Supporting Information for "Can rates of ocean primary production and biological carbon export be related through their probability distributions?"

B. B. Cael[1,2] Kelsey Bisson[3] and Christopher L Follett[1]

Contents of this file

S1. Re-examining efficiency-production relationships

We re-examine two posited relationships for the relationship between ef, the fraction of primary production exported out of the surface ocean, and depth-integrated production ($\mathcal{P}$, mg C m^{-2} d^{-1}) [*Laws et al.*, 2011; *Maiti et al.*, 2013]. In brief, both are based on analyses of compilations of coincident f and $\mathcal{P}$ measurements; ef is defined as $ef = f/\mathcal{P}$, and the relationship between ef and $\mathcal{P}$ is analyzed; *Maiti et al.* [2013] find an inverse relationship, $ef \sim \mathcal{P}^{-1}$, for $(F, \mathcal{P})$ data from the Southern Ocean, while *Laws et al.* [2011] find a sublinear relationship $ef \sim \mathcal{P}^{0.3}$ for $(f, \mathcal{P})$ data from the global ocean. Note that the relationship between f and p we describe in the main text is not directly comparable with the two relationships considered here (we do not consider $\mathcal{P}$); a comparison would require an analysis of the relationship between p and $\mathcal{P}$.

We examine the *Maiti et al.* [2013] relationship first. *Maiti et al.* [2013] find an inverse relationship between ef and $\mathcal{P}$ data from the Southern Ocean. Substituting the definition of ef as $f/\mathcal{P}$, this relationship becomes $f/\mathcal{P} \sim 1/\mathcal{P}$. It is known that ratios of uncorrelated variables can produce spurious correlations; in particular, for two uncorrelated variables A and B, it has been demonstrated in the aquatic sciences literature that comparing their ratio AB^{-1} with B can produce spurious inverse relationships and correlations [*Berges*, 1997]. This suggests that the measurements of f and $\mathcal{P}$ in their data compilation could be unrelated. We find this to be the case; see Figure 1. The Pearson correlation of these data is not significant at the 95% confidence level.

We then turn to the relationships posited by *Laws et al.* [2011]. *Laws et al.* [2011] analyzed a dataset of coincident f, $\mathcal{P}$, and temperature T (°C) measurements compiled from locations across the global ocean by *Dunne et al.* [2005]. Table S1 summarizes our findings, which are that in the *Dunne et al.* [2005] data, $\mathcal{P}$ is too weakly related to ef to extract a meaningful relationship, and that the relationships posited by *Laws et al.* [2011] appear artificially strong in their analyses either because they only consider a subset of the data or because they fit binned data, which is known to spuriously amplify relationships (see Figure S1) [*Wainer et al.*, 2006].

Laws et al. [2011] posited two relationships for ef as a function of $\mathcal{P}$ and T (their equations 2 and 3). However, one relationship was only tested on a subset of the dataset analyzed, and the other was only tested on binned data; thus these analyses do not address the question of how well the ef data are explained by the $\mathcal{P}$ and T data.

We fit the equations reported by *Laws et al.* [2011], with the parameters reported in *Laws et al.* [2011], to the data analyzed in *Laws et al.* [2011] (compiled by *Dunne et al.* [2005]). This yielded $r^2 = 0.16$ and $r^2 = -0.59$, respectively (a negative r^2 indicates a worse fit than assuming a constant ef). When the same equations are fit without fixing the parameters, both have an $r^2 = 0.47$. However, both include a linear dependency on temperature which alone (i.e. neglecting dependency on $\mathcal{P}$; see Table S1) yields an $r^2 = 0.45$, while in each equation neglecting the temperature dependency in each equation (see Table S1) yields $r^2 = 0.07$. Indeed, Figure S2 demonstrates this straightforwardly, indicating that $\mathcal{P}$ can explain little of the variation in ef in these data.

Equation	Parameters	Ind. vbls.	r^2
$ef = \frac{(a-bT)\mathcal{P}}{c+\mathcal{P}}$	fixed	$T, \mathcal{P}$	0.16
$ef = \frac{(a-bT)\mathcal{P}}{c+\mathcal{P}}$	free	$T, \mathcal{P}$	0.47
$ef = \frac{a\mathcal{P}}{b+\mathcal{P}}$	free	$\mathcal{P}$	0.07
$ef = (a-bT)\mathcal{P}^c$	fixed	$T, \mathcal{P}$	−0.59
$ef = (a-bT)\mathcal{P}^c$	free	$T, \mathcal{P}$	0.47
$ef = a\mathcal{P}^b$	free	$\mathcal{P}$	0.07
$ef = a - bT$	free	T	0.45

Table S1. Summary of our analyses of *Dunne et al.* [2005] data. First column is the model equations used to predict ef from $\mathcal{P}$ and/or T; equations are those considered by *Laws et al.* [2011] or simplifications thereof. Second column indicates whether the parameters in the equation were fit to the data (free) or if their values were taken from *Laws et al.* [2011] (fixed). Third column indicates which variable(s) ef is a function of in each case. Fourth column indicates the percentage of variation in ef explained by the independent variables via each equation.

Therefore it appears that in the data analyzed by *Laws et al.* [2011], there is no useful relationship between ef and $\mathcal{P}$, and that instead the correspondence between independent and dependent variables for their equations 2 and 3 is due to a combination of i) the dependence of ef on T, ii) considering only a subset of the data, and/or iii) analyzing binned data. Note that we do not take issue with their choice to analyze a subset of the data (see their text for details); we merely intend to point out that their equation 2 should not be taken as a global relationship for ef and $\mathcal{P}$ given the above considerations. The lack of a clear relationship between ef and $\mathcal{P}$ could be due to similar reasons

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

0886-6236/18/$12.00

as in the *Maiti et al.* [2013] data, or for other reasons. Our intention herein is not to explain the relationships (or lack thereof) between these variables, but rather to demonstrate that the existing literature on the relationship between ef and $\mathcal{P}$ does not provide satisfactory relationships between production and efficiency (and therefore production and export flux). Figure S1 shows how spurious relationships can emerge from comparing ratios of uncorrelated variables or from binning data.

S2. Compilation of f data

We compiled global shallow (~200m depth and shallower, a depth sufficient to capture instances of deep euphotic and/or mixed layers – most measurements were at 100m or 150m) flux measurements of particulate organic carbon (mg C m^{-2} d^{-1}) from all available sources, including large databases (from programs like BATS, HOT, JGOFS, VERTEX, and so forth), publications, and personal communication. In cases where numerical data was not tabulated in publications the data was digitized from available plots. This yielded a dataset with 1,770 independent measurements from sediment traps and ^{234}Th flux measurements. We supply the variables of interest: particulate organic carbon flux [mg C m^{-2} d^{-1}], depth [m], flux type [trap or thorium], date [in YYYYMMDD], latitude and longitude, and reference. In the supplemental data file, references are numbered 1...62; we provide a supplemental key (PDF file) indicating which number from the data file corresponds to which reference from the main text. To ensure data fidelity and to maximize the number of available observations no adjustments to the data were conducted (e.g. normalizing to a particular depth, averaging by location or time, choosing cutoffs for trap duration or trap type, and so forth). Comparison of the probability distributions of data from 100m vs. 150m, and of the thorium data vs. the trap data, indicates that conflating measurements across sampling depths or methods did not affect our results.

S3. Convergence of the Central Limit Theorem

We briefly give an example of how the convergence of the Central Limit Theorem [*Pitman*, 1999] can be surprisingly rapid.

We take a 100,000 × 3 matrix where each value is sampled uniformly from the unit interval $[0, 1]$, then sum each row, thereby generating 100,000 samples of x_i, where each x_i is the sum of three draws from a uniform distribution (a distribution quite unlike the normal distribution!). Figure S3 shows the PDF of x_i, which looks strikingly like a normal even for the very small value of $n = 3$. Of course, the exact value of n where convergence occurs will depend on many things, i.e. the distributions of the numbers being summed over, correlations in those numbers, the criterion whereby the distribution is considered converged, & so forth, but this gives a demonstration of the nonintuitive rapidity of the Central Limit Theorem's convergence.

S4. Derivation of Equations 5 and 6

Let $f = e^C p^\alpha$, which after taking the logarithm of both sides can be written

$$\ln f = C + \alpha \ln p \tag{1}$$

where the equality is taken in the statistical sense. Because $\ln f$ can be expressed as a continuous monotonic function of $\ln p$, their probability distributions are related through their differentials,

$$P_f(\ln f)\ d\ln f = P_p(\ln p)\ d\ln p \tag{2}$$

Dividing both sides by $d\ln p$,

$$P_f(\ln f)\ \frac{d\ln f}{d\ln p} = P_p(\ln p) \tag{3}$$

which by equation 1 implies

$$\alpha P_f(\ln f) = P_p(\ln p) \tag{4}$$

Dividing by α and replacing $\ln p$ by Equation 1 yields

$$P_f(\ln f) = \frac{1}{\alpha} P_p\Big(\frac{\ln f - C}{\alpha}\Big) \tag{5}$$

Then, substituting in that $P_p(\ln p)$ is a normal distribution with parameters μ_p and σ_p,

$$P_f(\ln f) = \frac{1}{\alpha\sqrt{2\pi}\sigma_p^2} \exp\Big(\frac{(\frac{1}{\alpha}\ln f - \frac{1}{\alpha}C - \mu_p)^2}{2\sigma_p^2}\Big) \tag{6}$$

which can be rearranged as

$$P_f(\ln f) = \frac{1}{\sqrt{2\pi}(\alpha\sigma_p)^2} \exp\Big(\frac{(\ln f - C - \alpha\mu_p)^2}{2\alpha^2\sigma_p^2}\Big) \tag{7}$$

which is a normal distribution with parameters $(\alpha\mu_p + C)$ and $(\alpha\sigma_p)$. Thus the probability denstiy function of f is a lognormal distribution, and its parameters μ_f and σ_f are related to those of p by

$$\mu_f = \alpha\mu_p + C, \quad \sigma_f = \alpha\sigma_p. \tag{8}$$

S5. Lower cutoff for p measurements

When analyzing the complete dataset of p measurements described in the text, we first must discard the 1,173 measurements where $p \leq 0$. Here we are only interested in positive measurements; note that much of the ocean (in principle, all of the ocean below the euphotic depth) has $p \approx 0$. Then, for small values of p, even very small errors or truncation of significant digits can strongly affect the log-transformed p values; to illustrate, a +0.1 error on a measurement of 10.1 causes a difference in log-space of < 0.01, but on a measurement of 0.1 causes a difference in log-space of > 0.69. Therefore we must impose a lower threshold, below which we neglect all samples. Samples below the threshold may not conform to the lognormal for various reasons, including error introduced by rounding, measurement errors, measurement sensitivity, because the null model is a poor one at low p values, or because the lognormal overlaps with a distribution of measurements for samples where $p = 0$ (which because of measurement error will not be a delta-function at $p = 0$).

The choice of threshold is not well-constrained; we use a threshold of 10% of the empirical p distribution's peak height (in log-space), which because the peak occurs at ~5 μg C L^{-1} d^{-1} means the threshold is set at 0.5 μg C L^{-1} d^{-1}. That this threshold separates measurements which neatly follow a lognormal distribution from those which do not can be seen easily in Figure S4a. By computing relative residual error of a lognormal fit as a function of p value, we confirm this cutoff value to be sensible; see Figure S4b. To compute relative residual error, we sort the samples with $p > 0$ into logarithmically spaced bins, and then fit a lognormal distribution to the data above the threshold (see Section S5). For each bin i, the lognormal predicts ℓ_i samples to be in that bin, while there are actually m_i samples in that bin. The relative residual error is then given by

$(m_i - \ell_i)/m_i$. Figure S4b shows that the lognormal underpredicts the number of samples by $\geq 40\%$ for all bins below the threshold, while the average over- or under-prediction for bins above the threshold is 5% and no bin exceeds 12%. This strongly indicates that the samples below the threshold do not follow the same distribution as those above the threshold, i.e. there is a distributional shift at the threshold (given that the samples above the threshold are well-described as lognormally distributed; see Section 4.1 in the main text).

Thus, in the manuscript we only discuss measurements of $p > 0.5$ μg C L^{-1} d^{-1}. As noted in the main text, neglecting measurements below the threshold is reasonable because they constitute <0.1% of all production in the data compilation, and our results are not sensitive to the value of the threshold as we repeated the analyses with the threshold varied by a factor of 2 (i.e. at 0.25 and 1 μg C L^{-1} d^{-1}) and found that this had no impact on the results. Given a complete description of measurement error introduced by the ^{14}C method, one could incorporate samples below the threshold into the following analysis by convolving the lognormal distribution with the measurement error distribution for ^{14}C, as is done in image processing [*Wiener*, 1949]. For consistency, we treat the f data the same way, only analyzing data greater than 10% of peak height (3.6 mg C m^{-2} d^{-1}), and also find that factor of two changes in the threshold value used made no difference to our results.

S6. Kuiper's statistic and other CDF statistics

The standard way to compare a distribution to empirical data is to compare their cumulative distribution functions (CDF) [*Flannery*, 1992]. A variety of metrics to do so exist, though the presence of a lower threshold makes some unavailable, and simulations suggest that all have comparable statistical power for lognormal distributions of large sample size [*Stephens*, 1974]. We use Kuiper's statistic:

$$v = \max\Big(C(p) - c(p)\Big) + \max\Big(c(p) - C(p)\Big)$$

where $C(p)$ is the empirical CDF of p and $c(p)$ is the lognormal fit, i.e. v is the sum of the maximum distance of c above and below C. This modification of the original CDF statistic (the Kolmogorov-Smirnov statistic; see below) is preferred in many applications because it balances ease of interpretation with sensitivity to the tails of the distribution [*Flannery*, 1992]. We note that, like the threshold, our results are not sensitive to this choice of metric; we repeated the subsequent analyses with two other metrics [*Flannery*, 1992], the original Kolmogorov-Smirnov statistic (D):

$$D = \max\Big|C(p) - c(p)\Big|$$

and the Anderson-Darling statistic (D^*):

$$D^* = \max\frac{\big|C(p) - c(p)\big|}{\sqrt{c(p)(1 - c(p))}}$$

and found that this had no impact on the results.

The standard techniques for calculating a p-value from these statistics are not appropriate here. This is due to many factors; the p and f data we test are not randomly distributed in time, depth, or latitude-longitude; measurement error exists and is not fully constrained. Also, sample sizes are large (i.e. n in the thousands or tens of thousands, and CDF metrics are known to be biased against large sample sizes [*Field*, 2009]). Instead, the Kuiper test must be used heuristically to estimate moments of lognormal fits to the p data and to quantify the goodness-of-fits of these lognormal distributions - higher values meaning worse fits. Visual comparison of the PDF of the p datasets with their lognormal fits suffices to demonstrate that in all cases p is well-described as lognormally-distributed, especially given the non-randomness of the p measurements. All of the above also applies to f. We find that for the global p and f distributions and the subregions' p distributions, $v \leq 0.0400$ in all cases, i.e. total deviations above and below between the empirical and hypothesized CDF's sum to $\leq 4\%$, corroborating the visual comparisons. As $v \sim n^{-1/2}$ [*Stephens*, 1974; n.b. this also holds for D and D^*], the value of $v = 0.0400$ for the global distribution of f (n = 1033) is comparable to the value $v = 0.0107$ for that for the global distribution of p (n = 38334) when accounting for differences in sample size. Using the tests described in *Stephens* [1974], v, D, and D^* indicate that the time series' p distributions are inconsistent with a lognormal distribution, and that the time series' and biomes' f distributions are consistent with a lognormal distribution, though the smaller sample sizes for these latter datasets ($n \in [115, 447]$) afford less statistical power to reject a lognormal distribution.

Figures

Figure S1. Illustration of spurious relationships. a) scatterplot of x and y, two uncorrelated random variables sampled from a uniform distribution ($n = 200$). b) y/x plotted against x (dashed line is $1/x$; compare to *Maiti et al.* [2013]). c) X and Y are binned averages of x and y, into five bins; dashed line is two-parameter fit ($Y = 0.45; X^{-0.08}; r^2 = 0.99$); compare to *Laws et al.* [2011]. 1/8 random sample sets will have a monotonic relationship such as the one seen here if binned into five bins ($2^{-4} = 1/16$ will be monotonically increasing and 1/16 will be monotonically decreasing; this number increases rapidly even in response to very weak correlations, e.g. $\rho = 0.01$).

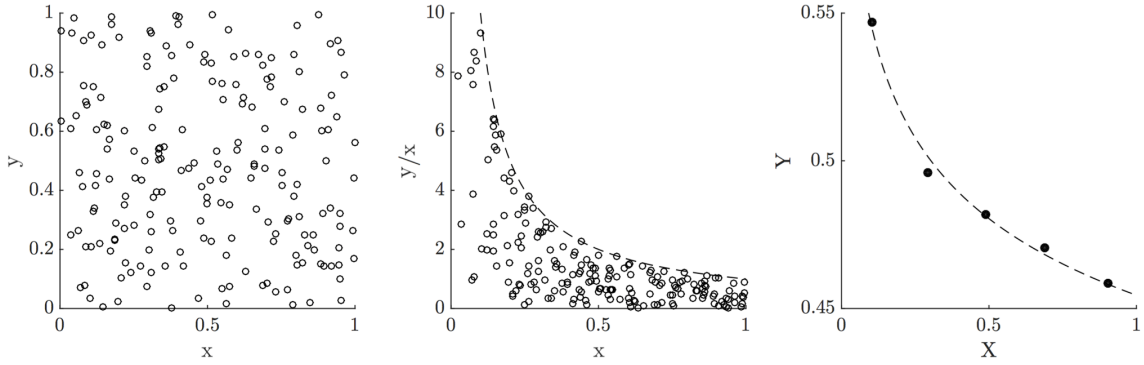

Figure S2. $\mathcal{P}$ and ef data from *Laws et al.* [2011], originally compiled by *Dunne et al.* [2005]. Metrics reported are from Pearson's correlation: the fraction of variance of explained (R^2) and the P-value of the correlation.

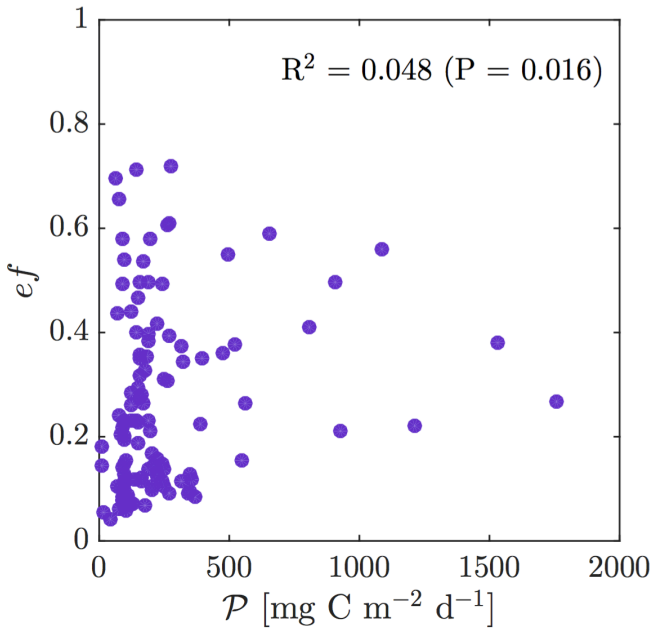

Figure S3. PDF of sums of triplets drawn from the unit uniform distribution, with normal distribution overlaid, illustrating the rapid convergence of the Central Limit Theorem.

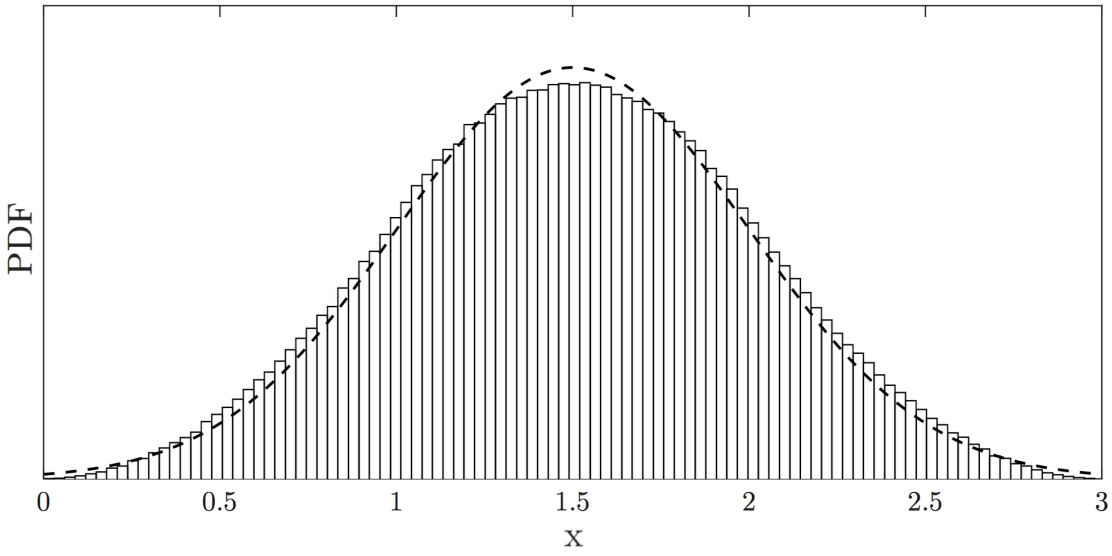

Figure S4. Identification of the lower threshold for p samples considered. a) Bars are PDF of measurements of $p > 0$; dashed black line (for both panels) is the threshold identified at 0.5 μg C L^{-1} d^{-1}, or ~10% of the peak in the log-transformed distribution. Green line is a lognormal probability distribution function above the threshold. b) Bars are the relative residual error of the lognormal distribution shown in panel (a).