



Evaluation of fluorometrically-derived chlorophyll *a* as a satellite ocean color validation product using statistical metrics

AIMEE R. NEELEY,^{1,2,*} IVONA CETINIĆ,^{1,3} AND CRYSTAL THOMAS^{1,2}

¹NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

²Science Systems and Applications, Inc., Lanham, Maryland, USA

³GESTAR II, Morgan State University, Baltimore, Maryland, USA

*aimee.neeley@nasa.gov

Abstract: The fluorescence of chlorophyll *a* (Chla) has long been considered a reliable estimate of phytoplankton biomass in aquatic environments. The in vitro fluorometric method for measuring Chla is a simple, fast, and cost-effective way to estimate phytoplankton biomass. High performance liquid chromatography measurements (HPLC) of Chla (HChla) have been used historically for ocean color validation. The applicability of fluorometrically-derived Chla (FChla) to satellite ocean color validation has not been fully evaluated. To this end, we developed statistical metrics to evaluate the uncertainty of fluorometric measurements of Chla by comparing them to coincident measurements made by HPLC. The distribution of uncertainties demonstrates occasional trends in the discrepancies between FChla and HChla, from which we developed a set of recommendations for future collection and validation efforts.

© 2025 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

The Coastal Zone Color Scanner (CZCS) on Nimbus-7 launched in 1978 by the National Aeronautics and Space Administration (NASA) was the first space-borne mission to measure ocean color and derive the concentration of total phytoplankton pigments, i.e., Chla, in the ocean from remote sensing reflectance [1–3]. This proof-of-concept mission proved that such observations of global phytoplankton biomass and productivity were possible [4]. Following CZCS's success, numerous, more advanced ocean color instruments were developed and launched, including Sea-viewing Wide Field-of-view Sensor (SeaWiFS) in 1997, the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard Terra and Aqua in 2000 and 2002, respectively, the Suomi National Polar-orbiting Partnership (SNPP) that incorporated the first Visible-Infrared Imaging Radiometer Suite (VIIRS), the Medium Resolution Imaging Spectrometer (MERIS), Second-generation Global Imager (SGLI), and the Ocean and Land Colour Instrument (OLCI) [5]. Owing to the advancements in ocean color science, in situ measurements of Chla have been invaluable for ground-truthing satellite-derived phytoplankton pigments used as a global proxy for phytoplankton biomass.

The concentration of the photosynthetic pigment Chla has long been considered a reliable estimate of phytoplankton biomass [6–8]. An early accepted method used to measure Chla was by spectrophotometric determination using the trichromatic or monochromatic equation [9]. A fluorometric method was later developed and showed to be a simple, rapid, and inexpensive method that could be applied both in the laboratory and in the field [8,10]. In this “acidification” method (updated in [11]), the fluorescence of the pigment extract is measured before and after the addition of a small volume of acid, which removes the magnesium atom from Chla, converting it to pheophytin *a*; from the two readings the concentrations of Chla and pheopigments can be determined. A method was later developed in which the optical configuration of the fluorometers

included narrower excitation and emission filters than those used for the conventional acidification method, thereby reducing the interference of other fluorescing accessory pigments and eliminating the need for the addition of acid [12]. Chromatographic analysis of pigments has evolved from the early methods of thin-layer chromatography [13] into state-of-the-art HPLC methods. HPLC analysis has revolutionized the way we measure not only Chla but many accessory pigments that can be used to model the phytoplankton community composition in an aquatic environment [14–18]. HPLC methods for measuring phytoplankton pigments are thought to be more accurate than fluorometric methods. Fluorometric methods exhibit sources of error that impact the accuracy of fluorometry attributed to interfering pigments and method and extraction types [19]. As such, HPLC methods for measuring phytoplankton pigments are thought to be more accurate than fluorometric methods.

Numerous community programs and independent studies were designed to facilitate activities that would assess the robustness and uncertainties of in situ Chla measurements for the downstream purpose of calibrating and validating ocean color sensors and algorithms [20]. For instance, a Sensor Intercomparison and Merger for Biological and Interdisciplinary Ocean Studies (SIMBIOS) round robin experiment was conducted in 2002 that compared laboratory methods for measuring Chla. The resulting report suggested that the nuances of the different HPLC and fluorometric methods could lead to the discrepancies between fluorometrically-measured and HPLC-measured Chla, abbreviated here as FChla and HChla, respectively [21]. Other method assessment studies have examined the impacts of sample storage, extraction, and analysis techniques on uncertainties in the measurements of Chla. Long-term freezer (-20° C) storage of samples can cause Chla degradation by as much as 50%, leading to poor analytical precision compared to samples that were extracted and measured immediately after filtration [22]. In a recent study by [23], the intercomparison between eight different Swedish water quality monitoring laboratories that practiced different sample storage, extraction and Chla measurement techniques showed that the variability of the measurements among laboratories was likely caused by differing storage and extraction techniques. These studies and others have been essential to quantifying or sourcing the uncertainties in FChla and HChla [16,24]. Yet, to date, no studies have addressed the agreement between these two measurements of Chla using statistical metrics for both independently collected samples and diluted HPLC extracts.

The quality of ground truth data, quantified using their associated uncertainties, are transferred to models and algorithms during development. HChla has been historically used for the purpose of ocean color satellite validation because its accuracy, precision, and uncertainties have been well established and quantified through intercalibration activities, such as the SeaWiFS Analysis Round-Robin Experiment (SeaHARRE) activities [24]. Our understanding of discrepancies between HChla and FChla must be better described and quantified through uncertainty metrics to determine if FChla can be used interchangeably with HChla for algorithm and model validation. Given that much of the Chla data submitted and distributed by NASA's SeaWiFS Bio-optical Archive and Storage System (SeaBASS) repository is FChla, it is critical to assess the quality of these data for algorithm validation and development, particularly for NASA's recently-launched the Plankton, Aerosol, Cloud and ocean Ecosystem (PACE) mission [25]. Moreover, Chla retrievals from missions more suitable for inland and coastal water quality measurements, such as Sentinel-2 (MSI) [26] and OLCI [5] will also benefit from such an assessment. In this study, we aimed to apply statistical metrics on coincident FChla and HChla measurements from field samples to develop a mechanism to quantify the level of agreement between FChla and HChla. Through this exploratory analysis, we applied the uncertainties from each measurement to compute difference metrics and evaluate the circumstances under which HChla and FChla can be used interchangeably for algorithm development and validation.

2. Methods

2.1. Study areas and sample collection methods

Samples for HChla and FChla, were collected in duplicate or triplicate during six different field campaigns. For this study, the term ‘replicates’ indicates the collection of more than 1 filter from each sampling event. The sample set information is summarized in Table 1. Each field campaign has been labeled with a corresponding letter than will be used throughout this study. Chla data were partitioned into two different categories based on collection site type: ‘Ocean’ and ‘River’ (Fig. 1). Briefly, ‘Ocean’ samples were collected off the Eastern Coast of the United States in February 2013 (Mannino et al. 2013) (A), off the coast of Hawaii in June 2012 (McClain et al. 2012) (B), North Atlantic and Northeastern US coast in July/August 2014 (C) and in the Eastern North Atlantic in May 2021 (Nelson and Siegel 2021) (D). Freshwater and estuarine samples (‘River’ group of samples) were collected from various rivers in Maine [27] (E) and Harpswell Sound in Maine (F) [28]. All water samples were filtered through 25 mm GF/F filters of nominal pore size 0.7 µm, under gentle vacuum filtration (<5 mm Hg). Filters collected for the detection and quantification of pigments were flash frozen in liquid nitrogen and stored at -80° C until analysis.

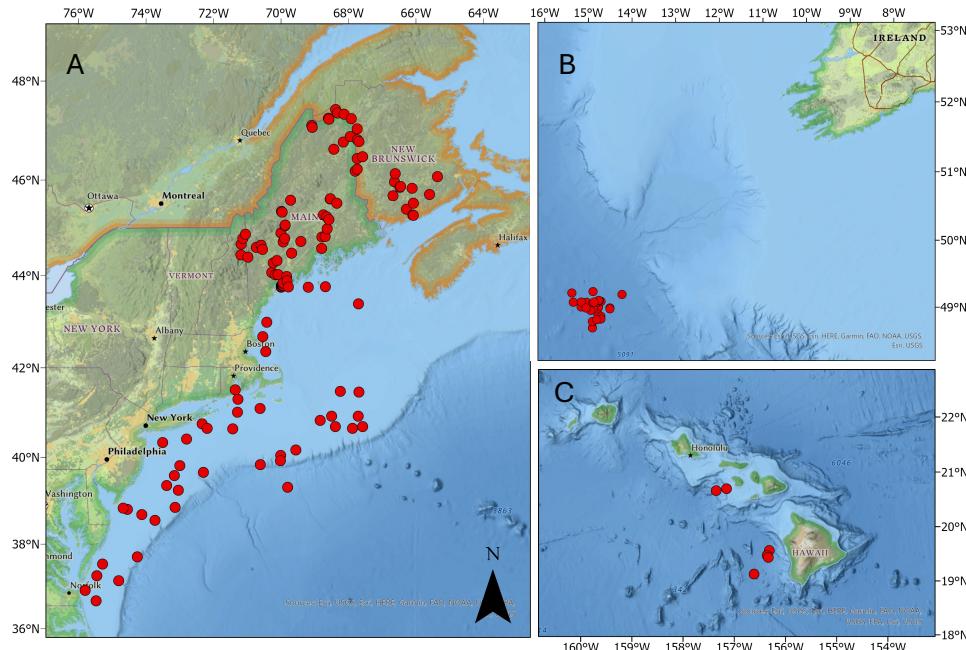


Fig. 1. Maps of sample locations for all data used in this study, with red dots indicating the collection locations. Panel A: Western North Atlantic and Maine inland waters (Sample sets A, C, E, F). Panel B: Eastern North Atlantic (in proximity of station PAP; Sample set D). Panel C: in proximity to Hawai'i (Sample set B).

2.2. HPLC analysis method

Phytoplankton pigments were measured using HPLC following the procedures of [15] and further described in [24]. Briefly, 2.5 mL of 100% acetone (including the internal standard Vitamin E acetate) and 0.15 mL of water were added to the filters (with final acetone concentration of 90%). The internal standard, Vitamin E acetate, was later used for determining extraction volumes. Samples were disrupted using an ultrasonic probe, followed by extraction soaking time of four

Table 1. Sample sets used in this analysis.

Sample set	Location	Dates	Category	doi
A	U. S. East Coast	February 2013	Ocean	10.5067/SeaBASS/ECOMON
B	Coastal Hawaii	June 2012	Ocean	10.5067/SeaBASS/MURI_HI
C	North Atlantic/Northeastern US Coast	July/August 2014	Ocean	N/A
D	Eastern North Atlantic Ocean	June 2021	Ocean	10.5067/SeaBASS/EXPORTS
E	Maine Rivers	2011-2014	River	10.5067/SeaBASS/THREERIVERS
F	Harpowell Sound, Maine	April – July 2011	River	10.5067/SeaBASS/BOWDOINBUOY

hours at -25° C. Samples were clarified through a 0.45 µm pore size syringe filter before analysis. The pigments were analyzed on an Agilent RR1200 HPLC with a 3.5 µm 4.6 × 150 mm Eclipse XDB C8 column, refrigerated autosampler compartment, thermostatted column compartment, quaternary pump with in-line vacuum degasser, and photo-diode array detector with deuterium and tungsten lamps, which collect in-line visible absorbance spectra for each pigment. The mobile phase consisted of a linear gradient from 5% to 95% solvent B over 27 minutes, for which solvent A is 70 parts methanol/30 parts 28 mM tetrabutylammonium acetate (pH 6.5), and solvent B is methanol. The column temperature was maintained at 60° C. Total Chla (Chla in this paper) concentration was calculated as the sum of monovinyl Chla, divinyl Chla, chlorophyllide *a* and Chla allomers and epimers as defined in [24].

2.3. Fluorometric methods

As shown in Table S1 (Supplement 1), fluorometric analyses were performed in three different laboratories. Laboratories 1 and 3 collected and analyzed triplicate filters for fluorometric analysis [8,10]. In addition to fluorometric analysis, Laboratory 2 also conducted HPLC analysis wherein a portion of each HPLC extract was diluted and analyzed using a fluorometer according to EPA method 445.0 [11]. These dilutions were performed using glass gastight syringes (for the extract) and an organic solvent bottle-top dispenser (for the diluent, Dispensette Organic). The laboratories calculated FChla and pheopigment concentrations based on the method described in [10].

2.4. Statistical analyses

The mean and standard deviation (σ) of the replicate filters for each sample were computed. In cases where samples were not collected with replicates, the mean σ ($\bar{\sigma}$) of all values of σ was used to represent the average uncertainty from the sample set. Values of $\bar{\sigma}$ were computed as the root square of the average variance ($\bar{\sigma}^2$) from all the existing replicates in that specific experiment or sample set.

Linear least squares regression analysis was performed using MATLAB's Curve Fitting Toolbox. Bisquare weights were applied during the regression analysis, where the weights of each data point were based on its distance from the line of fit. The Y-intercept was set to zero indicating no Chla is present in the solvent. The linear regression coefficients were determined with 95% confidence bounds. Histogram plots were created using MATLAB's hist function and log-transformed concentrations with 50 equally spaced bins.

For each sample type ('River' or 'Ocean'), Bland-Altman plots were used to compare the agreement between FChla and HChla [29]. The difference (D_i) between HChla and FChla was computed and displayed on the y-axis, while the mean of all HChla and FChla pairs was displayed on the x-axis. The data within sample type groups were not normally distributed as indicated by a one-sided Kolmogorov-Smirnov test performed in MATLAB. As such, the interquartile range

(IQR) was computed using a MATLAB-based Bland-Altman function, which generates both Bland-Altman and correlation scatter plots and provides statistical metrics for agreement [30]. The limits of agreement were computed as +/- 1.45*IQR.

An additional uncertainty visualization method used in this study is described in [31], where a zeta score metric is presented in a modified version of the Bland-Altman plot. Here, the zeta score (ζ_i) is defined as the pairwise differences, D_i , normalized to the uncertainties of both measurements following Eq. (1):

$$\zeta_i = \frac{D_i}{\sqrt{u(\text{HChla}_i)^2 + u(\text{FChla}_i)^2}} \quad (1)$$

where $u(\text{HChla}_i)$ and $u(\text{FChla}_i)$ represent the σ for HChla and FChla, respectively, paired i^{th} values. Pair-wise differences were corrected using a correction factor, CF_i , following Eq. (2):

$$CF_i = 1 - DO_i \quad (2)$$

where DO_i is defined as the degree of overlap computed using the methods described in [31,32]. As DO_i approaches 1, less weight is applied to D_i as statistically there is no difference between the pairs. As DO_i approaches 0, more weight is applied to D_i because the differences are larger. When $DO_i = 0$, the pairs are statistically different. The corrected zeta score, ζ'_i , is then computed following Eq. (3):

$$\zeta'_i = \frac{CF_i D_i}{\sqrt{u(\text{HChla}_i)^2 + u(\text{FChla}_i)^2}} \quad (3)$$

As defined in [31], when comparing the two methods, the following thresholds define the quality of the comparisons for either ζ_i or ζ'_i :

- 1) Values of $|\zeta| \leq 2$ = satisfactory
- 2) Values of $2 \leq |\zeta| \geq 3$ = questionable
- 3) Values of $|\zeta| \geq 3$ = unsatisfactory

The pairwise comparison metrics mean absolute error (MAE) and bias were computed to quantify the average differences and error between the HChla-FChla pairs. MAE and bias were computed following Eq. (4) and (5):

$$\text{bias} = \frac{1}{N} \sum_{i=1}^N D_i \quad (4)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |D_i| \quad (5)$$

The corrected bias and MAE (bias' and MAE') were computed following Eq. (6) and 7:

$$\text{bias}' = \frac{1}{N} \sum_{i=1}^N CF_i D_i \quad (6)$$

$$\text{MAE}' = \frac{1}{N} \sum_{i=1}^N CF_i |D_i| \quad (7)$$

3. Assessment

3.1. Linear regressions and scatter plots

When applied to all data ($N = 1731$), the coefficient of determination ($r^2 = 0.996$) and the regression coefficient (0.983) indicated that almost all variation in HChla could be explained by FChla and a positive, near 1:1 relationship between FChla and HChla (Fig. S1A; Table S2, [Supplement 1](#)). Moreover, the low value of Root Mean Square Error (RMSE) also indicated a good fit (0.191). However, the high value of the Sum of Squares Error (SSE; 62.84) indicated a large unexplained variation between the values of HChla and FChla that was not captured by the regression analysis. The slope, sitting at a value marginally lower than 1, indicated that FChla was slightly overestimated compared to HChla (Table S2, [Supplement 1](#)). The scatter plot of the log-transformed data indicated a lot of scatter along the 1:1 line (Fig. S1B, [Supplement 1](#)), likely owing to the variability within the River data, which constituted a large percentage of the data points ($N = 1575$ or 91%). Histogram plots of the entire data set indicated that the distribution of the HChla observations were very similar to the distribution of FChla, with the FChla values skewing towards higher concentrations (Fig. S2, [Supplement 1](#)). When compared to global distributions in the NASA bio-Optical Marine Algorithm Dataset (NOMAD; [25]), the data set from this study exhibited a narrower distribution centered around Chla concentrations of 1 - 1.5 mg/m³. Generally, the range of Chla concentrations in NOMAD were comparable to this study (Fig. S2, [Supplement 1](#)).

Given the disproportionate contributions of the Ocean and River data to total FChla-HChla pairs, we determined that the results of the statistical analyses could be skewed by the larger data set. Therefore, it was determined that separate analyses to test the River and Ocean data would be more informative. To better understand the variability observed in the entire dataset, we evaluated the relationship between the FChla-HChla pairs within the River and Ocean sample groups (Fig. 2(A), (B), respectively). The River data covered a larger dynamic range in concentrations (~0.05 - 50 mg/m³) when compared to the Ocean dataset (~0.02–7 mg/m³). The scatter around the 1:1 line demonstrated that some of the FChla-HChla pairs within the River group did not agree well. However, no visible bias or trend was observed. The scatter plot for the Ocean data did indicate good agreement between FChla and HChla, where most data points were clustered on or near the 1:1 line, with a few data points indicating an overestimation of HChla compared to FChla at concentrations less than ~0.100 mg/m³. It is important to note that the Ocean data set included an order of magnitude fewer data points ($N = 156$) than the River data ($N = 1575$). Nevertheless, the level of agreement between the Ocean pairs was encouraging.

Linear regression analysis was applied to the River and Ocean FChla-HChla pairs separately (Fig. 3). The coefficients for both groups were determined with 95% confidence bounds. For the River data, the value of the regression coefficient r^2 indicated that almost 100% of the variation in HChla could be explained by the variation in FChla with a positive, near 1:1 relationship between FChla and HChla (Fig. 3(A); Table S2, [Supplement 1](#)). Moreover, the low value of RMSE (0.190) also indicated a good fit. However, the high value of SSE (53.52) indicated large unexplained variation between the values of FChla and HChla that could not be explained by the regression. Similarly, the Ocean data also exhibited a good fit with a low RMSE (0.077), an r^2 close to 1 and near 1:1 relationship between FChla and HChla (Fig. 3(B); Table S2, [Supplement 1](#)). In contrast to the River data, the value of SSE of the Ocean data was much lower, indicating less unexplained variation than the River data. The bias for all of the data combined was low (0.002) with the River pairs indicating no bias and the Ocean pairs indicated a bias of 0.017. MAE was highest in the River data (0.352) and lowest in the Ocean data (0.099). The MAE for the entire data set was 0.329, likely biased by the error in the River data set. Bias' and MAE' showed some improvement compared to the non-corrected values (Table S2, [Supplement 1](#)).

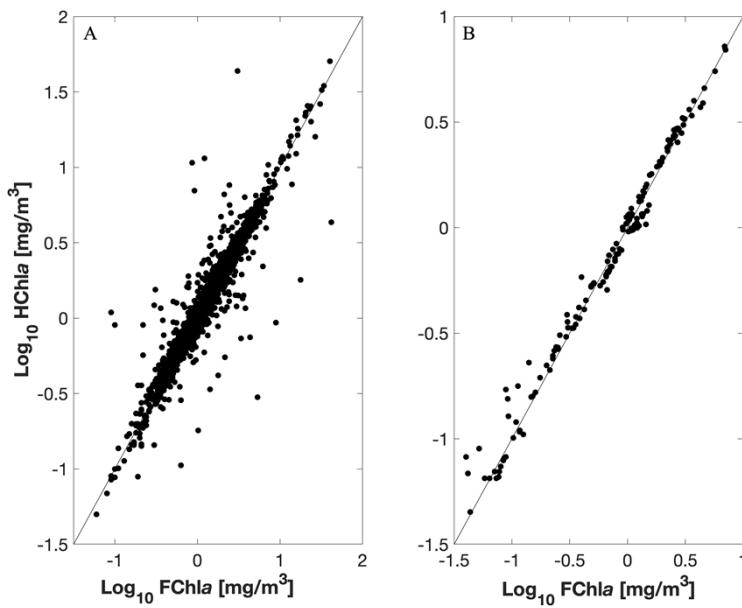


Fig. 2. Scatter plots of log-transformed FChla-HChla pairs for the River (A) and Ocean (B) data sets.

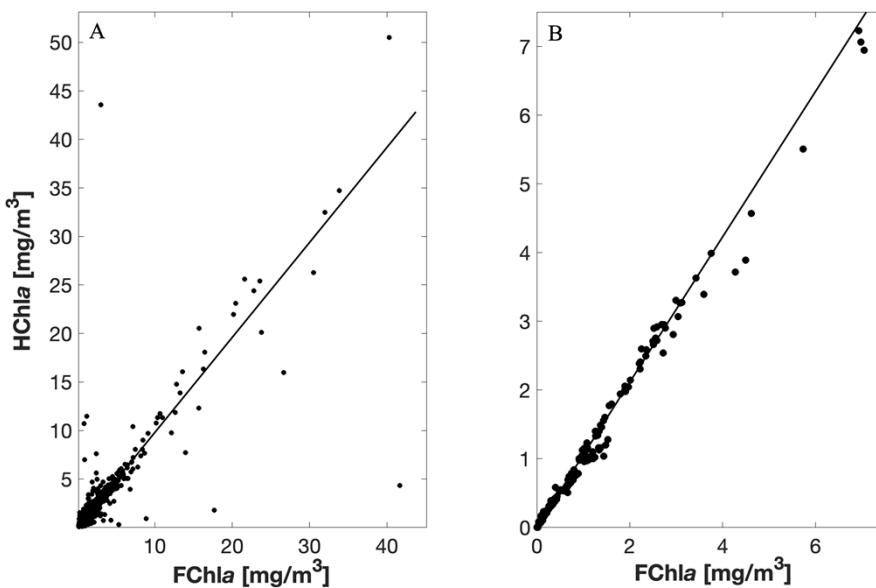


Fig. 3. Linear regression analysis of FChla-HChla pairs for the River (A) and Ocean data (B).

To further investigate possible trends in the FChla and HChla discrepancies, and to understand the threshold at which FChla and HChla (dis)agree, we evaluated the level of (dis)agreement as a function of mean Chla ($\overline{\text{Chla}}$) across the entire dataset by further partitioning the $\overline{\text{Chla}}$ from the Ocean and River data sets into four bins (Table S2, [Supplement 1](#)). Bias ranged between -1.7% and 6.2%, the higher of which was observed in the River $\overline{\text{Chla}} \geq 3 \text{ mg/m}^3$ and lowest in the Ocean $< 1 \text{ mg/m}^3$. MAE ranged between 0.037 and 1.372, the highest value observed in the River $\overline{\text{Chla}} \geq 3 \text{ mg/m}^3$ and the lowest in the Ocean $< 1 \text{ mg/m}^3$. Bias' demonstrated some improvement in most categories except River and Ocean $\overline{\text{Chla}} \geq 3 \text{ mg/m}^3$, and MAE' values were improved in all the $\overline{\text{Chla}}$ categories (Table S2, [Supplement 1](#)).

3.2. Bland Altman and zeta score plots

Bland-Altman analysis was performed to further assess the (dis)agreement between FChla and HChla pairs. The analysis of the River dataset determined a bias of 0 (solid line), indicating no trend in the difference between the two measurements (Fig. 4). The limit of agreement (LOA) was between -0.31 and 0.30 mg/m^3 (dotted lines), indicating that values of D_i beyond this range, 21% of datapoints or 338 FChla/HChla pairs did not have good agreement. The River data also exhibited a high kurtosis (400), and a positive skewness (1.671), likely driven by the outliers or differences between FChla and HChla in excess of $\pm 5 \text{ mg/m}^3$.

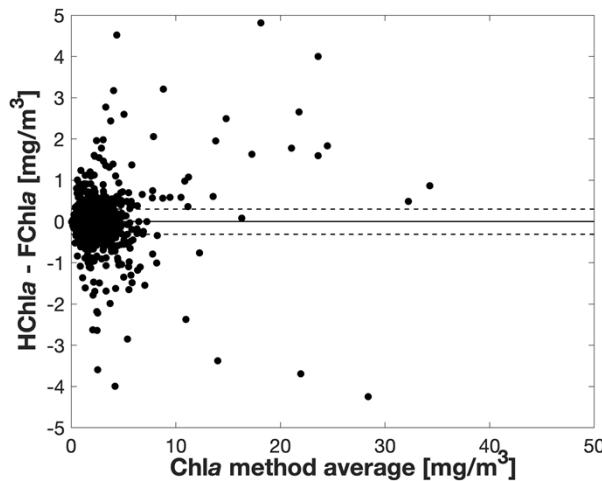


Fig. 4. Bland-Altman plot comparing differences between HChla and FChla varying with the method average values of Chla for the River data (y-axis limit ± 5).

For the Ocean dataset, the LOA was between -0.15 and 0.22 (dotted lines), where nineteen FChla - HChla pairs, or 12% of the data points did not indicate good agreement (Fig. 5). The bias was computed as 0.02 (solid line), indicating a slightly positive bias in the differences between the two measurements. The Ocean data also exhibited smaller value of kurtosis (6.470) and slightly negative value of skewness (-0.866), driven by a few outliers $< -0.02 \text{ mg/m}^3$. Note the gray data points in Fig. 5 that exhibit a negative trend while the rest of the data exhibit a positive trend. The negative trend is likely due to an overestimation of FChla owing to the interference of Chlc during the measurement. This will be discussed in more detail in section 4.1.

Standard zeta scores (ζ) were computed for both the River and Ocean data sets (Table 2). Given the high volume of data points in the River data set, the scores were reduced for Fig. 6 to only include those scores that were improved by applying the correction factor (ζ'). For the River data set, the absolute values ζ or $|\zeta|$ indicated that most of the data pairs, 1147 or 73%, had acceptable (green range, Fig. 6(A)), 135 or 8.5% had questionable (yellow range, Fig. 6(A)), and 293 pairs

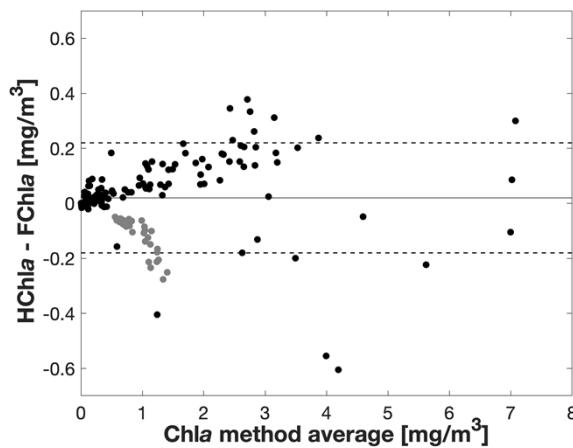


Fig. 5. Bland Altman plot comparing differences between HChla and FChla varying with the method average values of Chla for the Ocean data. Data points in gray belong to sample set D.

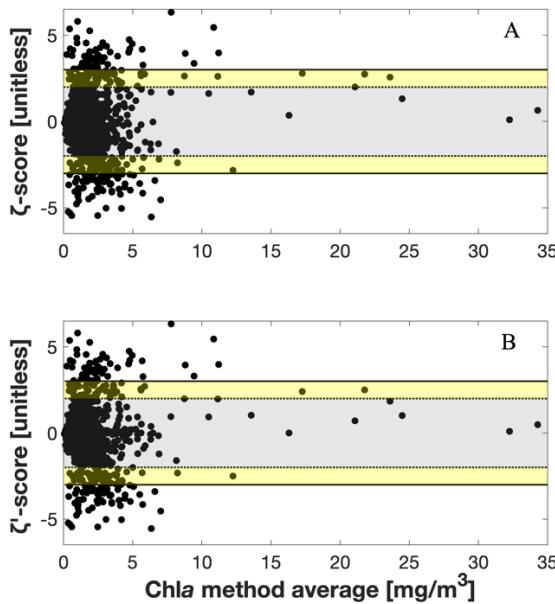


Fig. 6. Zeta scores (A) and corrected zeta scores (B) for the River data plotted against the method average. Gray shading indicates scores $|\zeta| \leq 2$, yellow indicates scores $2 \leq |\zeta| \geq 3$ and, white indicates scores $|\zeta| \geq 3$. Data were reduced to show only improved scores.

(18.6%) had poor agreement (red range, Fig. 6(A); Fig. S3A, [Supplement 1](#)). The corrected zeta scores, $|\zeta'|$, exhibited an improvement over $|\zeta|$, with the acceptable values increasing to ~76% of the data points ($n = 1190$), and questionable and poor data pairs decreasing to ~6.5% ($n = 103$) and ~18% ($n = 282$) respectively (Fig. 6(B); Fig. S3B, [Supplement 1](#)). For the Ocean data, $|\zeta|$ of 95 data pairs (61%) were acceptable, 25 (~16%) were questionable, and 36 (~23%) were poor (Fig. 7(A); Table 2). Although this evaluation indicated that the agreement of the FChla-HChla pairs were largely acceptable, the $|\zeta'|$ values (Fig. 7(B)) increased the number of acceptable data points to 106 (68%) and decreased the number of questionable to 14 (9%), and the number of

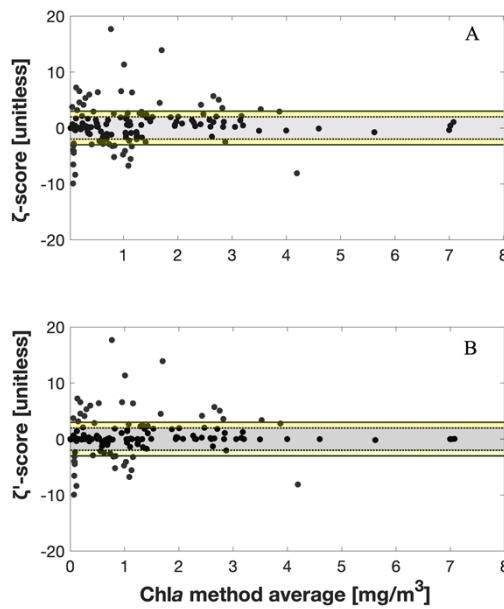


Fig. 7. Zeta scores (A) and corrected zeta scores (B) for the Ocean data comparing FChla and HChla varying with the method average values of Chla. Gray shading indicates scores $|\zeta| \leq 2$, yellow indicates scores $2 \leq |\zeta| \leq 3$, and white indicates scores $|\zeta| \geq 3$.

poor data points remained the same. To further distinguish water types, data were re-partitioned into five Chla bins. The highest percentage of acceptable values of $|\zeta'|$ were observed in the $0.3 \geq \overline{\text{Chla}} < 1$ category (90%). The percentage of acceptable values declined over last two categories ($1 \geq \overline{\text{Chla}} < 3$ and $\overline{\text{Chla}} \geq 3$) to 76% and 46%, respectively. Overall, the percentage of acceptable $|\zeta'|$ values were slightly improved over the $|\zeta|$ values by 1–9% in the partitioned Chla categories.

Table 2. Zeta score statistics and tallies for HChla and FChla.

Range	N	Mean ζ (sd)	Mean ζ' (sd)	$ \zeta < 2$	$2 \geq \zeta < 3$	$ \zeta \geq 3$	$ \zeta' < 2$	$2 \geq \zeta' < 3$	$ \zeta' \geq 3$
All	1731	0.543 (15.168)	0.560 (15.154)	1242	160	329	1296	117	318
River	1575	0.558 (15.865)	0.576 (15.853)	1147	135	293	1190	103	282
Ocean	156	0.393 (3.437)	0.390 (3.333)	95	25	36	106	14	36
$\overline{\text{Chla}} < 0.1$	24	-1.900 (5.396)	-1.043 (2.408)	16	1	7	17	1	6
$0.1 \geq \overline{\text{Chla}} < 0.3$	83	1.301 (8.488)	0.467 (1.986)	73	1	9	74	1	8
$0.3 \geq \overline{\text{Chla}} < 1$	616	0.760 (11.525)	0.071 (2.101)	529	29	58	557	19	40
$1 \geq \overline{\text{Chla}} < 3$	762	-0.245 (7.006)	-0.253 (6.805)	531	80	151	576	53	133
$\overline{\text{Chla}} \geq$	246	3.098 (35.480)	3.106 (35.470)	95	48	103	112	36	98

Box-density trace plots, also known as violin plots, were used in this study to graphically represent the distribution of the uncertainties, σ in this study, in the replicates for each sample set [33,34]. For the violin plots, the shaded circles represent the individual values of σ from the replicates for FChla (Fig. 8, Panel A) and HChla (Fig. 8, Panel B) with the kernel density estimation overlaid on the data plots. The shape of the uncertainty distribution, with the bulk of the data points creating a bulge at the bottom of the violin, indicates that the values of uncertainty largely fall in the lower end with a limited number of outliers driving the mean. These results give us confidence that generally, with some exceptions, values of FChla and HChla relatively agree with each other.

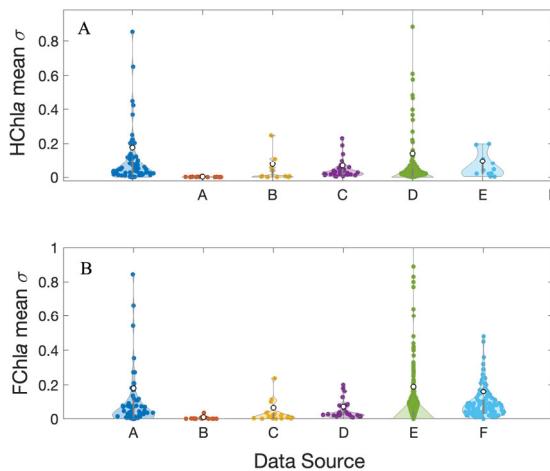


Fig. 8. Violin plots of σ for FChla (Panel A) and HChla (Panel B). Groups are labeled based on Table 1. Note the cutoff on the y axis for the E dataset (max values for σ HChla = 1.357 and FChla = 5.040). The white circle represents the $\bar{\sigma}$ used in this analysis. Each sample set is represented separately on the x-axis. The gray bar indicates the limits of the first and third quartiles, with outside points located at the top of the plot.

4. Discussion

4.1. Sources of discrepancies: analytical error

The discrepancies between FChla and HChla arise from both systematic and sampling uncertainties, as well as from the interference of other pigments (usually associated with limitation of the individual methods). In this study, HPLC extracts from a subset of the samples were run coincidentally on a fluorometer, thereby removing uncertainties associated with sampling and handling, allowing us to focus on the systematic/analytical errors only.

During the SIMBIOS Chla intercalibration exercise [21], the discrepancy between HChla and FChla for HPLC extracts fell between -14.8% and 6.41%, while, in this study the discrepancies for the ocean samples (FChla measured from HPLC extracts; $n = 134$) fell between: -18.9 to 18.1%. While this is a larger range than observed during the SIMBIOS intercalibration, most of the values from this study fall within their reported range. In our dataset, pair differences were distributed between negative and positive values, indicating that depending on the source or sample, HChla can be greater than or less than FChla [21].

The differences between FChla and HChla in the River data did not exhibit a notable trend (Fig. 5). On the other hand, the differences between FChla and HChla in the Ocean samples generally exhibited a positive trend with increasing concentration except for subset of samples (HPLC extracts belonging to collection D), which exhibited a negative trend. Interestingly, Van

Heukelem et al. (2002) noted that negative discrepancies observed in the SIMBIOS dataset were associated with higher Chlc:HChla (n.b. they could not relate the magnitude of the discrepancies to the presence of the accessory pigments).

The presence of Chlb, pheopigments, and Chlc at high levels relative to Chla can cause an underestimation or an overestimation, respectively, of FChla compared to HChla [12,19,35–38]. The excitation and emission filters of the standard fluorometric method are broad, 340–500 nm and > 665 nm, respectively, allowing for the fluorescence response of Chlb and Pheophytin b to be measured (the maximum excitation wavelength for Chlb is 470 nm). According to [36], when samples are acidified, maximum excitation wavelength for Chlb shifts to 440 nm, leading to higher emission of pheophytin b in the red end of the spectrum. When FChla is computed, the larger value of acidified fluorescence is subtracted from the initial fluorescence, leading to an underestimation of Chla [36]. The average error in the fluorometric method from samples collected in the ocean has been reported as 39% and ranged from -68 to 53% largely owing to the interference of accessory pigments Chlb and Chlc [19]. Moreover, other studies have reported an overestimation of FChla by ≤ 10% and an underestimation of pheophytin a when the ratio of Chlc:HChla was close to 1:1 [10,36]. The interference of Chlc largely occurs at high concentrations of diatoms and chrysophytes [8]. In our sample set D, Chlc:HChla ranged from 30 - 50%, higher than the other ocean samples where Chlc:HChla ranged from 0–28%. Chlb:HChla was generally less than 4%. In contrast, the range of Chlb:HChla values were greater in sample sets A-C (0–0.731) compared to sample set D (0.0138–0.593). Furthermore, sample set D was collected during a senescent diatom bloom in the North Atlantic that was likely caused by silicate limitation [39], similar to previous North Atlantic Blooms [35,40]. This evidence supports the hypothesis that the presence of Chlc is correlated with observed negative discrepancies in the sample set D, i.e., an overestimation of FChla, supporting the previous studies and observations discussed here.

4.2. Sources of discrepancies: heterogeneity and sampling - handling error

During this study, we have observed evidence that, in addition to systematic errors such as those observed during the SIMBIOS experiment, additional sources of discrepancy between FChla and HChla likely originate from patchiness (i.e., heterogeneity) in the distribution of phytoplankton cells as well as sampling and handling techniques. Microscale and fine scale patchiness of small phytoplankton occurs in the coastal and oligotrophic oceans owing to environmental conditions such as a stable water column and low wind shear and the concentration of nutrients and particles in coastal oceans [41]. Populations of the diazotroph *Trichodesmium*, which can aggregate into ‘tufts’ and ‘puffs’ are known for their heterogeneous distribution in the oligotrophic ocean, including the south and northern Pacific Ocean [40–44]. Sampling in these patchy, more phytoplankton-concentrated areas can lead to higher uncertainties in Chla measurements, whether HChla or FChla, particularly if replicate filters are not collected.

Diatom blooms in the North Atlantic are known for their patchiness owing to eddies that keep phytoplankton cells and nutrients near the ocean surface [45]. Moreover, eddies can create multiple, separate patches or blooms of diatoms that can exist simultaneously over large spatial scales [45–47]. Such largescale blooms typically include large chain diatoms, chlorophyll-containing fecal pellets, aggregates, and zooplankton that can also cause heterogeneity in replicate sample filters. For example, large chain forming diatoms, or a zooplankton is filtered onto one replicate filter but not the others. Additionally, larger diatoms sink faster than other phytoplankton cells. Such differential particle settling rates in Niskin bottles or sampling carboys can lead to bias in sampling, ultimately resulting in concentration discrepancies within same water mass (e.g., for particulate carbon as noted in [48]).

Although all samples used in this study were handled in a similar manner, variations in sample handling – collection, storage, and extraction – have the potential to be sources of error with

pigment analysis. Cell breakage and leakage of Chla and other pigments can occur during the process of filtration either by high particle loads on the filters, or if too much pressure or vacuum is applied to the sample [48]. Another source of sampling error is related to the type of pump used in underway systems on research vessels. The use of impeller pumps can lead to cell breakage and a decrease in plankton cell counts by 40% - 78% when compared to the use of a diaphragm pump [49]. Decomposition of the Chla occurs resulting in elevated concentration of pheophorbide *a* and chlorophyllide *a* (activity of chlorophyllase enzyme particularly in diatoms), both of which are detected using HPLC pigment analysis [49,50]. Chlorophyllide *a* is also observed coincident with diatom bloom senescence.

Storage and extraction procedures of sample filters collected for Chla measurements are also sources of error. [23] observed a ~51% decrease of Chla in samples that were frozen at -20° C. Other studies have also shown losses of Chla up to 28% following long-term storage at -20° C and short-term storage (5 days) at 4° C indicating that a majority of the loss occurs within the first two weeks of storage [8,23,51]. In contrast, pheopigments did not appear to be affected by the varying storage procedures [23,51]. Additionally, it was observed that the magnitude of the Chla loss was biased towards phytoplankton size distributions > 20 µm [23]. Flash-freezing and storage in liquid nitrogen or at -80° C are considered preferable storage options over -20° C and room temperature [23,51]. Methanol and dimethylformide (DMF) have the potential for higher extraction efficiency depending on the species present (Wright et al. 1997), but acetone remains the most common extraction solvent [52]. DMF is not often used due to its toxicity. Chlorophylls extracted in methanol quickly become allomerize [53] and can degrade at rates as high as 60% per day [24,53]. Samples extracted in acetone are more stable [24,53,54], and can degrade at rates as little as -0.2% per day [24].

5. Comments and recommendations

Accurate measurements of in situ Chla from aquatic environments are vital for calibration and validation efforts of space-borne ocean color missions. HChla measurements, over the years, have been a preferred over the FChla, due to the more precise methodology to yield in situ Chla. However, the HPLC method is not available to all laboratories, largely owing to the requirement of highly trained personnel and professional laboratory setup, making it cost prohibitive. FChla, is a simpler, faster, and more accessible (less costly) method, although it is considered less accurate. In this paper we demonstrated that in a majority of both oceanic and coastal environments with Chla concentrations $\leq 3 \text{ mg/m}^3$ Chla, these two methods can be used interchangeably. However, in other environments, such as in coastal areas with higher productivity and Chla concentrations (e.g., the North Atlantic bloom), uncertainties will likely be higher for FChla owing to the greater presence of interfering compounds (Chlb and Chlc) that do not impact HChla. The large ‘unexplained’ variation in the River data discussed in section 3.1 may actually be explained by the ‘patchiness’ in filter replicates along with high concentrations of interfering compounds (e.g., Chlb and Chlc) observed in coastal and inland water types, leading to the large variation between HChla and FChla.

This analysis was possible because replicates were collected as part of all these studies. Replication here was used to understand better the population dynamics, and estimate (when possible) reasons for discrepancies between the methods. In a similar manner, these replicates can be used to estimate a point in which heterogeneity in the water is too high, and to inform researchers when to use HChla over FChla. With that in mind, our recommendation is to routinely collect replicates. A subset of this samples with replicates can be measured using both fluorometry and HPLC. If the FChla values are relatively comparable to HChla, then the full set would not need to be analyzed by HPLC, thereby reducing analysis costs. The statistical metrics discussed in this study can also be used in data exercises (algorithm developments, validation,

and models) where replicates can help to determine which samples and Chla measurements should be used.

The derived Chla product from ocean color observations is one of the most used, space-based observations in science. It is an Essential Ocean Variable (EOV), and in most cases its distribution follows the FAIR principles of Findability, Accessibility, Interoperability, and Reusability [55–57]. In situ Chla measurements collected as a part of NASA-funded studies must be stored in the bio-optical archive SeaBASS, a long-term, publicly accessible repository in order to provide best possible data for validation and to develop new algorithms, e.g., NOMAD [25]. However, as we are attempting to make science, especially space-based science, open to all, the development of geographically local remote sensing algorithms for Chla in areas where HPLC-based methods are not available (or cost prohibitive), is challenging. These results demonstrate that fluorometric-based measurements are as valuable and as representable as HPLC. They can support the development, refinement, and validation of the current and future Chla algorithms from space, making data more accessible for all. In summary, the collection of replicate filters in patchy environments is important to capture particle variability. In highly productive or bloom environments, interfering compounds can lead to over or underestimation of FChla compared to HChla and FChla may be used with caution. Collection and storage procedures described in this, and other studies must be followed to ensure that the best and most consistent techniques are implemented for optimal sample preservation.

Funding. National Aeronautics and Space Administration (Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) Mission).

Acknowledgment. This project was funded by the Ocean Biology and Biogeochemistry Program at NASA. We acknowledge the PIs who worked hard to collect, analyze, and submit Chlorophyll *a* data sets to NASA's SeaBASS repository. We would also like to thank Lachlan McKenna for his expertise and providing code for the zeta score computations.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Table 1 of this paper and in the SeaBASS repository [25].

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. H. R. Gordon, D. K. Clark, J. L. Mueller, *et al.*, "Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: comparisons with surface measurements," *Science* **210**(4465), 63–66 (1980).
2. W. A. Hovis, D. K. Clark, F. Anderson, *et al.*, "Nimbus-7 Coastal Zone Color Scanner: system description and initial imagery," *Science* **210**(4465), 60–63 (1980).
3. I. S. Robinson, "Satellite observations of ocean colour," *Philos. Trans. Royal Soc. A* **309**, 415–432 (1983).
4. C. R. McClain, "Satellite Remote Sensing: Ocean Color," in *Encyclopedia of Ocean Sciences (Second Edition)*, J. H. Steele, ed. (Academic Press, 2009), pp. 114–126.
5. C. R. McClain, B. A. Franz, and P. J. Werdell, "Genesis and evolution of NASA's satellite ocean color program," *Front. Remote Sens.* **3**, 938006 (2022).
6. J. Krey, "Chemical methods of estimating standing crop of phytoplankton," *Cons. Perm. Int. Expl. Mer. Rapp. Proc.-Verb. Reun* **144**, 20–27 (1958).
7. C. J. Lorenzen, *A method for the continuous measurement of in vivo chlorophyll concentration*, in (Elsevier, 1966), Vol. 13, pp. 223–227.
8. C. S. Yentsch and D. W. Menzel, "A method for the determination of phytoplankton chlorophyll and pheophytin by fluorescence," *Deep-Sea Res. Oceanogr. Abstr.* **10**(3), 221–231 (1963).
9. S. t. Jeffrey and G. Humphrey, "New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton," *Biochemie und physiologie der pflanzen* **167**(2), 191–194 (1975).
10. O. Holm-Hansen, C. J. Lorenzen, R. W. Holmes, *et al.*, "Fluorometric determination of chlorophyll," *ICES J. Mar. Sci.* **30**(1), 3–15 (1965).
11. E. J. Arar and G. B. Collins, *Method 445.0: In Vitro Determination of Chlorophyll a and Pheophytin a in Marine and Freshwater Algae by Fluorescence* (1997).
12. N. A. Welschmeyer, "Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments," *Limnol. Oceanogr.* **39**(8), 1985–1992 (1994).
13. S. W. Jeffrey, "An improved thin-layer chromatographic technique for marine phytoplankton pigments," *Limnol. Oceanogr.* **26**(1), 191–197 (1981).

14. S. Roy, "High-performance liquid chromatographic analysis of chloropigments," *J. Chromatogr. A* **391**, 19–34 (1987).
15. L. Van Heukelom and C. S. Thomas, "Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments," *J. Chromatogr. A* **910**(1), 31–49 (2001).
16. L. Van Heukelom, C. S. Thomas, and P. M. Gilbert, *Sources of Variability in Chlorophyll Analysis by Fluorometry and High-Performance Liquid Chromatography in a SIMBIOS Inter-Calibration Exercise* (2002), p. 52.
17. S. W. Wright, S. W. Jeffrey, R. F. C. Mantoura, *et al.*, "Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton," *Mar. Ecol.: Prog. Ser.* **77**, 183–196 (1991).
18. M. Zapata, F. Rodríguez, and J. L. Garrido, "Separation of chlorophylls and carotenoids from marine phytoplankton: a new HPLC method using a reversed phase C8 column and pyridine-containing mobile phases," *Mar. Ecol.: Prog. Ser.* **195**, 29–45 (2000).
19. C. C. Trees, M. C. Kennicutt II, and J. M. Brooks, "Errors associated with the standard fluorimetric determination of chlorophylls and phaeopigments," *Mar. Chem.* **17**(1), 1–12 (1985).
20. C. R. McClain, *SIMBIOS Project 1999 Annual Report* (National Aeronautics and Space Administration, Goddard Space Flight Center, 1999), Vol. 209486.
21. L. Van Heukelom, *Sources of Variability in Chlorophyll Analysis by Fluorometry and High-Performance Liquid Chromatography in a SIMBIOS Inter-Calibration Exercise* (National Aeronautics and Space Administration, Goddard Space Flight Center, 2002), Vol. 211606.
22. J. R. Graff and T. A. Ryneanson, "Extraction method influences the recovery of phytoplankton pigments from natural assemblages," *Limnol. Oceanogr. Meth.* **9**(4), 129–139 (2011).
23. S. Kratzer, E. T. Harvey, and E. Canuti, "International Intercomparison of In Situ Chlorophyll-a Measurements for Data Quality Assurance of the Swedish Monitoring Program," *Front. Remote Sens.* **3**, 866712 (2022).
24. S. B. Hooker, L. Van Heukelom, C. S. Thomas, *et al.*, "The second SeaWiFS HPLC analysis round-robin experiment (SeaHARRE-2)," NASA Tech. Memo **212785**, 124 (2005).
25. P. J. Werdell and S. W. Bailey, "An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation," *Remote Sens. Environ.* **98**(1), 122–140 (2005).
26. A. Ansper and K. Alikas, "Retrieval of Chlorophyll a from Sentinel-2 MSI Data for the European Union Water Framework Directive Reporting Purposes," *Remote Sens.* **11**(1), 64 (2019).
27. C. Roesler, J. Lichter, and P. Camil, "NASA 3Rivers. SeaWiFS Bio-optical Archive and Storage System (SeaBASS)," NASA. Accessed: July 2021. <http://dx.doi.org/10.5067/SeaBASS/THREERIVERS/DATA001.>, (2011).
28. C. Roesler, "Bowdoin Buoy. SeaWiFS Bio-optical Archive and Storage System (SeaBASS)," NASA. Accessed: July 2021. <http://dx.doi.org/10.5067/SeaBASS/BOWDOINBUOY/DATA001.>, (2011).
29. M. J. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet* **327**, 307–310 (1986).
30. R. Klein, "Bland-Altman and Correlation Plot," (2019).
31. L. I. McKenna, I. Cetinić, and P. J. Werdell, "Development and validation of an empirical ocean color algorithm with uncertainties: A case study with the particulate backscattering coefficient," *J. Geophys. Res. Oceans* **126**(5), e2021JC017231 (2021).
32. R. D. Harmel, P. K. Smith, and K. W. Migliaccio, "Modifying Goodness-of-Fit Indicators to Incorporate Both Measurement and Model Uncertainty in Model Calibration and Validation," *Trans. ASABE* **53**, 55–63 (2010).
33. B. Bechtold, "Violin Plots for Matlab, Github Project,".
34. J. L. Hintze and R. D. Nelson, "Violin plots: a box plot-density trace synergism," *Am. Stat.* **52**(2), 181–184 (1998).
35. I. Cetinić, M. J. Perry, N. T. Briggs, *et al.*, "Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment," *J. Geophys. Res.* **117**(C6), C06028 (2012).
36. C. J. Lorenzen, "Chlorophyll b in the eastern North Pacific Ocean," *Deep-Sea Res., Part A* **28**(9), 1049–1056 (1981).
37. M. Vernet and C. J. Lorenzen, "The presence of chlorophyll b and the estimation of phaeopigments in marine phytoplankton," *J. Plankton Res.* **9**(2), 255–265 (1987).
38. M. E. Loftus and J. H. Carpenter, "A fluorometric method for determining chlorophylls a, b, and c," *J. Mar. Res.* **29**, 319–338 (1971).
39. L. Johnson, D. A. Siegel, A. F. Thompson, *et al.*, "Assessment of oceanographic conditions during the North Atlantic EXport processes in the ocean from RemoTe sensing (EXPORTS) field campaign," *Prog. Oceanogr.* **220**, 103170 (2024).
40. M. E. Sieracki, P. G. Verity, and D. K. Stoecker, "Plankton community response to sequential silicate and nitrate depletion during the 1989 North Atlantic spring bloom," *Deep Sea Res., Part II* **40**(1-2), 213–225 (1993).
41. R. W. Owen, "Microscale and finescale variations of small plankton in coastal and pelagic environments," *J. Mar. Res.* **47**(1), 197–240 (1989).
42. M. Benavides, L. Conradt, S. Bonnet, *et al.*, "Fine-scale sampling unveils diazotroph patchiness in the South Pacific Ocean," *ISME Commun.* **1**(1), 3 (2021).
43. U. Pfreundt, J. Slomka, G. Schneider, *et al.*, "Controlled motility in the cyanobacterium *Trichodesmium* regulates aggregate architecture," *Science* **380**(6647), 830–835 (2023).
44. Y. Wang, X. He, J. Huang, *et al.*, "The population of *Trichodesmium* cyanobacteria in the northern Pacific Ocean associated with Kuroshio Extension," *Aquat. Sci.* **84**(3), 44 (2022).

45. G. Chen and T. A. Rynearson, "Genetically distinct populations of a diatom co-exist during the North Atlantic spring bloom," *Limnol. Oceanogr.* **61**(6), 2165–2179 (2016).
46. I. Cetinić, M. Perry, E. D'Asaro, *et al.*, "Optical community index to assess spatial patchiness during the 2008 North Atlantic Bloom," *Biogeosciences Discussions* **11**, 12833–12870 (2014).
47. A. Mahadevan, E. D'Asaro, C. Lee, *et al.*, "Eddy-Driven Stratification Initiates North Atlantic Spring Phytoplankton Blooms," *science* **337**(6090), 54–58 (2012).
48. J. E. Chaves, I. Cetinić, G. Dall'Olmo, *et al.*, "Particulate organic carbon sampling and measurement protocols: consensus towards future ocean color missions," in *IOCCG Ocean Optics & Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation* (2021), Vol. 6.0.
49. I. Cetinić, N. Poulton, and W. H. Slade, "Characterizing the phytoplankton soup: pump and plumbing effects on the particle assemblage in underway optical seawater systems," *Opt. Express* **24**(18), 20703–20715 (2016).
50. R. Suzuki and Y. Fujita, "Chlorophyll decomposition in Skeletonema costatum: a problem in chlorophyll determination of water samples," *Mar. Ecol.: Prog. Ser.* **28**, 81–85 (1986).
51. R. F. C. Mantoura, S. W. Wright, S. W. Jeffrey, *et al.*, "Filtration and storage of pigments from microalgae," (1997).
52. J. L. Pinckney, D. F. Millie, L. V. Heukelem, *et al.*, eds., *Cambridge Environmental Chemistry Series* (Cambridge University Press, 2011), pp. 627–635.
53. M. A. van Leeuwe, L. A. Villerius, J. Roggeveld, *et al.*, "An optimized method for automated analysis of algal pigments by HPLC," *Mar. Chem.* **102**(3-4), 267–275 (2006).
54. M. Latasa, K. van Lenning, J. L. Garrido, *et al.*, "Losses of chlorophylls and carotenoids in aqueous acetone and methanol extracts prepared for RP-HPLC analysis of pigments," *Chromatographia* **53**(7-8), 385–391 (2001).
55. F. E. Muller-Karger, P. Miloslavich, N. J. Bax, *et al.*, "Advancing marine biological observations and data requirements of the complementary essential ocean variables (EOVs) and essential biodiversity variables (EBVs) frameworks," *Front. Mar. Sci.* **5**, 211 (2018).
56. F. E. Muller-Karger, E. Hestir, C. Ade, *et al.*, "Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems," *Ecol. Appl.* **28**(3), 749–760 (2018).
57. M. D. Wilkinson, M. Dumontier, Ij. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data* **3**(1), 160018 (2016).