Heliyon

ELSEVIER

Received:
24 August 2016
Accepted:
7 September 2016

Heliyon 2 (2016) e00156



Enigmatic relationship between chlorophyll a concentrations and photosynthetic rates at Station ALOHA

E.A. Laws [a,*], **R.R. Bidigare** [b,c], **D.M. Karl** [b,d]

[a] *College of the Coast & Environment, Department of Environmental Sciences, Louisiana State University, Baton Rouge, USA*

[b] *Daniel K. Inouye Center for Microbial Oceanography and School of Ocean and Earth Science and Technology, Department of Oceanography, University of Hawaii, Honolulu, USA*

[c] *Hawaii Institute of Marine Biology, University of Hawaii, Kaneohe, USA*

[d] *Department of Oceanography, University of Hawaii, Honolulu, USA*

* Corresponding author.

E-mail address: edlaws@lsu.edu (E.A. Laws).

Abstract

An ordinary least squares (OLS) analysis of the relationship between chlorophyll *a* (chl *a*) concentrations and photosynthetic rates at depths of 5 and 25 m at Station ALOHA produced a slope that was only 28% of the mean productivity index at those depths and an intercept at zero chl *a* that equaled 70% of the mean photosynthetic rate. OLS regression lines are known to produce a slope and intercept that are biased estimates of the true slope and intercept when the explanatory variable, *X*, is uncontrolled, but in this case the measurement errors and natural variability of the chl *a* concentrations were much too small to explain the apparent bias. The bias was traceable to the fact that the photosynthetic rates were determined by more than one explanatory variable, a source of variability that is typically overlooked in discussions of OLS bias. Modeling the photosynthetic rates as a function of the product of chl *a* and surface irradiance produced a much more accurate and realistic description of the data, but the OLS continued to be

http://dx.doi.org/10.1016/j.heliyon.2016.e00156

2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

biased, presumably because the photosynthetic rates were functions of factors in addition to chl *a* and surface irradiance (e.g., temperature, macronutrients, trace metals, and vitamins). The results underscore the need to recognize that the absence of bias in an OLS when *X* is not controlled implies that all scatter in the data about the OLS is due to errors in the dependent variable, an unlikely scenario. In most cases, resolution of the bias problem will require identification of the explanatory variables in addition to *X* that determine the dependent variable.

Keywords: Environmental science, Biogeoscience

1. Introduction

A major challenge in marine ecology has been to describe and understand functional relationships between and among populations and their environment. The subdiscipline of macroecology [1] has become even more important as we endeavor to document large-scale trends in ecosystem structure and dynamics under threats of anthropogenic change, including, but not limited to, greenhouse gas-induced warming. Satellite observations of ocean color provide a mechanism for detecting spatial and temporal distributions of phytoplankton on a global ocean scale based on established empirical relationships between ocean color (i.e., spectral radiance or reflectance) and chl *a* [2]. Over time, the algorithms have evolved, especially for chl *a* detection in the low nutrient, low biomass-containing oligotrophic gyres [3]. Satellite-based chl *a* observations have also been used to estimate photosynthetic rates on regional scales using functional relationships between primary production and chl *a* [4] and globally using more comprehensive bio-optical and physiological models [5, 6]. More recently, these satellite-based productivity models have been extended to estimate global net community production and the metabolic state of oligotrophic marine ecosystems like Station ALOHA [7], and in combination with food-web models, used to estimate global carbon export from the euphotic zone [8]. The accuracy of these satellite-based, productivity-carbon flux models relies entirely on the functional relationship between chl *a* and primary production.

Station ALOHA was established in 1988 at 22°45'N, 158°W in the North Pacific subtropical gyre as the site of A Long-term Oligotrophic Habitat Assessment (ALOHA), the most extensive of which has been the Hawaii Ocean Time-series (HOT), which involves four-day research cruises to the site on almost a monthly basis. The multi-decadal dataset available from those studies (http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html) has provided a wealth of information with which to test hypotheses about, inter alia, the functional relationship between chl *a* and primary production.

In the present study, we focused on the relationship between *in situ* photosynthetic rates measured by the ^{14}C method [9] and chl *a* concentrations at Station ALOHA

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

at depths of 5 and 25 m, where previous work has indicated that photosynthetic rates are light-saturated [10,11] and hence a function only of the biomass of the phytoplankton and of their physiological status.

2. Materials and methods

Water samples for photosynthetic rate measurements and pigment analyses were collected on more than 200 cruises from 20 January 1994 to 21 December 2013. The sampling and incubation protocols have previously been described by Karl and Hebel [12]. We analyzed the results from samples collected at depths of 5 and 25 m because average photosynthetic rates during the photoperiod at those two depths over the course of the HOT program are virtually identical, 6.56 ± 0.15 and 6.52 ± 0.16 mg C m^{-3} d^{-1}, respectively, where the error bounds are the standard deviations of the mean values based on 259 and 256 measurements, respectively. Our analysis is based on estimates of photosynthetic rates during the photoperiod (i.e., dawn-to-dusk). Measurements of chl *a* were made by high-pressure liquid chromatography [HPLC, [13]], and the reported values are the sum of monovinyl and divinyl chl *a*.

Validation assays for monovinyl chl *a* were run on 10 occasions in our laboratory. Each assay involved three amounts of chl *a* per injection, 94 ng, 281 ng, and 548 ng. The 10 HPLC assays were run over a time period of 18 days. To estimate the natural variability of chl *a* concentrations, we carried out a time-series study of chl *a* concentrations at Station ALOHA from 23 August to 7 September of 2012. Samples were collected daily from a depth of 5 m at 1200 h and every third day at 1200 and 1600 h.

3. Results

Examination of the relationship between the concentrations of chl *a* and photosynthetic rates revealed that 95% of the productivity indices (*PI*s; ratios of photosynthetic rates to chl *a*) fell within the range 3.1–12.1 g C g^{-1} chl *a* h^{-1} (Fig. 1). The mean and median *PI* were 7.0 and 6.9 g C g^{-1} chl *a* h^{-1}, respectively. To our surprise, there was very little correlation between chl *a* concentrations and photosynthetic rates. Although the correlation coefficient was significant at $p = 3 \times 10^{-11}$, an ordinary least squares (OLS) regression explained only 11.2% of the variance of the photosynthetic rates. The slope of the OLS regression line with chl *a* as the explanatory variable was 1.96 g C g^{-1} chl *a* h^{-1}, 28% of the mean and median *PI*s. The 95% confidence interval to the slope was 1.40–2.52 g C g^{-1} chl *a* h^{-1}. It is axiomatic that the photosynthetic rate must be zero when the chl *a* concentration is zero, and the OLS regression line in Fig. 1 clearly does not satisfy that constraint. Instead, the intercept of the regression line equaled 70% of the mean photosynthetic rate and was statistically significant at $p = 2 \times 10^{-52}$.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

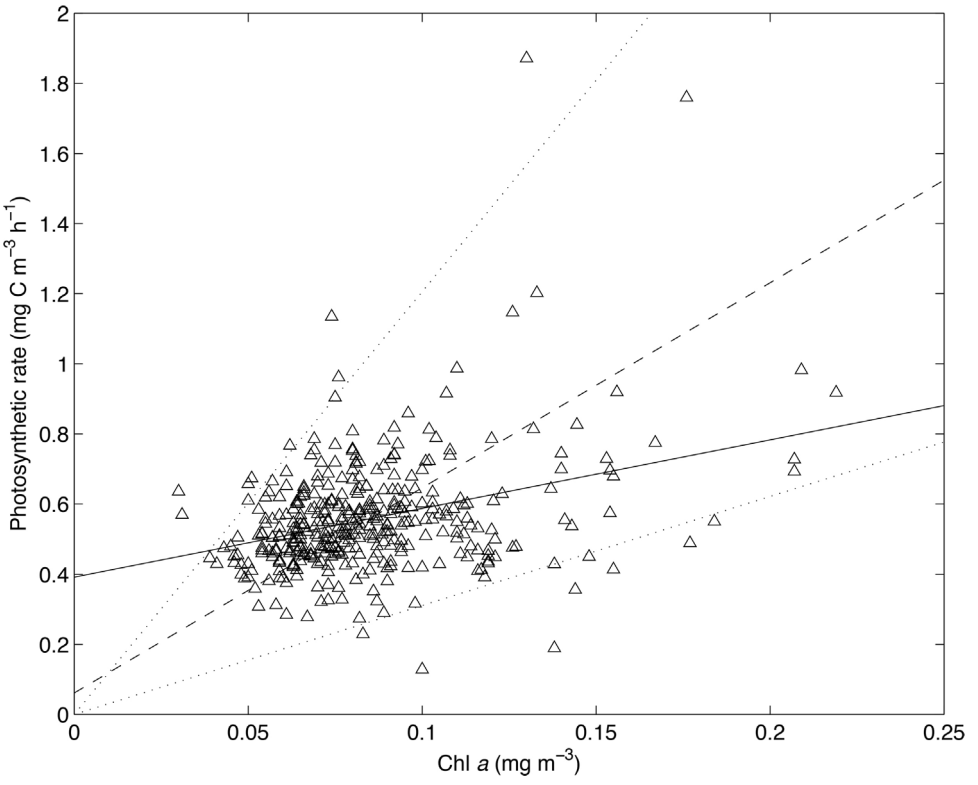


Fig. 1. Relationship between chl *a* concentrations and photosynthetic rates at Station ALOHA at depths of 5 and 25 m. The solid line is the OLS regression line. The two dotted lines define the region within which 95% of the productivity indices lie and correspond to *PI*s of 3.1 and 12.1 mg C mg^{-1} chl *a* h^{-1}. The dashed line is a model II geometric mean regression line fit to the data.

4. Discussion

If the OLS regression were unbiased, the slope of the line would be an estimate of the average *PI*, 6.9–7.0 g C g^{-1} chl *a* h^{-1}. In fact, the 95% confidence interval to the slope did not even overlap with the range of 95% of the *PI*s in the dataset. Our results suggest that the OLS regression lines in Fig. 1 is seriously underestimating the slope of the functional relationship between photosynthetic rates and chl *a* concentrations, but why? It is well known that the magnitude of the slope of the functional relationship between two variables X and Y will tend to be underestimated if there are errors in the explanatory variable X and X is not under the control of the investigator [14, 15]. Under these conditions, when the number of data in the dataset becomes sufficiently large, the value of the OLS slope, B, approaches a value given by

$$B = \beta\left(1 - \frac{V(\delta)}{V(X)}\right) \tag{1}$$

where β is the true slope, δ is the error in X, and $V(\delta)$ and $V(X)$ are the variances of δ and the measured values of X, respectively. If the error δ is independent of the

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

true value of X, X_t, then $V(X) = V(X_t) + V(\delta)$. Therefore $0 \leq \frac{V(\delta)}{V(X)} \leq 1$, and a consequence of the X values being uncontrolled observations is that the OLS slope approaches a value smaller in magnitude than the true slope, β. The error in X includes both measurement errors and what Ricker [14] characterized as natural variability, the error "inherent in the material being measured" [14,p. 410]. Ricker went on to point out that, "In practice, there seem to be few situations in biology where both Y and X are subject to measurement error alone... With biological materials very frequently 80% or more of the variability is natural" [14,p. 424,425].

Ricker's [14] recommendation for estimating the functional relationship between X and Y when the X observations are uncontrolled was to describe the data with a model II geometric mean regression line, in which case the slope of the line is equal to the OLS slope divided by the absolute value of the correlation coefficient between X and Y. That line is indicated by the dashed line in Fig. 1. The slope of the geometric mean regression line was 5.85 g C g^{-1} chl a h^{-1}, which is 84% of the mean PI, and the intercept, 0.06 g C m^{-3} h^{-1}, was much closer to zero than the intercept of the OLS regression line. However, the geometric mean regression line did a very poor job of accounting for the variance in the data. In fact, the variance about the geometric mean regression line was 33% larger than the variance about a horizontal line drawn through the mean of the photosynthetic rates. In other words, the geometric mean regression line accounted for −33% of the variance of the photosynthetic rates, i.e., it provided a worse description of the data than a horizontal line drawn through the mean of the photosynthetic rates.

To determine why the data were so scattered, we first tried to estimate the contribution of measurement error and natural variability to $V(\delta)$. The validation assays for monovinyl chl *a* produced standard deviations that were 0.5%, 1.0%, and 0.4% of the mean values of the 94 ng, 281 ng, and 548 ng standards, respectively. The mean and variance of the chl *a* data in Fig. 1 were 84.5 μg m^{-3} and 782 μg^2 m^{-6}, respectively. If V(δ) in Eq. (1) is the square of 1.0% of 84.5 μg m^{-3}, then the V(δ)/$V(X)$ ratio is less than 1×10^{-3} and cannot explain why the OLS slope would be seriously underestimating the true slope.

The cause of the bias in the OLS slope would therefore seem to be natural variability. We estimated the natural variability of the chl *a* concentrations based on the results of the time-series study from 23 August to 7 September 2012. There was no significant correlation between the times the samples were collected and the chl *a* concentrations ($p = 0.23$), and the ratio of the standard deviation (7.6 μg m^{-3}) to the mean (81 μg m^{-3}) of the chl *a* concentrations was 0.094. This implies a somewhat larger error than the estimate based on analytical precision but would still account for only a 8% bias in the OLS slope.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

An alternative estimate of natural variability can be obtained by noting that

$$\frac{dC}{dt} = (chl\,a)\left(\frac{1}{chl\,a}\,\frac{dC}{dt}\right) = (chl\,a)PI \tag{2}$$

where $\frac{dC}{dt}$ and (chl a) are the true photosynthetic rate and chl a concentration, respectively, and $\left(\frac{1}{chl\,a}\,\frac{dC}{dt}\right)$ is the true productivity index, PI. If the true PIs are constant, Eq. (2) will not create any problems, but if the true PIs are variable, an OLS of chl a versus photosynthetic rate will be biased, even in the absence of measurement errors. Eq. (2) can be rewritten

$$\frac{dC}{dt} = (chl\,a)(PI - \overline{PI} + \overline{PI}) = (chl\,a)\left(1 + \frac{\Delta PI}{\overline{PI}}\right)\overline{PI} \tag{3}$$

where $\overline{PI}$ is the average of the PIs, and $\Delta PI = PI - \overline{PI}$. The assumption of OLS is that the dependent variable is a linear function of an explanatory variable, but it is apparent from Eq. (3) that in this case the explanatory variable is not the chl a concentration. Instead, the explanatory variable is the chl a concentration plus $(chl\,a)\Delta PI/\overline{PI}$. Thus assuming that the explanatory variable is the chl a concentration creates an error in X, the error inherent in the material being measured that Ricker [14] discussed. If ΔPI is independent of the chl a concentration, then this error term is independent of the chl a concentration, and although in this case the error is very likely heteroscedastic, the analysis leading to Eq. (1) requires only that the δ be independent of the X_t. The implication of this analysis is that when the slope of the functional relationship between X and Y is not constant, the variability of the slope causes the OLS slope to be a biased estimate of the average value of the slope.

To determine how much bias this error might be causing, we first examined the relationship between chl a concentrations and photosynthetic rates during three-month time intervals (Fig. 2). There was no significant correlation ($p = 0.41$) between chl a and photosynthetic rates during months 2–4 (February through April) (Fig. 2A), but the correlations were significant ($p < 0.0003$) during the other three-month time intervals. The results were particularly noteworthy during months 5–7 (May through July). During that time interval, the mean PI was 8.5 g C g^{-1} chl a h^{-1}, and the slope of the OLS was 8.8 with a 95% confidence interval of 7.1–10.4 g C g^{-1} chl a h^{-1}. The intercept of the OLS was not significantly different from zero ($p > 0.05$). Thus there was no evidence of any bias in the OLS of chl a versus photosynthetic rate during months 5–7. The absence of bias implies that all of the scatter about the OLS for those months was due to errors in the photosynthetic rates and that ΔPI was either zero or a very small fraction of $\overline{PI}$ (Eq. (3)). The OLS for those three months accounted for 53% of the variance in the photosynthetic rates (Fig. 3B). Because errors in the dependent variable create scatter but no bias, the implication is that during months 5–7 errors in the

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

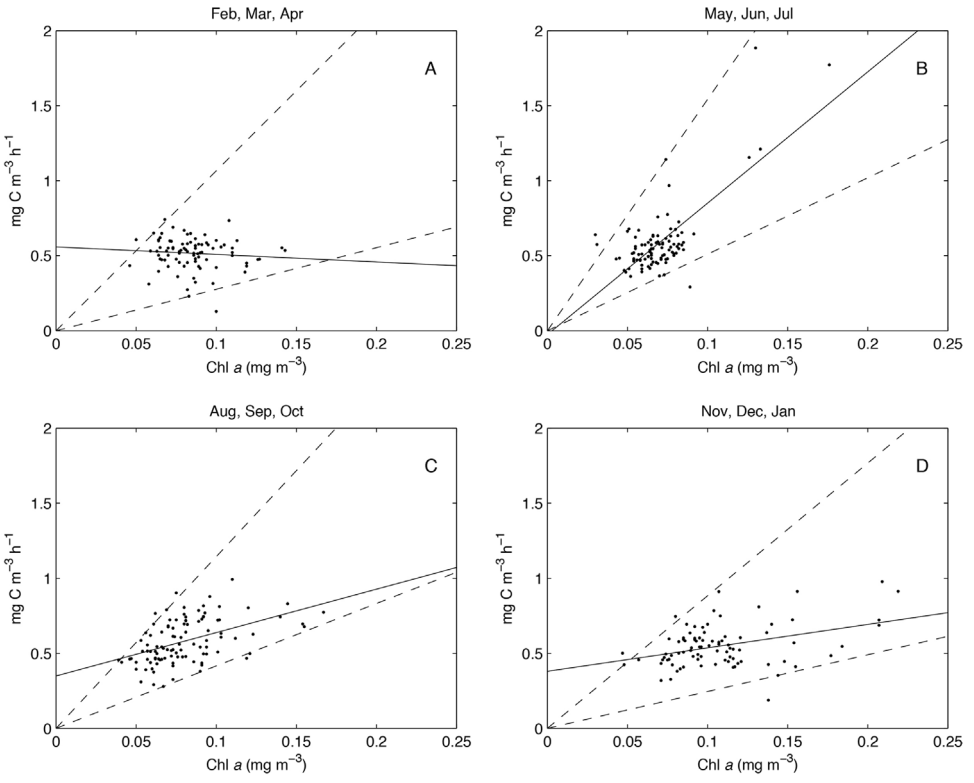


Fig. 2. Data from Fig. 1 sorted by three-month intervals.The mean *PI*s are 6.4, 9.4, 7.7, and 5.0 mg C mg^{-1} chl *a* h^{-1} in panels A, B, C, and D, respectively. Dashed lines are the 95% confidence intervals to the mean *PI*s in each panel, and the solid lines are the OLS regression lines fitted to the data in each panel.

photosynthetic rates were responsible for about 47% of the variance of the photosynthetic rates.

An analysis of the variance of the *PI*s revealed that between-month differences of the *PI*s were much greater than within-month differences (Fig. 3A). The fact that the OLS of the data from May through July was unbiased seems to reflect the fact that the between-month differences were unusually small during those months, $\overline{PI}$ was a maximum, and therefore $\Delta PI/\overline{PI}$ in Eq. (3) was smaller than during any other three-month time interval. Comparison of the pattern of monthly median *PI*s with monthly median surface photosynthetically active radiation (PAR) (Fig. 3B) revealed that the patterns were very similar. Maximum and minimum values occurred in May–June and December–January, respectively. In fact, an OLS regression line of monthly median PAR versus monthly median *PI* accounted for 82% of the variance of the latter (Fig. 4). In contrast, sea surface temperature at Station ALOHA reaches a maximum of about 26.5 °C in September and a minimum of about 23.3 °C in February–March [16] and is therefore out of phase with the PAR and PI data by about 2–3 months.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

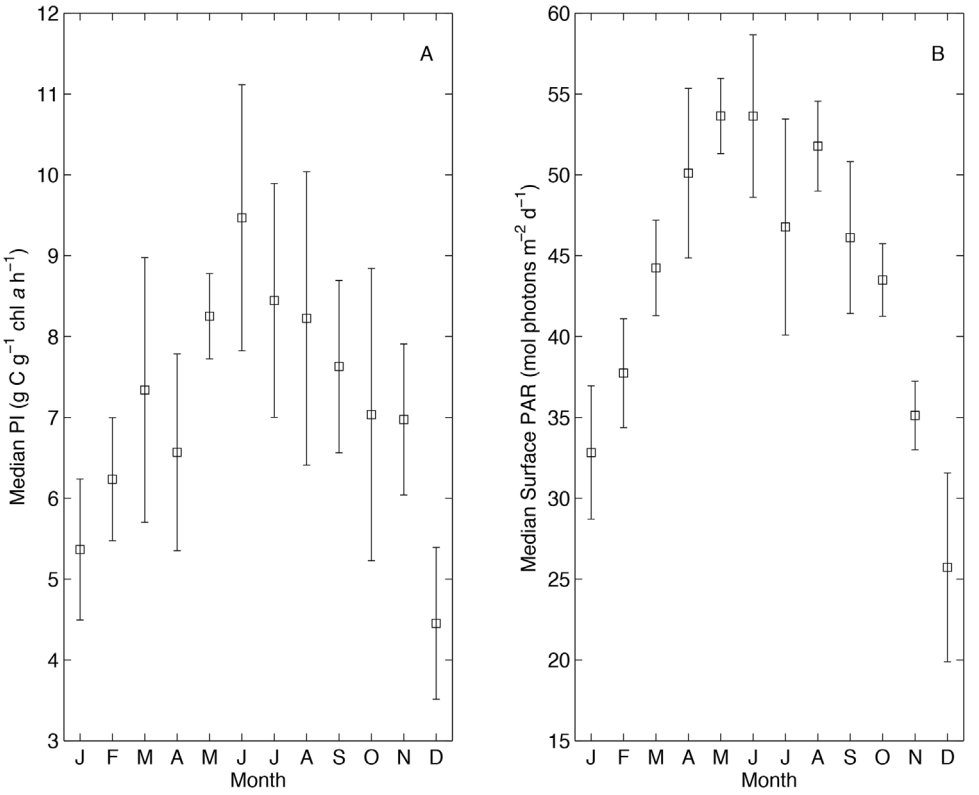


Fig. 3. Monthly median *PI*s (A) and surface photosynthetically active radiation (PAR, 400–700 nm) (B) at Station ALOHA. Error bars are median absolute deviations. Data in panels A and B correspond to the time intervals 1994–2013 and 1994–2003, respectively.

Based on the relationship between monthly median surface PAR and *PI* values at 5 and 25 m (Fig. 4), we hypothesized that the *PI*s could be approximated as being directly proportional to PAR on each sampling day. We therefore used the product of the chl *a* concentrations and the corresponding PAR values as the explanatory variables for regression purposes. Accurate estimates of surface PAR for the HOT program are available during the 10 years from 1994 through 2003, and results of a regression model of photosynthetic rate versus the product of chl *a* and PAR are shown in Fig. 5. In this case an OLS accounted for 33% of the variance in the data (versus 11% in Fig. 1), and the intercept of the regression line, 0.24 g C m^{-3} h^{-1}, was 60% of the OLS intercept in Fig. 1. A geometric mean regression passed directly through the origin (intercept = 0.00 g C m^{-3} h^{-1}) and accounted for 15% of the variance in the data (versus −33% in Fig. 1).Thus explicitly taking into account the cause of some of the natural variability of the relationship between photosynthetic rates and chl *a* concentrations dramatically improved the goodness of fit of both the OLS and geometric mean regression lines, and in the latter case the regression line appears to be a very good representation of the underlying functional relationship.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

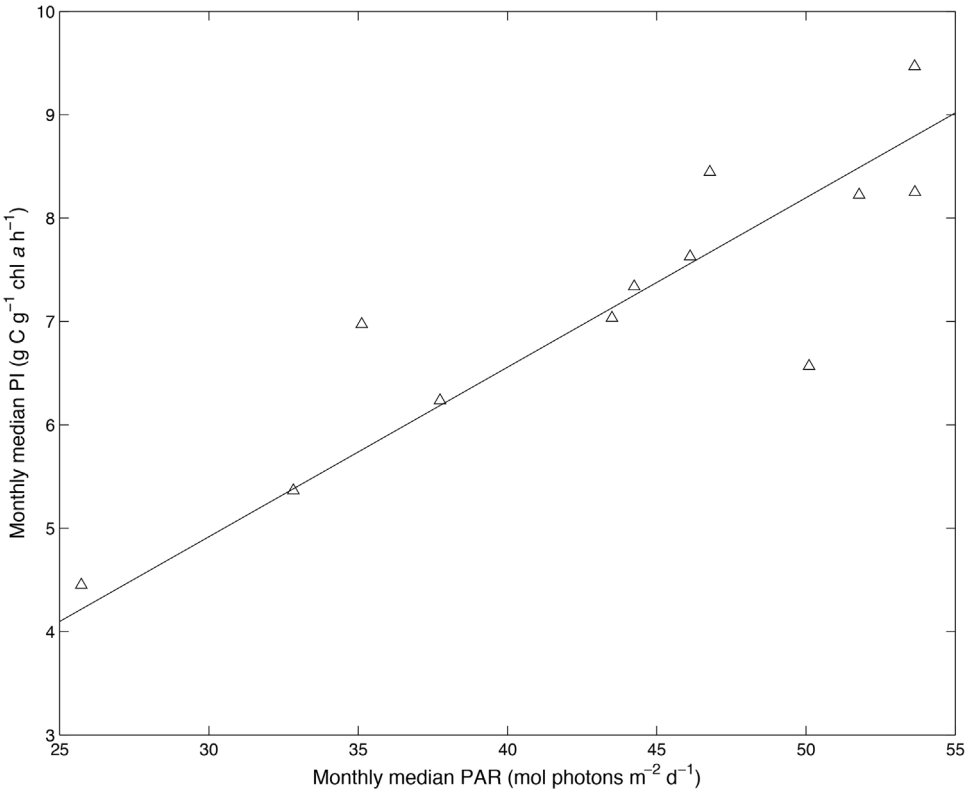


Fig. 4. Monthly median PAR versus monthly median PI. The straight line is a linear regression forced through the origin.

The example of natural variability cited by Ricker (1973) was the heights of brothers and sisters. While superficially of no relevance to photosynthetic rates, that example nevertheless illustrates several important points. First, the height of a brother is determined by factors other than the height of his sister, and the height of a sister is determined by factors other than the height of her brother. Thus no matter whether the height of brothers or sisters is assumed to be *X*, there is natural variability in *X* associated with the fact that *X* is only an approximation of the explanatory variable that determines the functional relationship between the heights of siblings. Second, repeated measurements of the heights of adult brothers and sisters at time intervals of days, weeks, or years will provide no clue as to the magnitude of this natural variability.

Unfortunately, natural variability in the context of regression analysis has too often been interpreted to mean the natural variability of only *X*, without regard to the possibility that *Y* may be a function of explanatory variables in addition to *X*. Thus, for example, Calbet and Prairie [17] argue that an OLS regression line fit to a log-log plot of primary production (PP) versus mesozooplankton biomass-specific ingestion rates is unbiased because "literature reports of within-samples variability in PP measurement provide a coefficient of variation of 10%... [and that]

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

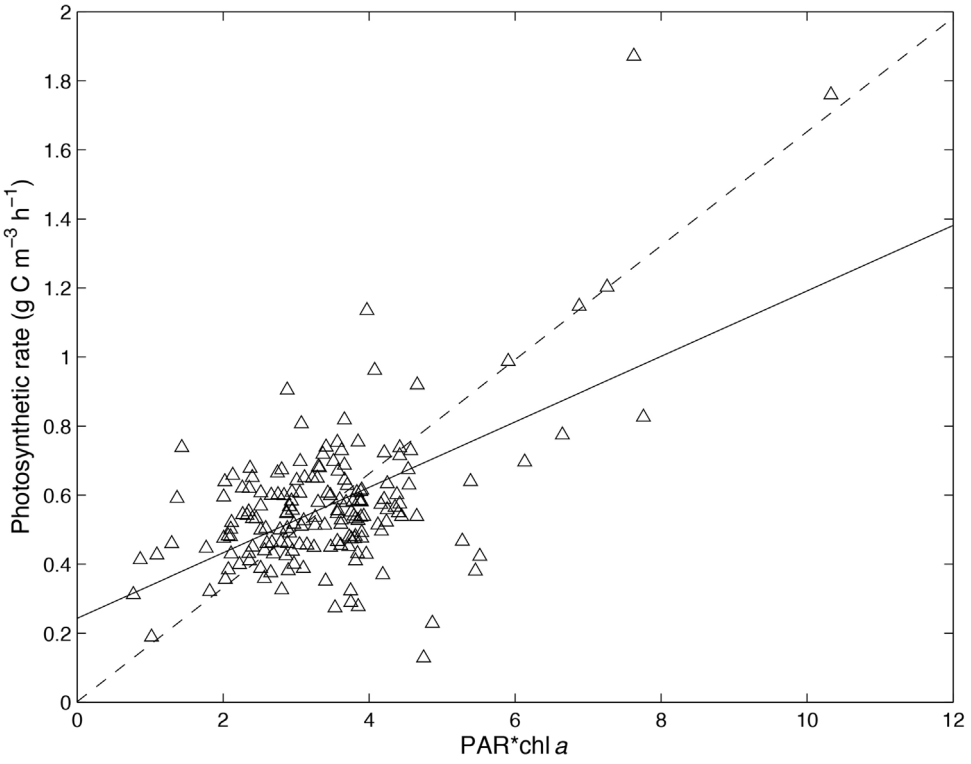


Fig. 5. The product of surface PAR (mol photons m^{-2} d^{-1}) and chl *a* (mg chl *a* m^{-3}) versus photosynthetic rate at depths of 5 and 25 m at Station ALOHA during 1994–2003. The straight line is an OLS regression. The dashed line is a model II geometric mean regression.

accounting for the error on the X axes does not raise significantly the slope of the relationship" [17,p. 1360]. Had we followed a similar line of reasoning, we would have concluded that the OLS in Fig. 1 was a good approximation of the underlying functional relationship because estimates of the natural variability of chl *a* concentrations can explain a bias of no more than about 8%. Such arguments, however, ignore the possibility that much of the natural variability is due to the fact that *Y* is a function of more explanatory variables than *X*, the result being variability in the slope of the relationship between *X* and *Y*. As is apparent from Eq. (3), variability in the slope of the relationship causes an effect that is mathematically indistinguishable from errors in *X*, the result being that an OLS analysis produces a slope that tends to underestimate the average slope of the true relationship. The variability of the slope is typically caused by the fact that variables in addition to *X* affect *Y*, as is the case with the heights of brothers and sisters. In the case of the photosynthetic rates and chl *a* concentrations at Station ALOHA, one of those additional variables appears to be surface PAR (Fig. 4). Other factors, such as temperature and concentrations of macronutrients, trace metals, and/or vitamins, may exert additional controls on photosynthetic rates in the surface waters of Station ALOHA. Likewise, it is quite possible that factors

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

other than primary production rates affect mesozooplankton biomass-specific ingestion rates, a possibility ignored by Calbet and Prairie [17].

Ondrusek and Bidigare [11] developed a mechanistic, full spectral bio-optical model to estimate primary production at Station ALOHA. Their results indicated that primary production in the subtropical North Pacific Ocean may have been systematically underestimated by the Behrenfeld and Falkowski [6] model. They occasionally observed high primary production rates that were not manifested as increased satellite-based chl *a* concentrations. Productivity indices are the product of the growth rates of the phytoplankton, $\frac{1}{C}\frac{dC}{dt}$, and the C/chl *a* ratio of the phytoplankton. At Station ALOHA, variability of phytoplankton growth rates and C/chl *a* ratios at 5 and 25 m obscure the functional relationship between chl *a* concentrations and photosynthetic rates, with the exception of the time interval from May through July (Fig. 2). Even if functional relationships had been clearly discernible from OLS regressions during three-month time intervals, the seasonal variability of the *PI*s would have been sufficient to seriously obscure the functional relationship between chl *a* and photosynthetic rates averaged over one year. The *PI*s varied significantly over three-month time intervals (Kruskal-Wallis test, $p < 0.001$), and the mean *PI*s within those three-month time intervals varied by almost a factor of 2, from 5.0 to 9.4 g C g^{-1} chl *a* h^{-1} (Fig. 3). Similarly, there is no predictable functional relationship between ^{14}C-based primary production (^{14}C-PP) and carbon export from the euphotic zone at Station ALOHA [18].

The lack of correlation between chl *a* and ^{14}C-PP at Station ALOHA undoubtedly results from both analytical errors and the variability of physiological/ecological processes. Analysis of replicate measurements of photosynthetic rates at 5 and 25 m at Station ALOHA on roughly 200 cruises from January 1994 through December 2013 indicated that the coefficients of variation were log-normally distributed with a median value of 8%, and we estimate the coefficient of variation of our chl *a* concentrations to be 9–10% (*vide supra*). However, the accuracy of the ^{14}C-PP rates is unknown and possibly varied over the 20-year observation period. The ^{14}C-method measures a rate that lies between gross primary production (GPP) and net primary production (NPP), and the relationships between ^{14}C-PP and GPP and NPP vary with depth and season [19]. A portion of that variability may be attributable to bottle confinement; however, this impact is difficult to quantify. Near-surface GPP at Station ALOHA appears to be at least twice ^{14}C-PP [19]. Thus *PI*s based on GPP would be more than twice the values reported here. Furthermore, GPP and *PI* are dependent upon the growth rate and C/chl *a* ratio of the phytoplankton assemblage. These parameters vary with light, nutrient concentrations, temperature, and phytoplankton community structure, all of which have varied at Station ALOHA over the 20-year observation period. Finally, we now recognize that oligotrophic ocean gyres are dynamic, non-steady-state ecosystems, where stochastic physical processes can lead to a decoupling of the

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

balance between production and loss of phytoplankton. Consequently, instantaneous assessments of chl *a* concentrations and ^{14}C-PP may fail to reveal fundamental ecological relationships within these oligotrophic systems.

The scatter in any graph of *X* versus *Y* is due to errors in *X* and errors in *Y*, with errors understood to include both measurement errors and errors "inherent in the material being measured" [14,p. 410]. If the errors in *Y* are independent of *X*, then an OLS of *X* versus *Y* when *X* is not controlled will be biased only by errors in *X*. To argue that an OLS regression line is unbiased when *X* is not controlled therefore implies that all scatter about the OLS is due to errors in *Y*. A more likely scenario is that *Y* is a function of more than one explanatory variable. The argument that there is negligible bias in an OLS of *X* versus *Y* when *X* is uncontrolled is frequently based on estimates of errors in *X* derived from repeated measurements of *X*. Such an argument ignores the possibility that factors in addition to *X* may determine *Y* and fails to consider the implication of the argument, namely that all the scatter about the OLS is due to errors in *Y*.

To avoid misinterpreting the results of an OLS when *X* is not controlled, it is best to consider the possibility that *Y* is a function of variables in addition to *X*. In the case of the Station ALOHA chl *a* and photosynthetic rate data, it is straightforward to identify the additional explanatory variables that determine the photosynthetic rates: the C/chl *a* ratios and growth rates of the phytoplankton. In other cases it may not be so obvious, and it is best to proceed with caution, particularly if the *X* and *Y* data are poorly correlated, because the OLS will be unbiased only if all scatter about the OLS is due to errors in *Y*. In the case of the Station ALOHA data, Ricker's [14] recommendation that the OLS slope be divided by the absolute value of the correlation coefficient produced a result that was more realistic but in fact provided a very poor fit to the data (Fig. 1). A much more satisfactory approach is to identify the other factors that control *Y* and incorporate them, insofar as possible, into the regression model (Fig. 5).

Ricker's comment that, "With biological materials very frequently 80% or more of the variability is natural" [14,p. 425] reflects in large part the realization that dependent variables of biological interest are frequently functions of more than one explanatory variable. Under such conditions, if a particular explanatory variable *X* in fact accounts for a small fraction of the natural variability, then an OLS may provide a very biased estimate of the functional relationship between *X* and *Y* if the *X* observations are uncontrolled.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Declarations

Author contribution statement

Edward A. Laws: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Robert R. Bidigare, David M. Karl: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by the National Science Foundation (OCE-1260164) and the Center for Microbial Oceanography: Research and Education (C-MORE; EF-0424599). David M. Karl was supported by the Gordon and Betty Moore Foundation (Marine Microbiology Investigator #3794), and the Simons Foundation (SCOPE). Robert R. Bidigare was supported by the Department of Energy (DE-EE0003371).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associated with this study has been deposited at http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html

References

[1] J.H. Brown, B.A. Maurer, Macroecology: The division of food and space among species on continents, Science 243 (4895) (1989) 1145–1150.

[2] M.J. Sauer, et al., Under the hood of satellite empirical chlorophyll a algorithms: revealing the dependencies of maximum band ratio algorithms on inherent optical properties, Opt. Express 20 (19) (2012) 20920–20933.

[3] C.M. Hu, Z. Lee, B. Franz, Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference, J. Geophys. Res. Oceans (2012) 117.

[4] R.W. Eppley, et al., Estimating ocean primary production from satellite chlorophyll: Introduction to regional differences and statistics for the Southern California Bight, J. Plankton Res. 7 (1) (1985) 57–70.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[5] T. Platt, S. Sathyendranath, Oceanic primary production: Estimation by remote-sensing at local and regional scales, Science 241 (4873) (1988) 1613–1620.

[6] M.J. Behrenfeld, P.G. Falkowski, Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnolo. Oceanogr. 42 (1) (1997) 1–20.

[7] T.K. Westberry, P.J.L. Williams, M.J. Behrenfeld, Global net community production and the putative net heterotrophy of the oligotrophic oceans, Glob. Biogeochem. Cycles 26 (2012) 17.

[8] D.A. Siegel, et al., Global assessment of ocean carbon export by combining satellite observations and food-web models, Glob. Biogeochem. Cycles 28 (3) (2014) 181–196.

[9] E. Steemann Nielsen, Measurement of the production of organic matter in the sea by means of carbon-14, Nature 167 (1951) 684–685.

[10] B.L. Li, et al., Size-dependent photosynthetic variability in the North Pacific Subtropical Gyre, Mar. Ecol. Prog. Ser. 440 (2011) 27–40.

[11] M.E. Ondrusek, et al., A predictive model for estimating rates of primary production in the subtropical North Pacific Ocean, Deep-Sea Research (Part II Topical Studies in Oceanography) 48 (8–9) (2001) 1837–1863.

[12] D.M. Karl, et al., The role of dissolved organic matter release in the productivity of the oligotrophic North Pacific Ocean, Limnol. Oceanogr. 43 (6) (1998) 1270–1286.

[13] R.R. Bidigare, L. Van Heukelem, C.C. Trees, Analysis of algal pigments by high-performance liquid chromtography, In: R.A. Andersen (Ed.), Algal Culturing Techniques, Academic Press, New York, 2005, pp. 327–345.

[14] W.E. Ricker, Linear regressions in fishery research, Journal of the Fisheries Research Board of Canada 30 (1973) 409–434.

[15] Y.T. Prairie, R.H. Peters, D.F. Bird, Natural variability and the estimation of empirical relationships: a reassessment of regression methods, Can. J. Fish. Aquat. Sci. 52 (4) (1995) 788–798.

[16] M.J. Church, M.W. Lomas, F. Muller-Karger, Sea change: Charting the course for biogeochemical ocean time-series research in a new millennium, Deep-Sea Res. II 93 (2013) 2–15.

[17] A. Calbet, Y.T. Prairie, Mesozooplankton grazing and primary production: Reply to the comment by Laws, Limnol. Oceanogr. 48 (3) (2003) 1359–1362.

 http://dx.doi.org/10.1016/j.heliyon.2016.e00156

2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

[18] D.M. Karl, et al., Temporal studies of biogeochemical processes determined from ocean time-series observations during the JGOFS era, In: M.J.R. Fasham (Ed.), Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change, Springer, New York, 2003, pp. 239–267.

[19] P.D. Quay, et al., Measuring primary production rates in the ocean: Enigmatic results between incubation and non-incubation methods at Station ALOHA, Glob. Biogeochem. Cycles 24 (2010) 14.

http://dx.doi.org/10.1016/j.heliyon.2016.e00156
2405-8440/© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).