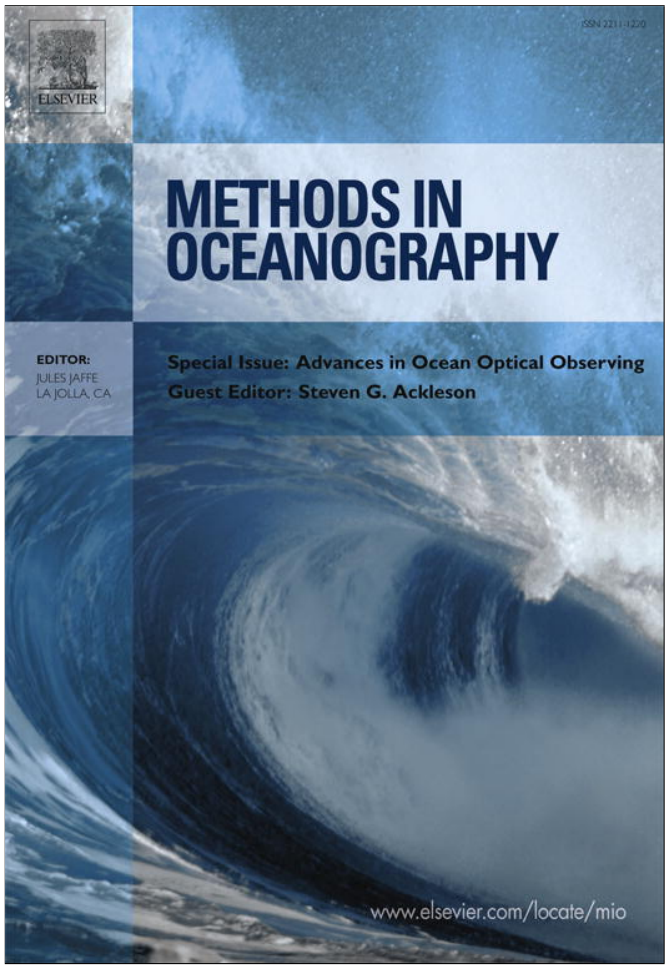
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Methods in Oceanography 7 (2013) 40–51

Contents lists available at ScienceDirect

Methods in Oceanography

journal homepage: www.elsevier.com/locate/mio

Full length article

Underway sampling of marine inherent optical properties on the Tara Oceans expedition as a novel resource for ocean color satellite data product validation☆

P. Jeremy Werdell[a,*], Christopher W. Proctor[a,b], Emmanuel Boss[c], Thomas Leeuw[c], Mustapha Ouhssain[d,e]

[a] NASA Goddard Space Flight Center, Mail Code 616.2, Greenbelt, MD 20771, USA

[b] Science Systems and Applications, Inc., 10210 Greenbelt Road, Suite 600, Lanham, MD 20706, USA

[c] School of Marine Sciences, University of Maine, 458 Aubert Hall, Orono, ME 04401, USA

[d] UMS 829 - Observatoire Océanologique - Centre National de la Recherche Scientifique, Villefranche-sur-Mer, France

[e] Universite Pierre et Marie Curie, Paris O6, Paris, France

ARTICLE INFO

Article history:
Available online 11 December 2013

Keywords:
Ocean color
Bio-optics
Remote sensing
Particle absorption

ABSTRACT

Developing and validating data records from operational ocean color satellite instruments requires substantial volumes of high quality *in situ* data. In the absence of broad, institutionally supported field programs, organizations such as the NASA Ocean Biology Processing Group seek opportunistic datasets for use in their operational satellite calibration and validation activities. The publicly available, global biogeochemical dataset collected as part of the two and a half year Tara Oceans expedition provides one such opportunity. We showed how the inline measurements of hyperspectral absorption and attenuation coefficients collected onboard the R/V Tara can be used to evaluate near-surface estimates of chlorophyll-a, spectral particulate backscattering coefficients, particulate organic carbon, and particle size classes derived from the NASA Moderate Resolution Imaging Spectroradiometer onboard Aqua (MODISA). The predominant strength of such flow-through

☆ This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author. Tel.: +1 301 286 1440.
E-mail address: jeremy.werdell@nasa.gov (P.J. Werdell).

2211-1220/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.mio.2013.09.001

Author's personal copy

measurements is their sampling rate—the 375 days of measurements resulted in 165 viable MODISA-to-*in situ* match-ups, compared to 13 from discrete water sampling. While the need to apply bio-optical models to estimate biogeochemical quantities of interest from spectroscopy remains a weakness, we demonstrated how discrete samples can be used in combination with flow-through measurements to create data records of sufficient quality to conduct first order evaluations of satellite-derived data products. Given an emerging agency desire to rapidly evaluate new satellite missions, our results have significant implications on how calibration and validation teams for these missions will be constructed.

Published by Elsevier B.V.

1. Introduction

Satellite ocean color instruments provide consistent and high-density data on temporal and spatial scales that far exceed current field and aircraft sampling strategies, often with time-series of sufficient length to allow retrospective analysis of long-term trends. For example, the daily, synoptic images captured by the NASA Sea-viewing Wide Field-of-view Sensor (SeaWiFS; 1997–2010) and Moderate Resolution Imaging Spectroradiometer onboard Aqua (MODISA; 2002–present) provide viable data records for observing decadal changes in biogeochemistry of both global and regional ecosystems (McClain, 2009). Briefly, satellite ocean color instruments measure the spectral radiance emanating from the top of the atmosphere at discrete visible and infrared wavelengths. Atmospheric correction algorithms are applied to remove the contribution of the atmosphere from the total signal and produce estimates of remote sensing reflectances ($R_{rs}(\lambda)$; sr^{-1}), the light exiting the water mass normalized to a hypothetical condition of an overhead Sun and no atmosphere (Gordon and Wang, 1994). Bio-optical algorithms are applied to the $R_{rs}(\lambda)$ to produce estimates of additional geophysical properties, such as the near-surface concentration of the phytoplankton pigment chlorophyll-a (C_a; $\mathrm{mg\,m^{-3}}$) and spectral marine inherent optical properties (IOPs), namely the absorption and scattering properties of seawater and its particulate and dissolved constituents (O'Reilly et al., 1998; Werdell et al., 2013). Time-series of these geophysical properties provide unparalleled resources for studying carbon stocks, phytoplankton population diversity and succession, and ecosystem responses to climatic disturbances on regional to global scales (e.g., Siegel et al., 2013).

Refining bio-optical algorithms and verifying ocean color satellite data products requires a substantial volume of *in situ* data to ensure their validity on global spatial and temporal scales (Werdell and Bailey, 2005; Bailey and Werdell, 2006). Previously, large volumes of high quality data were most successfully acquired via institutionally supported programs, such as the NASA Sensor Intercomparison and Merger for Biological and Interdisciplinary Oceanic Studies (SIMBIOS) activity (Fargion and McClain, 2003). During its six-year tenure, SIMBIOS enabled the assembly of 67,000 measurements from 1100 unique field campaigns collected by an assortment of 62 international researchers for inclusion in the NASA SeaWiFS Bio-optical Archive and Storage System (SeaBASS), the permanent archive for *in situ* data obtained under the auspices of the NASA Ocean Biology and Biogeochemistry Program (Werdell et al., 2003). While extremely useful, these data remain heterogeneously distributed in time and space—emphasizing seasonal biases (Spring-Fall) and the coastal and North Atlantic oceans. In the absence of a coordinated activity, organizations responsible for operational ocean color satellite missions, such as the NASA Ocean Biology Processing Group (OBPG; oceancolor.gsfc.nasa.gov) opportunistically seek *in situ* data records to support their algorithm development and satellite data product validation activities.

The Tara Oceans expedition (September 2009 to March 2012) provides one novel opportunity. Briefly, Tara Expeditions (oceans.taraexpeditions.org ; Karsenti et al., 2011) conducted a ~91,000 km voyage on the R/V Tara over two and a half years to capture a view of the global distribution of

Author's personal copy

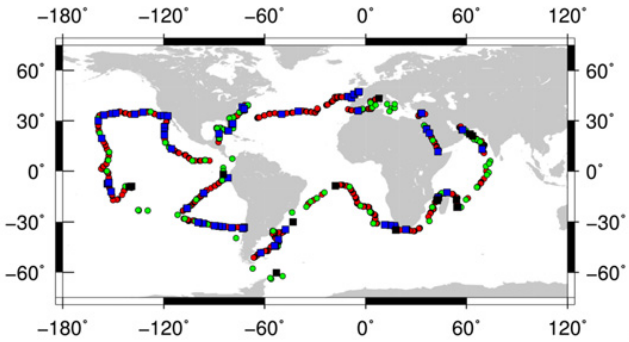


Fig. 1. Global distribution of Tara Oceans stations considered in this analysis. Red circles show all available AC-S hourly 15-minute bins ($N = 1708$). Blue squares show MODISA-to-AC-S match-ups ($N = 165$). Green circles show all available HPLC measurements ($N = 130$). Black squares show MODISA-to-HPLC match-ups ($N = 13$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

marine planktonic organisms (Fig. 1). As part of this expedition, a hyperspectral absorption and attenuation meter (WETLabs, Inc. AC-S) was outfitted within the flow-through system of the R/V Tara (Boss et al., 2013; Slade et al., 2010). Continuous sampling by this inline system alternated between whole seawater and seawater that passed through a 0.2 μm filter, providing calibration independent measurements of the absorption and attenuation of marine particles along the full cruise track (Slade et al., 2010). Ultimately, Tara Oceans collected 454 days of particulate optical properties along 70,000 km. Given their broad spatial and temporal distributions, these optical data records provide a highly unique resource to support operational ocean color satellite validation and bio-optical algorithm refinement activities.

Here, we evaluate the inline AC-S measurements collected as part of the Tara Oceans for use as "ground truth" for the validation of MODISA ocean color data products. To our knowledge, underway measurements of particulate absorption and attenuation coefficients have yet to be used to comprehensively evaluate the quality of satellite-derived C_a and IOPs, let alone particulate organic carbon and phytoplankton community structure. While this requires some modeling to estimate ocean color data products of interest, we show that satellite-to-*in situ* match-ups from these proxy estimations fall well within the envelope of standard match-ups that use direct quantification of biogeochemical variables (e.g., high performance liquid chromotography (HPLC) determination of phytoplankton pigment concentrations). In doing so, we demonstrate the value of flow-through absorption and attenuation meter systems as resources for accumulating substantial volumes of reliable, (potentially) spatially and temporally diverse, and reasonably low-cost data streams for ocean color satellite-to-*in situ* match-up analyses. Furthermore, we explore how continuous, flow-through sampling can assist in reconciling the comparison of different spatial scales (and sub-pixel variabilities) that confound standard satellite-to-*in situ* match-ups analyses (Bailey and Werdell, 2006).

2. Methods

2.1. Tara oceans in situ data

The R/V Tara hosted an inline system within its forecastle bilge that included a WET Labs, Inc. AC-S instrument and Sea-Bird Electronics SBE45 MicroTSG unit, as described in detail in Boss et al. (2013). With regards to the former, flowing seawater from 2 m below sea level entered the system at a Vortex debubbler before a three-way electrically actuated valve that sent the flow either directly to the AC-S instrument or through a 0.2 μm cartridge filter that preceded the AC-S instrument. We processed all data following the methods of Slade et al. (2010), which included residual temperature and salinity corrections. We calculated spectral particulate absorption ($a_p(\lambda) = a_{unfiltered}(\lambda) - a_{filtered}(\lambda)$; m^{-1}) and attenuation ($c_p(\lambda) = c_{unfiltered}(\lambda) - c_{filtered}(\lambda)$; m^{-1}) by interpolating between the filtered readings when unfiltered seawater was measured and performed a residual temperature correction to account for possible slight differences in temperature between the filtered and unfiltered samples (Slade et al., 2010). This provides calibration-independent estimates of $a_p(\lambda)$ and $c_p(\lambda)$,

Author's personal copy

as instrumental drifts and residual calibration errors persist in both the filtered and unfiltered measurements and can therefore be removed by subtracting the former from the latter (as long as the instrumental drift has a significantly longer timescale than the timescale of switching between measurements). We averaged all data into one-minute bins to suppress the high frequency variability (instrument noise plus sample inhomogeneity) that can often mask any low frequency variability of interest. This ultimately resulted in over 310,000 final spectra of absorption and attenuation over 375 days (attenuation was available for 375 of 454 days (Boss et al., 2013)).

We collected near-surface water samples for HPLC pigment analysis at 130 locations along the cruise track of the R/V Tara. For each sample, we vacuum-filtered 2 L of water through 25 mm (in some cases, 47 mm) diameter Whatman GF/F glass filters with 0.7 μm capacity. We stored samples in liquid nitrogen, then at −80°C, until their analysis at the Laboratoire d'Océanographie de Villefranche (LOV; France). At LOV, the filters were extracted in 3 mL (6 mL for 47 mm filters) 100% methanol, disrupted by sonication, and clarified two hours later by vacuum filtration. Within 24 h, the extracts were analyzed by HPLC using a complete 1200 Agilent Technologies system according to the protocol described in Ras et al. (2008).

2.2. Modeling of ocean color data products

Marine IOPs and C_a are the principle geophysical variables derived from satellite measurements of ocean color. Historically, C_a provides the standard climate data record from ocean color satellite time-series (Ras, 2011). More recently, the NASA and the International Ocean Colour Coordinating Group (IOCCG) invested significant effort in improving remotely sensed retrievals of marine IOPs (Werdell et al., 2013; IOCCG, 2006), including those that provide indices of phytoplankton and marine particle community structure (Brewin et al., 2011). With the goal of evaluating MODISA-derived C_a and IOP data records, we generated estimates of C_a, the spectral backscattering coefficients of particles ($b_{bp}(\lambda)$; m^{-1}), and the spectral slopes of $b_{bp}(\lambda)$ (η; unitless) and $c_p(\lambda)$ (γ; unitless) from the Tara inline AC-S time-series.

Following Bricaud et al. (1998), C_a can be related to spectral absorption of phytoplankton ($a_{ph}(\lambda)$; m^{-1}) in the open ocean via a power-law:

$$a_{ph}(\lambda) = A(\lambda)C_a^{B(\lambda)}. \tag{1}$$

Phytoplankton absorption in the red can be estimated using the line height method of Davis et al. (1997) as modified by Boss et al. (2007):

$$a_{ph}(676) = a_p(676) - [39/65a_p(650) + 26/65a_p(715)]. \tag{2}$$

As in Bricaud et al. (1998) and Boss et al. (2013), we developed a statistical relationship between log-transformed $a_{ph}(676)$ and C_a. To do so, we identified 52 HPLC samples collected within 1 h of the AC-S stations prepared for satellite-to-*in situ* match-ups analysis (see Section 2.3 below). Analysis of Type II linear regression yielded:

$$a_{ph}(676) = A(676){C_a}^{B(676)} = 0.0152{C_a}^{0.9055}, \tag{3}$$

which corresponds well with a relationship reported for oceanic assemblages of phytoplankton by Bricaud et al. (1998) ($A = 0.0180$; $B = 0.816$) and for a similar dataset by Boss et al. (2013) ($A = 0.0160$; $B = 0.865$). The correlation coefficient and root mean square error for our derived relationship are 0.88 and 45%, respectively. We produced estimates of C_a via inversion of Eq. (3).

Second, $b_{bp}(\lambda)$ relates to $b_p(\lambda)$ via:

$$b_{bp}(\lambda) = \tilde{b}_{bp}(\lambda)b_p(\lambda) \tag{4a}$$

$$b_{bp}(\lambda) = \tilde{b}_{bp}(\lambda)[c_p(\lambda) - a_p(\lambda)], \tag{4b}$$

where $\tilde{b}_{bp}(\lambda)$ is the dimensionless backscattering ratio that describes the proportion of light scattered in the backward hemisphere by particles, and $b_p(\lambda)$ is the spectral scattering of particles, defined as

Author's personal copy

the difference between attenuation and absorption ($=c_p(\lambda) - a_p(\lambda)$; m^{-1}). Twardowski et al. (2001) proposed a spectrally independent relationship between $\tilde{b}_{bp}(\lambda)$ and C_a:

$$\tilde{b}_{bp} = 0.0096\, C_a^{-0.253}, \tag{5}$$

which yields values of 0.02, 0.011, and 0.008 for C_a of 0.05, 0.5, and 2 mg m^{-3}, respectively. Note that Whitmire et al. (2007) confirmed the spectral independence of Eq. (5) using a diverse *in situ* dataset and reported an average $\tilde{b}_{bp}$ of 0.01. Using C_a from Eq. (3) as input into Eq. (5), $b_{bp}(\lambda)$ can be therefore be estimated as:

$$b_{bp}(\lambda) = (0.0096\, C_a^{-0.253})\, [c_p(\lambda) - a_p(\lambda)]. \tag{6}$$

Finally, we calculated the dimensionless power-law slope for spectral particulate backscattering (η) and attenuation (γ) as:

$$b_{bp}(\lambda) = b_{bp}(\lambda_0)\, [\lambda/\lambda_0]^{-\eta} \tag{7a}$$

$$c_p(\lambda) = c_p(\lambda_0)\, [\lambda/\lambda_0]^{-\gamma}, \tag{7b}$$

using the non-linear minimization approach of Levenberg–Marquardt and $\lambda_0 = 440$ nm. While the power-law function has been found to fit $c_p(\lambda)$ well and to be linked to size distribution parameters (Eq. (7b); Boss et al., 2001), we acknowledge that the validity of a similar relationship for $b_{bp}(\lambda)$ (Eq. (7a)) remains uncertain and requires future research (Slade et al., 2011).

2.3. Satellite data product validation

We generated Level-2 satellite-to-*in situ* match-ups for MODISA using the operational OBPG validation infrastructure (seabass.gsfc.nasa.gov/seabasscgi/search.cgi). We prepared the *in situ* AC-S data for comparison with the satellite measurements by generating 15 min averages of C_a, $b_{bp}(\lambda)$, γ, and η centered on 1:30 PM (in the time zone local to the R/V Tara), which coincided generally with the daily local overpass of MODISA. Satellite data processing and quality assurance for these match-ups followed Bailey and Werdell (2006). Specifically: (a) temporal coincidence was defined as $+/-3$ h; (b) satellite values were the filtered mean of all unmasked pixels in a 5×5 box centered on the *in situ* target; and (c) satellite values were excluded when the median coefficient of variation for unflagged pixels within the box exceeded 0.15. With regards to (a), the satellite-to-*in situ* time difference never exceeded one hour, given our use of averages at local 1:30 PM. We processed MODISA data using its R2013.0 (February 2013) reprocessing configuration. We considered the following MODISA-derived geophysical variables:

- C_a from O'Reilly et al. (1998) (OC3M);
- $b_{bp}(\lambda)$ from Werdell et al. (2013);
- particulate organic carbon (POC; mg m^{-3}) from Stramski et al. (2008); and,
- relative particle size class (PSC; %) from Uitz et al. (2006) and Hirata et al. (2011).

Note that the operational OBPG version of OC3M includes the modifications presented in oceancolor.gsfc.nasa.gov/ANALYSIS/ocv6/. We also downloaded the OBPG operational MODISA-to-*in situ* match-up results for qualitative comparison with those from Tara Oceans. Finally, as MODISA maintains a $\sim$1.1 km^2 footprint at nadir, we calculated the coefficient of variation (COV = standard deviation/mean) for C_a and a 10 nm bandwidth around $b_{bp}(650)$ from the AC-S at 1 km^2 bins along the cruise track to help assess the role of sub-pixel variability in the MODISA-to-*in situ* Tara Oceans match-ups.

3. Results

Direct comparisons of satellite-derived and *in situ* measurements provide estimates of the accuracy and precision of the satellite data products (Bailey and Werdell, 2006). Overall, the MODISA and AC-S-derived C_a compared favorably, particularly for $C_a < 4$ mg m^{-3} (Fig. 2A, Table 1). The slope of

Author's personal copy

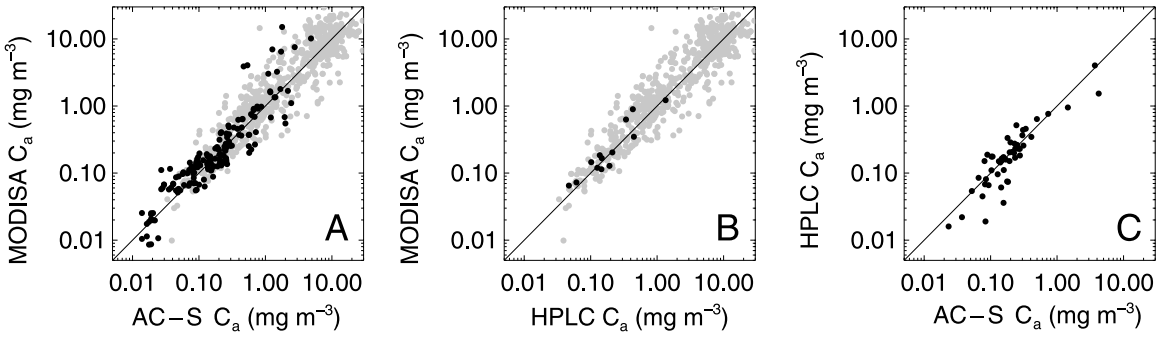


Fig. 2. Comparisons of satellite-derived and *in situ* C_a. (A) MODISA versus AC-S C_a. (B) MODISA versus HPLC C_a. (C) HPLC versus AC-S C_a, not limited to match-ups with MODISA. In (A) and (B), the black circles show the Tara Oceans satellite-to-*in situ* match-ups, and the gray circles show all available MODISA C_a match-ups available from the OBPG and SeaBASS. The solid line shows a 1:1 relationship. Table 1 presents complementary regression statistics.

Table 1
Statistics for MODISA versus AC-S C_a and $b_{bp}(\lambda)$ match-ups determined using Type II linear regression. N, r^2, Slope (SE), Ratio, and MPD indicate sample size, coefficient of determination, regression slope (standard error), median satellite-to-*in situ* ratio, and absolute median percent difference. We calculated all statistics using log-transformed data, with the exception of Ratio and MPD. We calculated MPD as the median of $[200\% * (\text{MODISA}_i - \text{AC-S}_i)/(\text{MODISA}_i + \text{AC-S}_i)]$ for a population of i match-ups. The bottom row shows regression statistics for coincident (+/−30 min) *in situ* C_a derived from the AC-S versus HPLC.

Product	N	r^2	Slope (SE)	Ratio	MPD
C_a – AC-S	165	0.83	1.08 (0.04)	1.08	33.7
C_a – AC-S (< 4 mg m^{-3})	156	0.85	0.98 (0.03)	1.06	31.7
C_a – HPLC	13	0.87	1.03 (0.12)	1.13	25.0
$b_{bp}(412)$	167	0.64	1.09 (0.03)	1.04	22.5
$b_{bp}(443)$	167	0.66	1.12 (0.03)	0.98	19.6
$b_{bp}(488)$	167	0.68	1.15 (0.03)	0.87	18.6
$b_{bp}(531)$	167	0.70	1.16 (0.03)	0.77	25.8
$b_{bp}(547)$	167	0.70	1.16 (0.03)	0.75	28.6
$b_{bp}(667)$	167	0.72	1.24 (0.03)	0.64	43.3
C_a – AC-S vs. HPLC	52	0.78	1.10 (0.08)	1.01	24.5

the Type II linear regression between AC-S and MODISA C_a was slightly positive (1.08), however, the r^2 (0.83) nearly matched that reported by the OBPG for operational MODISA match-ups analyses that include all available SeaBASS data (0.88) (Table 2). Visually, the AC-S match-ups fell within the cloud of all available SeaBASS match-ups (Fig. 2A). When we excluded MODISA-derived $C_a > 4$ mg m^{-3} (less 9 stations of 165 total), the slope reduced to almost unity (0.98) and the r^2 rose to 0.85 (Table 1). In both scenarios, MODISA demonstrated a slight positive bias, with median satellite-to-*in situ* ratios of ∼1.07 and median absolute percent differences (MPD) of ∼32% (the caption for Table 1 presents our calculations of MPD). Per Eq. (3), our AC-S estimates of C_a directly follow our derived, statistical relationship between HPLC estimates of C_a and AC-S estimates of $a_{ph}(676)$.

In the satellite ocean color paradigm, the oceanographic community currently considers HPLC-derived C_a as the state-of-the-art for bio-optical algorithm development and data product validation (Ras et al., 2008; Hooker et al., 2005). Despite the small sample size, the MODISA and HPLC-derived C_a compared very well, with a slope near unity (1.03), a ratio of 1.13, and a MPD of 25% (Table 1). Visually, these match-ups stations fell well within the bounds of the global data set of all available SeaBASS match-ups (Fig. 2B, Table 2). The AC-S and HPLC-derived C_a complemented each other (Fig. 2), which is not surprising given that the latter provided the power-law coefficients provided in Eq. (3) for AC-S processing. Despite a ratio of unity, HPLC values from the clearest waters ($C_a < 0.02$ mg m^{-3}) imparted a positive slope (1.10) and elevated MPD (24.5%) for this comparison. Eliminating the three lowest values (all < 0.02 mg m^{-3}) reduced the slope and MPD to 1.05 and 21.8%, respectively.

Comparisons of MODISA and AC-S-derived $b_{bp}(\lambda)$ largely matched those reported by the OBPG (Tables 1 and 2). The r^2 ranged from 0.64 to 0.72, which improved upon the operational OBPG results (0.57 to 0.62). The MPD and ratios, however, degraded to ranges of 18.6 to 43.3 and 0.64 to 1.04,

Author's personal copy

Table 2
Statistics for MODISA versus *in situ* C_a and $b_{bp}(\lambda)$ match-ups as reported by the OBPG for all available data in SeaBASS (oceancolor.gsfc.nasa.gov/seabasscgi/search.cgi) as of June 2013. Methods and abbreviations as Table 1.

Product	N	r^2	Slope (SE)	Ratio	MPD
C_a	760	0.88	1.02 (0.13)	1.12	30.9
$b_{bp}(412)$	63	0.61	1.05 (0.09)	0.94	17.3
$b_{bp}(443)$	63	0.62	1.08 (0.09)	0.94	17.3
$b_{bp}(488)$	63	0.62	1.11 (0.09)	0.95	17.2
$b_{bp}(531)$	64	0.62	1.12 (0.09)	0.97	17.9
$b_{bp}(547)$	63	0.61	1.14 (0.10)	0.97	18.5
$b_{bp}(667)$	63	0.57	1.17 (0.10)	1.00	19.9

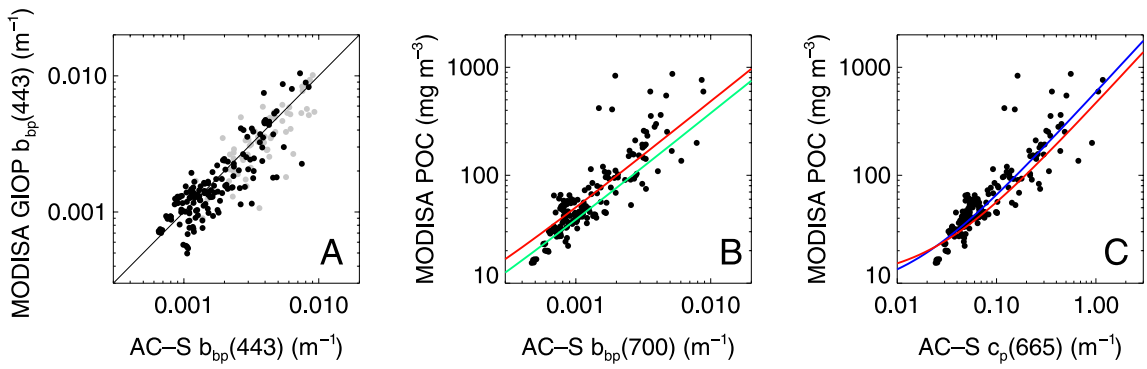


Fig. 3. Comparisons of satellite-derived and *in situ* IOPs and POC. (A) MODISA versus AC-S $b_{bp}(443)$. The solid line shows a 1:1 relationship and Table 1 presents complementary regression statistics. (B) MODISA POC versus AC-S $b_{bp}(700)$ ($N = 167$). The red and green solid lines show the relationships presented in Stramski et al. (2008) and Loisel et al. (2001). (C) MODISA POC versus AC-S $c_p(665)$ ($N = 167$). The red and blue solid lines show the relationships presented in Stramski et al. (2008) and Behrenfeld and Boss (2006). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively (in contrast to 17.2 to 19.9 and 0.94 to 1.0). Similar to that reported in Werdell et al. (2013), the slopes and ratios showed spectral dependence (inverse relative to each other), which indicates potential parameterization issues within the remote-sensing model (GIOP-DC; e.g., Raman effects are ignored). As for the satellite-to-*in situ* C_a match-ups, the AC-S results visually fall within the full dynamic range of operational OBPG results (Fig. 3A). However, the sample sizes for the AC-S match-ups exceeded those for the operational OBPG match-ups by almost three-fold (Tables 1 and 2), and many of these AC-S match-ups fell well below the lowest $b_{bp}(\lambda)$ values available in SeaBASS. Meso- and eutrophic samples dominate the population of data archived in SeaBASS (Werdell and Bailey, 2005), whereas oligotrophic conditions dominate the world's oceans.

Many empirical relationships between POC, $b_{bp}(\lambda)$, and $c_p(\lambda)$ have been proposed (e.g., Stramski et al., 2008; Loisel et al., 2001; Behrenfeld and Boss, 2006; Cetinić et al., 2012). As *in situ* POC were not collected as part of Tara Oceans, we explored these relationships to evaluate the use of optical data as proxy "ground truth" measurements for comparison with MODISA-derived POC. The relationship between MODISA POC and AC-S $b_{bp}(700)$ generally mimicked two common relationships developed for near surface waters (Fig. 3B). While other relationships exist (as reported in Cetinić et al. (2012)), we arbitrarily chose relationships for the surface layer, as ocean color satellite instruments do not "see" far below the first optical depth (i.e., the layer over which light attenuates to ~37% of its magnitude at the surface) and a detailed evaluation of POC–IOP relationships exceeded the scope of this paper. Over their full dynamic ranges, $b_{bp}(700)$ and POC behaved similarly to both Loisel et al. (2001) ($r^2 = 0.76$; root mean square error (RMSE) $= 0.14$ for log-transformed data) and Stramski et al. (2008) ($r^2 = 0.76$; RMSE $= 0.14$). Likewise, the relationship between MODISA POC and AC-S $c_p(665)$ followed two common relationships developed for near surface waters (Fig. 3C). Over their full dynamic ranges, $c_p(665)$ and POC generally followed both Behrenfeld and Boss (2006) ($r^2 = 0.80$; RMSE $= 0.14$) and Stramski et al. (2008) ($r^2 = 0.80$; RMSE $= 0.15$).

Author's personal copy

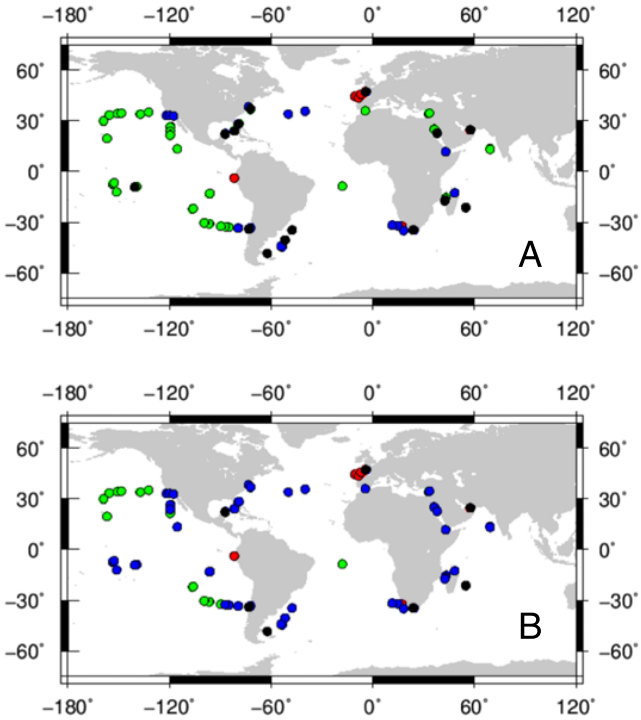


Fig. 4. Dominant PSCs as determined by Uitz et al. (2006) (top) and Hirata et al. (2011) (bottom). Red, blue, and green circles indicate micro-, nano-, and picoplankton, respectively. Black circles indicate stations where a dominant PSC could not be determined. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We expect both γ and η provide an index of particle size (e.g., Boss et al., 2001; Behrenfeld and Boss, 2006; Loisel et al., 2001, 2006), with lower values indicating larger mean particle sizes. The C_a-based PSC algorithms proposed by Uitz et al. (2006) and Hirata et al. (2011) provided straightforward (easily implemented) methods for estimating phytoplankton sizes from MODISA radiometric data. Both report the relative concentrations of micro- (>20 μm), nano- (2–20 μm), and picoplankton (<2 μm). We compared average γ and η for the MODIS-derived dominant size classes. We considered a station to be dominated by a PSC when the relative presence of that class exceeded 45% (Fig. 4). Under this definition, 31 and 16 match-ups stations remained unclassified by Uitz et al. (2006) and Hirata et al. (2011), respectively. For the satellite-to-*in situ* match-ups identified as dominated by micro-, pico-, and nanoplankton by Uitz et al. (2006), the AC-S reported average γ of 0.42, 0.83, and 0.88, respectively (Fig. 5). For the three PSCs assigned by Hirata et al. (2011), the AC-S reported average γ of 0.48, 0.87, and 0.83. Likewise, for the satellite-to-*in situ* match-ups identified as dominated by micro-, pico-, and nanoplankton by Uitz et al. (2006), the AC-S reported average η of 0.10, 0.62, and 0.68, respectively (Fig. 5). For the three PSCs assigned by Hirata et al. (2011), the AC-S reported average η of 0.13, 0.67, and 0.64. The MODIS-derived PSCs, γ, and η all converged to discriminate between stations with the largest microphytoplankton and the smaller nano- and picophytoplankton—e.g., Fig. 5 shows decreasing η and γ with increasing contributions of microplankton and increasing η and γ with increasing contributions of nano- and picoplankton. Effectively discriminating between the two smallest PSCs, however, remained ambiguous for both the remote-sensing and *in situ* optical methods.

Finally, to assess the role of sub-pixel variability in the MODISA-to-*in situ* Tara Oceans match-ups described above, we calculated the COVs for $b_p(650)$ and C_a for 1 km^2 bins along the R/V Tara cruise track (Fig. 6). The median COV for $b_p(650)$ was 0.012 (equivalent to 1.2%), with a 99th percentile of 0.1 (10%). The median COV for C_a was 0.06 (6%), with a 99th percentile of 0.4 (40%). Recall that MODISA maintains a 1.1 km^2 pixel footprint at nadir. In practice, this exercise demonstrated minor within-satellite-pixel variability along the Tara Oceans cruise track and suggested that sub-pixel variability cannot fully explain the mismatch between satellite-derived and *in situ* measurements for 99% of the match-up stations considered in this analysis.

Author's personal copy

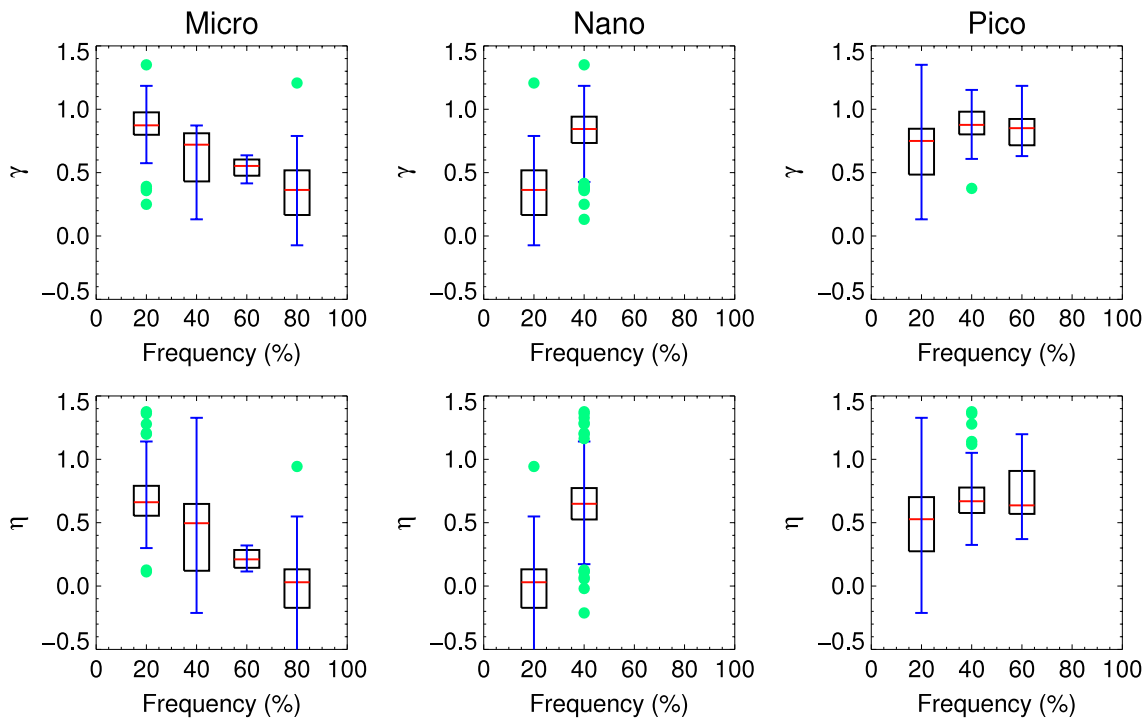


Fig. 5. Box and whisker plots for γ (top row) and η (bottom row) versus PSC frequency as determined by Uitz et al. (2006). The black boxes indicate the range from the first to third quartiles for frequency bins from 10–30, 30–50, 50–70, and 70%–90%. The red lines indicate the median. The blue lines show the range from the minimum to maximum value. The solid green circles indicate outliers (defined as more than 1.5 times the lower or upper quartile). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

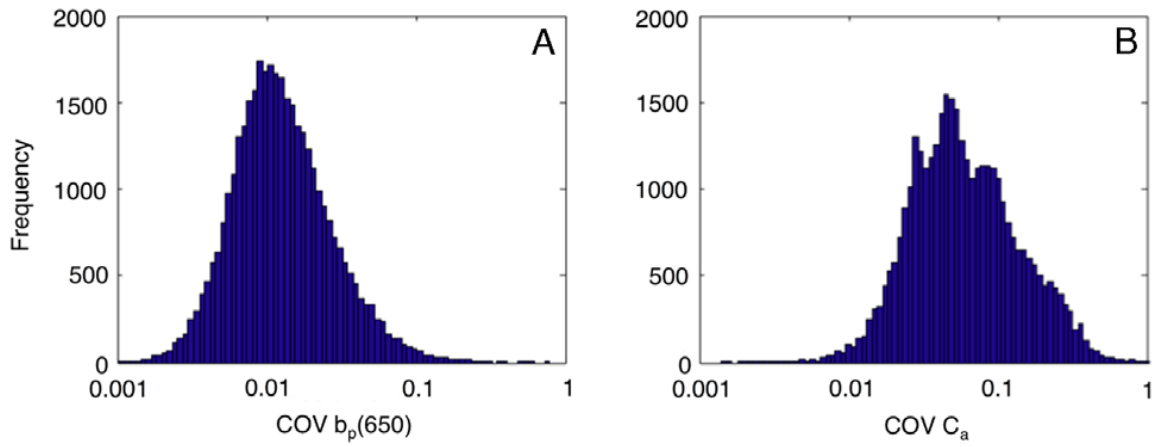


Fig. 6. Coefficients of variation (COV; unitless) of AC-S measurements merged into 1 km^2 spatial bins for $b_p(650)$ (A) and C_a (B). Sample size is 36,000. The median COVs for $b_p(650)$ and C_a are 0.012 and 0.06, respectively. The 99th percentiles for $b_p(650)$ and C_a are 0.1 and 0.4, respectively.

4. Discussion

Developing and validating biogeochemical data records from operational ocean color satellite instruments requires substantial volumes of high quality *in situ* data (Werdell and Bailey, 2005; Bailey and Werdell, 2006; Fargion and McClain, 2003). Given their potentially large temporal and spatial scales, underway expeditions such as Tara Oceans offer significant potential to increase data volumes, provided that data collected on these expeditions are relevant to and of sufficient quality for ocean color satellite calibration and validation activities. We showed how the inline measurements of hyperspectral absorption and attenuation coefficients collected onboard the R/V Tara can be used to evaluate MODISA estimates of C_a, $b_{bp}(\lambda)$, POC, and PSCs. With regards to such validation activities, the

Author's personal copy

most obvious strength of such flow-through measurements remains their sampling rate. The 375 days of AC-S measurements resulted in >165 viable MODIS-to-*in situ* match-ups (Table 1). Furthermore, the continuous sampling provided sufficient data to confirm that sub-pixel variability does not drive mismatches between the satellite and *in situ* measurements (Fig. 6). A clear weakness, however, remains the need to apply bio-optical models to estimate biogeochemical quantities of interest (e.g., C_a from $a_p(\lambda)$ and $c_p(\lambda)$). Our results indicate this can be done somewhat reliably and effectively, however, they also point to model assumptions and parameterizations that can be improved upon.

Chlorophyll-a can be reliably estimated using $a_p(\lambda)$ via the line-height method (Eq. (2); Boss et al., 2007), given reasonable estimates of how C_a varies with $a_{ph}(\lambda)$ (Eq. (3); Bricaud et al., 1998). We derived a cruise-specific relationship between C_a and $a_{ph}(\lambda)$, and realized MODISA-to-*in situ* match-ups of comparable quality to those from HPLC-derived C_a. The number of AC-S match-ups exceeded that for the HPLC match-ups by 100-fold (165 versus 14), which reinforces the benefit of inline time-series when in need of high volumes of data (say, during the first year of a new satellite mission). However, improved results might be achieved with refined line-height methods (Roesler and Barnard, 2013) and progressively more robust estimates of C_a from $a_{ph}(\lambda)$. Given the near consistency of $\tilde{b}_{bp}$ in the open ocean (Twardowski et al., 2001; Whitmire et al., 2007), $b_{bp}(\lambda)$ can also be reliably estimated from $b_p(\lambda)$. Not only did the MODISA-to-AC-S match-ups of $b_{bp}(\lambda)$ modestly agree with those reported by the OBPG for all available SeaBASS data (Tables 1 and 2; Fig. 3A), but the sample size of the former exceeded the latter three-fold. While $b_{bp}(\lambda)$ modeled from $a_p(\lambda)$ and $c_p(\lambda)$ remains several steps removed from the direct measurement of $b_{bp}(\lambda)$ (Twardowski et al., 2001), our results indicate they are of sufficient quality for use in ocean color validation activities, at least early in a new mission for preliminary assessment of remotely-sensed variables.

Many published relationships between POC, $b_{bp}(\lambda)$, and $c_p(\lambda)$ exist, however, most were developed using spatially-limited datasets and with varied *in situ* measurement protocols (Cetinić et al., 2012). But, a paucity of POC measurements exists in SeaBASS (although, we acknowledge that larger datasets exist elsewhere), indicating a need to pursue optical proxies for use in satellite validation activities. Our AC-S-derived IOPs provided a reasonable first-order verification of satellite-derived POC (Fig. 3). The relationships between POC-and-$b_{bp}(\lambda)$ and POC-and-$c_p(\lambda)$ both largely followed previously published relationships for the near-surface ocean. Improved agreement between satellite POC and *in situ* optics might be realized, however, if discrete *in situ* measurements had been made of POC as part of Tara Oceans. As we did for the AC-S-derived C_a using HPLC analyses, such discrete measurements could be used to tune POC-to-IOP relationships for the expedition cruise track. As for POC, large volumes of direct measurements of PSCs remain uncommon in most public databases. To overcome this, the ocean color community commonly uses HPLC-derived pigments as proxy indicators of phytoplankton community structure (see, e.g., the discussion of diagnostic pigment analyses presented in Uitz et al. (2006) and Hirata et al. (2011) and references therein). But, the optical parameters η and γ provide alternate proxy indicators of particle sizes (Boss et al., 2001; Loisel et al., 2006). On average, the underway AC-S estimates of η and γ appear to effectively discriminate between the largest and smallest particles (microphytoplankton versus nano- and picoplankton) (Fig. 5). The use of these *in situ* proxies for PSCs remains an ongoing, unresolved issue (both using HPLC and optics), as does achieving convergence in remote-sensing methods to estimate PSCs (e.g., Brewin et al., 2011; Fig. 4). Our results contribute to an emerging body of research on the validation of remotely sensed PSCs, however, significant work remains to truly understand relationships between varied indirect estimates of particle sizes.

We did not engage in this activity to ultimately propose that underway measurements of IOPs replace discrete biogeochemical measurements or vertical profiles in ocean color algorithm development and satellite data product validation activities. Rather, we hoped to demonstrate that underway measurements provide a previously unexploited means for achieving substantial volumes of complementary, high quality data for use in these activities. Tuning the AC-S estimates of C_a to sparsely and discretely sampled HPLC-derived C_a during the Tara Oceans expedition extended its sample size of satellite-to-*in situ* match-ups from 13 to 165. Furthermore, in the absence of a broad, institutionally supported program such as SIMBIOS, opportunistically outfitting sea-going vessels (of any purpose) with an inline system could provide high volumes of data with low costs relative to other data acquisition strategies. For example, for the 18-month overlap between the SIMBIOS Program and the

Author's personal copy

MODISA mission (mid 2002 through 2003), the sample size of MODISA-to-*in situ* C_a match-ups averaged 155 per year. In the four years that followed the conclusion of SIMBIOS (2004–2007), this rate fell to 104 C_a match-ups per year. In three of those years (2005–2007), the average rate fell further to 54 per year. In contrast, in the two and a half year overlap between Tara Oceans and MODISA, the sample size of MODISA-to-AC-S-derived C_a match-ups averaged 66 per year. In our current era of imposing rapid validation requirements on new satellite missions – and our community-wide need to develop remote sensing algorithms and validate satellite data records on unprecedented temporal and spatial scales – our results have significant implications on how forthcoming calibration and validation teams for existing and upcoming satellite missions will be constructed.

Acknowledgments

We thank S. Bailey, A. Chase, J. Ras, and H. Claustre for their helpful advice. We also thank staff at the Goddard Space Flight Center Ocean Ecology Laboratory for their support and J. Loftin, S. Searson, H. Le Goff, and S. Kandels for their handling of the AC-S during Tara Oceans. Finally, we thank the following people, institutions, and sponsors who made this singular expedition possible: CNRS, EMBL, Genoscope/CEA, UPMC VIB, Stazione Zoologica Anton Dohm, UNIMIB, ANR, FWO, BIO5, Biosphere 2, agnes b., the Veolia Environment Foundation, Region Bretagne, World Courier, Cap L'Orient, the Foundation EDF Diversiterre, FRB, the Prince Albert II de Monaco Foundation, Etienne Bourgois, and the Tara Foundation teams and crew. Tara Oceans could not have happened without the support of the Tara Foundation and the Tara Consortium. This is contribution no. 9 of the Tara Oceans Expedition 2009–2012. Funding for the collection and processing of the AC-S dataset was provided by NASA Ocean Biology and Biogeochemistry Program under grants NNX11AQ14G and NNX09AU43G to the University of Maine.

References

NRC 2011. Assessing Requirements for Sustained Ocean Color Research and Operations. National Academies Press, p. 100.

IOCCG 2006. Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications, Reports of the International Ocean Colour Coordinating Group No. 5, p. 126.

Bailey, S.W., Werdell, P.J., 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sens. Environ. 102, 12–23.

Behrenfeld, M.J., Boss, E., 2006. Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass. J. Mar. Res. 64, 431–451.

Boss, E.S., Collier, R., Larson, G., Fennel, K., Pegau, W.S., 2007. Measurements of spectral optical properties and their relation to biogeochemical variables and processes in Crater Lake, Crater Lake National Park, OR. Hydrobiol. 574, 149–159.

Boss, E., Pegau, W.S., Gardner, W.D., Zaneveld, J.R.V., Barnard, A.H., Twardowski, M.S., Chang, G.C., Dickey, T.D., 2001. Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf. J. Geophys. Res. 106, 9509–9516.

Boss, E., Picheral, M., Leeuw, T., Chase, A., Karsenti, E., Gorsky, G., Taylor, L., Slade, W., Ras, J., Claustre, H., 2013. The characteristics of particulate absorption, scattering and attenuation coefficients in the surface ocean; Contribution of the Tara Oceans expedition. Methods Oceanogr. 7, 52–62.

Brewin, R.J.W., Hardman-Mountford, N.J., Lavender, S.J., Raitsos, D.E., Hirata, T., Uitz, J., Devred, E., Bricaud, A., Ciotti, A., Gentili, B., 2011. An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. Remote Sens. Environ. 115, 325–339.

Bricaud, A., Morel, A., Babin, M., Allali, K., Claustre, H., 1998. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: analysis and implications for bio-optical models. J. Geophys. Res. 103, 31033–31044.

Cetinić, I., Perry, M.J., Briggs, N.T., Kallin, E., D'Asaro, E.A., Lee, C.M., 2012. Particulate organic carbon and inherent optical properties during 2008 North Atlantic bloom experiment. J. Geophys. Res. 117, http://dx.doi.org/10.1029/2011JC007771.

Davis, R.F., Moore, C.C., Zaneveld, J.R.V., Napp, J.M., 1997. Reducing the effects of fouling on chlorophyll estimates derived from long-term deployments of optical instruments. J. Geophys. Res. 102, 5851–5855.

Fargion, G.S., McClain, C.R., 2003. SIMBIOS Project 2003 Annual Report. NASA Tech. Memo. 2003–212251, 202 pp.

Gordon, H.R., Wang, M., 1994. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Appl. Opt. 33, 443–452.

Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., Yamanaka, Y., 2011. Synoptic relationships between surface chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences 8, 311–327.

Hooker, S.B., Van Heukelem, L., Thomas, C.S., et al., 2005. The second SeaWiFS HPLC analysis round robin experiment (SeaHARRE-2). NASA Tech. Memo. 2005–212785, 124 pp.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., et al., 2011. A holistic approach to marine ecosystems biology. PLoS Biol. 9, e1001177. http://dx.doi.org/10.1371/journal.pbio.1001177.

Author's personal copy

Loisel, H., Bosc, E., Stramski, D., Oubelkheir, K., Deschamps, P.Y., 2001. Seasonal variability of the backscattering coefficient in the Mediterranean Sea based on satellite SeaWiFS imagery. Geophys. Res. Let. 28, 4203–4206.
Loisel, H., Nicolas, J-M., Sciandra, A., Stramski, D., Poteau, A., 2006. Spectral dependency of optical backscattering by marine particles from satellite remote sensing of the global ocean. J. Geophys. Res. 111, http://dx.doi.org/10.1029/2005JC003367.
McClain, C.R., 2009. A decade of satellite ocean color observations. Ann. Rev. Mar. Sci. 1, 19–42.
O'Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. J. Geophys. Res. 103, 24937–24953.
Ras, J., Uitz, J., Claustre, H., 2008. Spatial variability of phytoplankton pigment distribution in the subtropical South Pacific Ocean: comparison between in situ and modelled data. Biogeosciences 5, 353–369.
Roesler, C.S., Barnard, A.H., 2013. Optical proxy for phytoplankton biomass in the absence of photophysiology: rethinking the absorption line height. Methods Oceanogr. 7, 79–94.
Siegel, D.A., Behrenfeld, M.J., Maritorena, S., McClain, C.R., Antoine, D., et al., 2013. Regional to global assessments of phytoplankton dynamics from the SeaWiFS mission. Rem. Sens. Environ. 135, 77–91.
Slade, W.H., Boss, E., Dall'Olmo, G., Langner, M.R., Loftin, J., Behrenfeld, M.J., Roesler, C., Westberry, T.K., 2010. Underway and moored methods for improving accuracy in measurement of spectral particulate absorption and attenuation. J. Atmos. Ocean. Tech. 27, 1733–1746.
Slade, W.H., Boss, E., Russo, C., 2011. Effects of particle aggregation and disaggregation on their inherent optical properties. Opt. Express 19, 7945–7959.
Stramski, D., Reynolds, R.A., Babin, M., Kaczmarek, S., Lewis, M.R., et al., 2008. Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and Eastern Atlantic Oceans. Biogeosciences 5, 171–201.
Twardowski, M.S., Boss, E., Macdonald, J.B., Pegau, W.S., Barnard, A.H., Zaneveld, J.R.V., 2001. A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. J. Geophys. Res. 106, 14129–14142.
Uitz, J., Claustre, H., Morel, A., Hooker, S.B., 2006. Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. J. Geophys. Res. 111, http://dx.doi.org/10.1029/2005JC003207.
Werdell, P.J., Bailey, S.W., 2005. An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sens. Environ. 98, 122–140.
Werdell, P.J., Bailey, S., Fargion, G., Pietra, C., Knobelspiesse, K., Feldman, G., McClain, C., 2003. Unique data repository facilitates ocean color satellite validation. EOS Trans. AGU 84, 377.
Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., et al., 2013. Generalized ocean color inversion model for retrieving marine inherent optical properties. Appl. Opt. 52, 2019–2037.
Whitmire, A.L., Boss, E., Cowles, T.J., Pegau, W.S., 2007. Spectral variability of the particulate backscattering ratio. Opt. Express 15, 7019–7031.