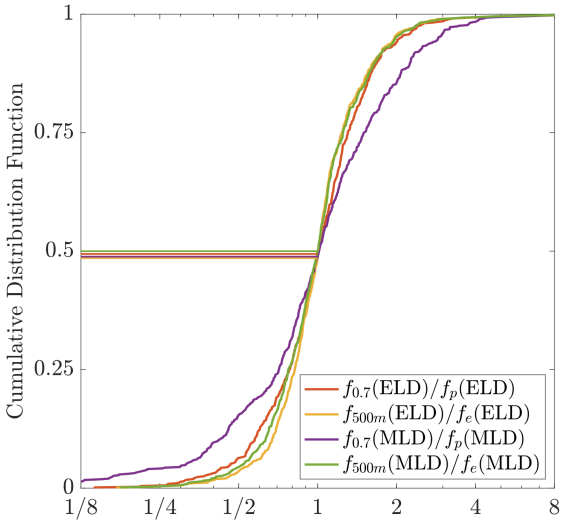


Figure S2. Cumulative distribution function of fixed-parameter vs. fit-parameter normalizations. Horizontal lines indicate the fraction of profiles in each case where the fixed-parameter normalization underestimates relative to the fit-parameter normalization. That these are all close to 0.5 indicates that the approximations of 500 m for ℓ and 0.7 for b introduce negligible bias.

iii. Statistical routine

The statistical routine addresses whether the difference in two models' quality of fit to data has the capacity to be statistically significant at those data's sample number and depth spacing, relative to the uncertainty in the models' quality of fit. For a set of $f(z)$ measurements, data are simulated at the measurements' sample depths from a 'true' model f_T thought to be generating the data; the true model is also fit by a 'false' model f_F to be rejected. The parameters for the true model are taken from fitting the $f(z)$ data as above, and measurement 'noise' with characteristic magnitude ν is superimposed on the simulated data. The root-mean-square error (RMSE) of the simulated data relative to each model provides a metric of goodness-of-fit. Repeating this process many times produces a distribution of RMSE for each model relative to the simulated data, i.e. p_T(RMSE) and p_F(RMSE).

The median RMSE for the true model will be ν by construction, and the false model's RMSE will on average be larger by construction, say $\nu + \varepsilon$. For any one simulated profile, however, the RMSE of the false model may be smaller, because the random noise can result in the false model having a better fit. This difference in these models' median RMSE, ε, may or may not be significantly detectable relative to the variation in RMSE for individual simulated profiles. The probability distributions p_T(RMSE) and p_F(RMSE) quantify uncertainty in the actual goodness-of-fit of f_T and f_F to the simulated data; if these distributions overlap, we cannot be 100% confident that f_T is a better fit, i.e. that $\varepsilon \geq 0$. The distributions yield a bootstrap percent confidence in the 'true' model versus the 'false' one, depending on their overlap. This process is systematically repeated with each model as 'true' and 'false' to yield a 'true'-model-versus-'false'-model table of tolerable measurement variabilities for a given confidence level (as in the figure in the main text), or confidence levels of 'false' model rejection for a given measurement variability (by systematically increasing ν until the estimated percent confidence goes below the chosen confidence level).

The routine can be described stepwise as follows:

1. Choose a set of $f(z)$ data, a bootstrap iteration number N, and a measurement variability level ν.

(The $f(z)$ data can be either a single profile or a composite of profiles. For our analyses we found $N = 10^5$ to be sufficient. We chose $\nu = 0.2$ in the following analyses, corresponding to the lower end of reported measurement variability for sediment trap and ^{234}Th measurements [Buesseler et al., 2000; Stanley et al., 2004; Buesseler et al., 2007]; ν can also be directly estimated from composite profile data, such as that from figure in the main text, where $\nu \approx 0.37 - 0.4$.)

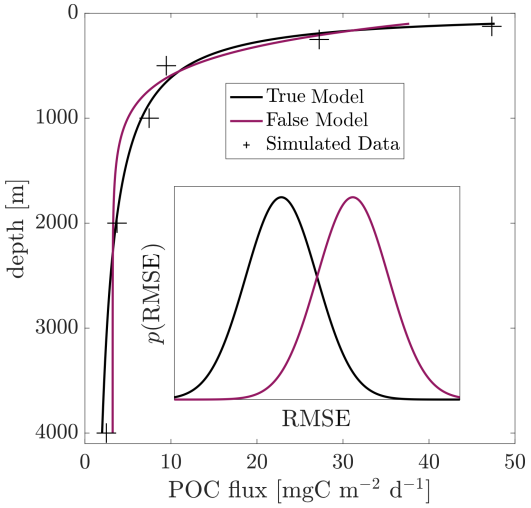
2. Choose a 'true' model f_T and a 'false' model f_F. Select parameters for f_T by fitting to the $f(z)$ data, then select parameters for f_F by fitting to f_T sampled at the depths and resolution of the $f(z)$ data. (For our analyses we fit by minimizing relative error as described in section (ii.c) above. We therefore also simulated measurement variability multiplicatively, i.e. superimposed multiplicative Gaussian measurement error with log-standard deviation ν, and calculated RMSE in terms of relative error. We also repeated these analyses minimizing absolute error, simulating measurement variability additively, and calculating RMSE in terms of absolute error. Doing so yielded similar results. f_F is fit to f_T rather than to $f(z)$ because the objective is to determine how closely different models can fit each other relative to measurement variability, which is included in $f(z)$.)

3. Simulate N measurement profiles sampled from f_T by sampling f_T at the depths and resolution of the $f(z)$ data N times and each time superimposing random variability with characteristic magnitude ν. Compute the RMSE for f_T and f_F for each of the N simulated profiles to generate the probability distributions $p_T(\text{RMSE})$ and $p_F(\text{RMSE})$.

4. Determine the percent confidence with which f_F can be rejected in favor of f_T from maximum percentile range at which $p_T(\text{RMSE})$ and $p_F(\text{RMSE})$ are disjoint. (For instance if the 75th percentile of $p_T(\text{RMSE})$ equals the 25th percentile of $p_F(\text{RMSE})$, f_F can be rejected with 50% confidence).

5. Repeat steps 2-4 for each valid combination of true and false models. (Not including cases where the true model is the same as or a special case of the false model, e.g. exponential vs. exponential, ballast, double, or stretched).

6. If determining tolerable variability thresholds for a chosen confidence level, repeat steps 1-5 above, systematically increasing ν until the estimated percent confidence is lower than the chosen confidence level.



% Confidence in rejecting false model

T vs:	E	P	B	R	D	S	G
E	·	99	·	99	·	·	53
P	90	·	39	7	2	2	57
B	96	46	·	68	·	50	82
R	88	5	35	·	1	<1	4
D	85	17	54	11	·	11	44
S	88	2	38	2	5	·	50
G	31	28	7	2	1	1	·

Figure S3. a) Illustration of the statistical routine. b) Percent bootstrap confidences in rejecting the 'false' (column) model in favor of the 'true' (row) one, from applying the statistical routine illustrated (a) to an artificial profile with observations at $z = [125, 250, 500, 1000, 2000, 4000]$m and 20% measurement variability. Initials are first letter of each model.

Figure S3 illustrates the routine visually. A true model (there, a power-law, whose parameters are taken from the figure in the main text) is fit by a false model (there, a ballast); data is then generated at sample depths (there $z = [125, 250, 500, 1000, 2000, 4000]$m and $\nu = 0.20$). RMSEs of many simulated profiles (only one of which is shown in the figure) result in probability distributions $p_T(\text{RMSE})$ and $p_F(\text{RMSE})$, and the percent confidence of rejecting the false model is determined by the overlap of these probability distributions. Each model is used as both true and false, to generate a model-model-table. For this

artificial profile, an exponential can be rejected versus most other models with reasonable confidence, but other model-model comparisons are generally well below typical significance thresholds.

Table S2 summarizes the result of applying the statistical routine to the profile database. With an assumed measurement variability of 20% (i.e. $\nu = 0.2$) and chosen confidence level of 90%, the table shows the number of profiles out of the total of 722 for which each model could be significantly rejected relative to each other. In general, very few profiles have the necessary vertical resolution and spacing. Significant rejection is largely restricted to profiles that have both high vertical resolution ($n \geq 6$) as well as a combination of shallow, mesopelagic, and bathypelagic observations.

Table S2. Number of profiles (out of a total of 722) for which the percent bootstrap confidence in rejecting the 'false' (column) model in favor of the 'true' (row) one is greater than 90%, assuming a measurement uncertainty of 20%. Initials are first letter of each model.

T vs:	E	P	B	R	D	S	G
E	·	6	·	7	·	·	0
P	6	·	1	0	2	0	0
B	13	7	·	9	·	4	7
R	7	0	1	·	2	0	3
D	11	5	1	8	·	3	3
S	2	1	1	2	3	·	0
G	0	0	2	4	4	0	·

Table S3 illustrates that increasing vertical resolution uniformly increases the percent confidences, but only to a limited extent. When the vertical resolution in Figure S3 is increased to $n = 20$ (logarithmically spaced measurements from 100-4000m), the percent confidences all increase, but many of the models can fit each other so closely that the percent confidence of rejection is still very small in many cases. Table S4 illustrates the importance of vertical extent of measurement profiles for comparing curves; when the same $n = 20$ measurements are logarithmically spaced over 100-1000m instead of 100-4000m, the percent confidence of rejection of curves is drastically reduced; not even the exponential can be significantly rejected in many cases. It is also worth noting that this level of vertical resolution is rarely, if ever, achieved in real observations.

In general the application of this routine to many artificial and actual measurement profiles broadly demonstrates that these models can fit each other so closely relative to the variability in observations that the small differences in their goodness-of-fit are insufficient to significantly reject one model in favor of another. The exception to this is the exponential model for profiles extending into the bathypelagic, where the exponential is known to perform poorly. We find these results to be robust across different combinations of model parameters, different types and magnitudes of simulated variability, and different measurement resolution and spacing.

Table S3. Same as Figure S3b except $n = 20$.

T vs:	E	P	B	R	D	S	G
E	·	99	·	99	·	·	70
P	99	·	53	11	3	10	76
B	99	57	·	77	·	61	95
R	98	7	46	·	7	1	66
D	97	18	70	12	·	11	64
S	98	3	52	3	9	·	68
G	70	98	37	99	2	2	·

Table S4. Same as Table S3 except $\max(z) = 1$km.

T vs:	E	P	B	R	D	S	G
E	·	20	·	4	·	·	8
P	54	·	7	2	1	1	20
B	15	27	·	14	·	5	2
R	39	2	4	·	1	1	8
D	44	7	1	3	·	4	12
S	46	<1	6	1	1	·	14
G	12	12	1	2	<1	9	·

Supplemental References

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

Berelson, W. M. (2001). Particle settling rates increase with depth in the ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(1-3), 237-251.

Middelburg, J. J. (1989). A simple rate model for organic matter decomposition in marine sediments. *Geochimica et Cosmochimica Acta*, 53(7), 1577-1581.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed

layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, 109(C12).

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.