

Validation of Ocean Color Remote Sensing Reflectance Using Autonomous Floats

GREGORY P. GERBI,^a EMMANUEL BOSS,^b P. JEREMY WERDELL,^c CHRISTOPHER W. PROCTOR,^{c,d}
 NILS HAËNTJENS,^b MARLON R. LEWIS,^{e,f} KEITH BROWN,^f DIEGO SORRENTINO,^f
 J. RONALD V. ZANEVELD,^g ANDREW H. BARNARD,^g JOHN KOEGLER,^g HUGH FARGHER,^h
 MATTHEW DEDONATO,^h AND WILLIAM WALLACE^h

^a Skidmore College, Saratoga Springs, New York

^b University of Maine, Orono, Maine

^c NASA Goddard Space Flight Center, Greenbelt, Maryland

^d Science Systems and Applications, Inc., Lanham, Maryland

^e Dalhousie University, Halifax, Nova Scotia, Canada

^f Atlantic, Halifax, Nova Scotia, Canada

^g WET Labs, Philomath, Oregon

^h Teledyne Webb Research, North Falmouth, Massachusetts

(Manuscript received 28 March 2016, in final form 28 July 2016)

ABSTRACT

The use of autonomous profiling floats for observational estimates of radiometric quantities in the ocean is explored, and the use of this platform for validation of satellite-based estimates of remote sensing reflectance in the ocean is examined. This effort includes comparing quantities estimated from float and satellite data at nominal wavelengths of 412, 443, 488, and 555 nm, and examining sources and magnitudes of uncertainty in the float estimates. This study had 65 occurrences of coincident high-quality observations from floats and MODIS *Aqua* and 15 occurrences of coincident high-quality observations floats and Visible Infrared Imaging Radiometer Suite (VIIRS). The float estimates of remote sensing reflectance are similar to the satellite estimates, with disagreement of a few percent in most wavelengths. The variability of the float–satellite comparisons is similar to the variability of in situ–satellite comparisons using a validation dataset from the Marine Optical Buoy (MOBY). This, combined with the agreement of float-based and satellite-based quantities, suggests that floats are likely a good platform for validation of satellite-based estimates of remote sensing reflectance.

1. Introduction

Satellite-based radiometry missions of the ocean require validation of data products by in situ measurements to assess and improve, if necessary, the accuracy of satellite-derived quantities (Mueller et al. 2003b). The most fundamental of the quantities estimated by satellites are water-leaving radiance L_w and remote sensing reflectance R_{rs} (the ratio of upwelling radiance to downwelling irradiance at the ocean surface). These are used to determine optical and biological quantities in the upper ocean using theoretical and empirical algorithms. In addition to product validation, satellite missions must undergo vicarious calibration of their radiometers through

which in situ measurements are used to determine a single set of spectrally dependent calibration factors (possibly with weak time dependence) for application to all locations and times (Clark et al. 1997; Franz et al. 2007).

The two major instrumented deep-water locations for vicarious calibration and product validation have been the Marine Optical Buoy (MOBY) offshore of Lanai, Hawaii (Clark et al. 2003), and Bouée pour l'acquisition de Séries Optiques à Long Terme (BOUSSOLE) in the Mediterranean offshore of Nice, France (Antoine et al. 2008). For validation, these are complemented by shallow-water sites of the ocean color component of the Aerosol Robotic Network (Zibordi et al. 2009). Vicarious calibration is generally performed with more exacting quality control standards than product validation and in environments with limited spatial variability and well-understood atmospheric conditions (Franz et al. 2007; Voss et al. 2010; Zibordi et al. 2015). Product validation, however, benefits from a broad set of observations over a wider range of

Corresponding author address: Gregory P. Gerbi, Physics Department, Skidmore College, 815 North Broadway, Saratoga Springs, NY 12866.
 E-mail: ggerbi@skidmore.edu

natural variability (Hooker et al. 2007; Werdell et al. 2007). Although the established sites provide excellent continuous in situ data, for validation it is advantageous to collect data from additional open ocean sites that more completely span the natural range of environmental conditions.

This paper explores the use of autonomous profiling floats (Davis et al. 2001) for in situ radiometry measurements. By drifting freely in the ocean and moving vertically by changing their buoyancy, profiling floats allow coverage of a wide range of locations, ocean optical properties, and atmospheric conditions. Floats currently provide more hydrographic data than any other platform in the ocean (Roemmich et al. 2009), and they are commonly deployed for periods of several years. Early in their development, a few floats were equipped successfully with radiometers (Mitchell et al. 2000). More recently they have carried other optical instruments used for estimating quantities, such as beam attenuation, absorption by colored dissolved materials, and chlorophyll concentration (e.g., Bishop et al. 2004; Boss et al. 2008; Xing et al. 2012; Estapa et al. 2013; Xing et al. 2014). Organelli et al. (2016) used autonomous floats to measure vertical profiles of downwelling irradiance in the upper 250 m of the ocean. Many of the strengths and difficulties associated with using autonomous floats to measure optical properties were reviewed in a report by the International Ocean-Colour Coordinating Group (IOCCG 2011).

In this study we describe the use of autonomous profiling floats as a new platform for validation of satellite estimates of remote sensing reflectance. Although our dataset is not sufficient to confirm the suitability of floats for vicarious calibration, this use may be possible in the future. For this analysis we examine the reliability of the float estimates in comparison to the satellite estimates. In the following sections, we describe our methods (section 2), present results (section 3), examine sizes and causes of uncertainty (section 4), discuss the suitability of floats for validation (section 5), and offer brief conclusions (section 6).

2. Methods

We used a set of six autonomous profiling floats deployed in the Mediterranean Sea, Pacific Ocean, and Atlantic Ocean to estimate radiometric quantities, chiefly R_{rs} at four wavelengths, nominally 412, 443, 488, and 555 nm. We compared these in situ estimates to estimates made by the ocean color measurements made by MODIS *Aqua* and the Visible Infrared Imaging Radiometer Suite (VIIRS) on board *Suomi-NPP* that have been vicariously calibrated using MOBY observations. The following subsections describe the floats and float measurements, the satellite measurements, quality control, and the statistics that we used in the analysis.



FIG. 1. Photograph of Mediterranean A float after recovery near BOUSSOLE. The optics package is the blackbody attached to the side of the float. The radiometers are marked with arrows (red: downwelling irradiance, orange: upwelling radiance). (Photograph by Florent Besson.)

a. Float vehicle and instrumentation

The vehicle is the Autonomous Profiling Explorer (APEX) profiling float manufactured by Teledyne Webb Research with computational hardware generation APF9i. The standard APEX firmware was modified to handle additional instruments and to accomplish the sampling goals of the mission. Each deployment had slightly different firmware, as improvements were made between deployments.

Instruments on the floats measured physical and optical quantities, including salinity, temperature, pressure, downwelling irradiance, upwelling radiance, several inherent optical properties, and oxygen concentration. Only radiometric quantities and pressure are used in this study. Both radiometers were model OCR-504, manufactured by Satlantic. The upward-looking radiometer (irradiance) was on top of the float and occupied the highest position to have an unobstructed view of the sky (Fig. 1).

TABLE 1. Summary of central sensor wavelengths (mm) used in direct comparisons. Float deployments are described in Table 2.

MODIS <i>Aqua</i>	412	443	488	555
VIIRS	410	443	486	551
Mediterranean A	411.3	443.8	489.7	554.1
Mediterranean B	412.0	443.4	489.7	554.2
Hawaii A	412.1	443.9	489.7	554.7
Hawaii B	411.8	443.4	489.6	554.1
Atlantic A	412.2	443.8	489.4	555.0
Atlantic B	412.2	443.8	489.4	555.0

The downward-looking radiometer (radiance) was on the base of the optics package (assembled by WET Labs), which was attached to the side of the float. The base of the radiometer was slightly higher than the base of the float, but the float was outside the field of view of the radiometer.

The radiometer bands were chosen to be close to the bands on MODIS *Aqua* and have a 10-nm bandwidth (full width at half maximum response amplitude). We made direct comparisons of R_{rs} between floats and satellites without making any corrections for slight differences in wavelength (a few nanometers) or sensor bandwidth. Throughout this paper we label each band with the wavelength of the equivalent band on MODIS *Aqua* (Table 1).

b. Float deployments

We deployed floats in three pairs in this study, one pair near the BOUSSOLE optical mooring in the Mediterranean Sea (Antoine et al. 2008), one pair near the MOBY optical mooring near Hawaii (Clark et al. 2003), and one pair northwest of Bermuda (Table 2). In each deployment, two floats were released relatively close to each other in space and time, within several kilometers and one day (Fig. 2). The floats generally remained near each other for several days after deployment, but they eventually diverged onto different paths. Two-way communication via Iridium allowed variation of sampling and float behavior for each profile, although we kept most parameters constant for the majority of each deployment.

c. Float behavior and sampling

Each float profile is divided into an ascent phase (rising to the surface from depth) and a buoy phase (drifting at the surface). In standard operation the floats were parked at ~ 1000 m to minimize fouling (following IOCCG 2011) and were profiled once every 2 days. For the near-surface measurements used in this study, the target ascent rate was 4 cm s^{-1} . Radiometric data were sampled continuously at 1 Hz, and during each 1-Hz sample, the radiometers integrated over approximately 0.933 s. In standard APEX sampling, pressure measurements are made at specified depth locations limited to about 1-m resolution. For the BOUSSOLE and Hawaiian floats, the coarse depth resolution and variability in ascent rate was a source of depth uncertainty. To improve the depth estimates for the 1-Hz optical data in the deployment of the Atlantic floats, we modified the APEX firmware to provide higher-frequency pressure measurements near the surface. These measurements were taken approximately once every 3 s (1/3 Hz) with longer gaps (about 15 s) at times when T and S measurements are made (1–2-m spacing).

After reaching the surface and finishing their ascent, the floats waited for about 10 min before sampling in their buoy phase. This delay allowed time for the floats to fully inflate their oil and air bladders and determine location. We targeted the local solar time of 1330 for surfacing, and the floats were generally within 30 min of this target. During the buoy phase, the downwelling irradiance sensor was about 0.3 m above the sea surface, and the upwelling radiance sensor was about 1.12 m below the surface. Uncertainty in the mean depth (ignoring bobbing) of the radiometer below the surface is likely no more than a few centimeters.

d. Float estimates of water-leaving radiance

For each profile, the L_w was computed using measurements of the upwelling radiance made during the buoy phase, $L_u(z_b)$, where z_b is the vertical location of the upwelling radiance sensor during the buoy phase measurement. The term L_u is extrapolated to the surface using the diffuse attenuation coefficient for upward

TABLE 2. Summary of deployments. Latitude and longitude are those for the first profile.

Float name	Location	Start date	End date	No. profiles	Lat	Lon
Mediterranean A	Near BOUSSOLE	12 Jul 2011	13 Sep 2011	29	43.36°	7.91°
Mediterranean B	Near BOUSSOLE	12 Jul 2011	15 Sep 2011	28	43.35°	7.93°
Hawaii A	Near MOBY	17 Dec 2011	20 Jul 2013	332	20.81°	-157.19°
Hawaii B	Near MOBY	18 Dec 2011	8 Nov 2013	360	20.81°	-157.19°
Atlantic A	Northwest of Bermuda	3 May 2012	29 Sep 2013	300	33.18°	-65.71°
Atlantic B	Northwest of Bermuda	3 May 2012	12 Nov 2012	132	33.28°	-65.75°

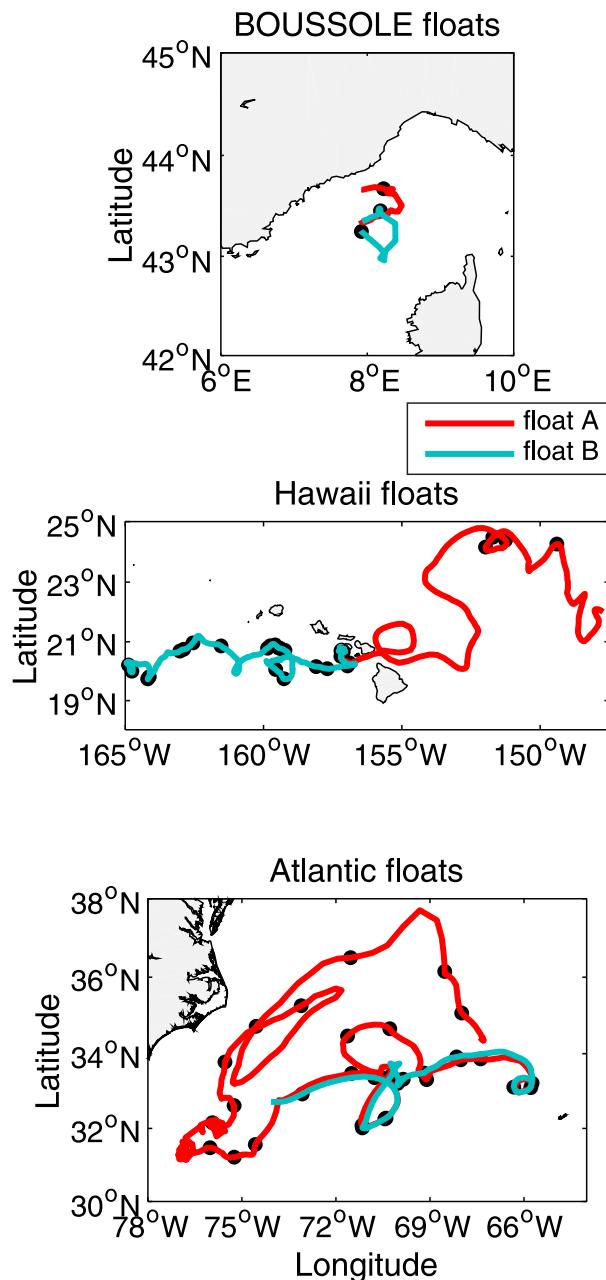


FIG. 2. Maps showing float trajectories for all deployments. Lines show paths over full lifetimes of floats. Dots show locations of profiles used in this study.

radiance K_L that was estimated during the float's ascent (described in more detail below). Both L_u and K_L depend on wavelength, but for compact notation the wavelength dependence is implied. In all calculations we only used samples whose tilts were smaller than a threshold of 5° on either axis. We note that determining the representative tilt for each measurement in this study is problematic. The tilt sensors made single quasi-instantaneous measurements for each 0.933-s radiance

measurement, so the tilt was likely not constant over the full measurement interval. It is possible that measurements with a 5° tilt were at tilts a few degrees higher or lower during their sampling interval but still likely within the NASA recommendation of $\pm 10^\circ$ (Mueller et al. 2003a).

Each float is azimuthally asymmetric, and the float itself can cast a shadow with relatively sharp edges. For buoy phase measurements, we eliminated observations at unfavorable orientations relative to the sun by using observations of L_u only at times when the optics package and sun were within $\pm 90^\circ$ of alignment, where 0° is the direction at which the optics package is directly between the float and the sun. A shading correction similar to that described by Leathers et al. (2004) was used for the buoy phase measurements (described further below). For the ascent phase measurements, we used observations at all relative azimuths, regardless of the likelihood of shading, and we did not perform a shading correction.

We describe the ascent phase measurements first, the buoy phase measurements second, and extrapolation to and through the sea surface third.

1) DIFFUSE ATTENUATION COEFFICIENT

The diffuse attenuation coefficient K_L was computed from observations made during the float ascent. For the Atlantic floats (the two for which we have high-frequency pressure measurements), the median ascent rate between 13.5 and 1.5 m was 4.3 cm s^{-1} with a standard deviation of 1.3 cm s^{-1} . Ninety-five percent of the ascent rates were slower than 5.8 cm s^{-1} . Because the radiometers integrated for 0.933 s, each sample represents an average over 4.0 and 5.4 cm for these median and limiting ascent rates, respectively. These ascent rates and integrated sampling were chosen, in part, to minimize variability caused by wave focusing (Stramska and Dickey 1998; Zibordi et al. 2004; D'Alimonte et al. 2010).

The ascent observations of L_u were grouped into four bins of 3-m thickness between depths of 13.5 and 1.5 m. Within each bin, we estimated parameters with a least squares fit of the model

$$L_u(z - z_m) = \widehat{L}_u(z_m) e^{(z - z_m)K_L}, \quad (1)$$

where z is the vertical coordinate, positive upward, with $z = 0$ at the sea surface (Fig. 3). The quantity on the left side is an observation, and the quantities $\widehat{L}_u(z_m)$ and K_L on the right side are parameters estimated from the model fit. For this fitting we took z_m as the mean depth of all the observations within the bin. The diffuse attenuation coefficient estimated for the shallowest bin was used to propagate $L_u(z_b)$ to the surface. Values of K_L for deeper bins were used for quality control.

We used observations at all relative azimuths to maintain data density during ascent. Eliminating certain headings left many profiles with only a small number of useable observations. As will be discussed in section 4, this is likely to increase the uncertainty in K_L estimates, but it is unlikely to contribute to the systematic bias in K_L estimates.

2) NEAR-SURFACE RADIANCE AND SHADING CORRECTION

During the buoy phase, the radiance sensor was at a nominal depth of $z_b = -1.12$ m. This depth is likely to have varied by several centimeters as the float bobbed in the sea surface or was passed by waves, but because the float's pressure sensor is exposed during buoy phase, we have no subsurface pressure measurements with which to estimate the magnitude of the depth variations. Sampling intervals varied, but for most profiles in the Hawaiian and Atlantic deployments, we collected 50–100 buoy phase samples over 2–5 min (each with a duration of 0.933 s). For BOUSSOLE profiles, floats sampled every second for about 10 min, giving several hundred buoy phase samples per profile.

We made a shading correction for each buoy phase sample following the method of Leathers et al. (2004). This method ignores skylight, assumes no scattering, and assumes that absorption is proportional to the diffuse attenuation coefficient. The floats have a more complicated shape than the cylindrical buoys described by Leathers et al., which we modeled as three cylinders. Each casts a shadow beneath the radiometer, and the deepest shaded depth was used in the shading correction. Because the radiometer is located outside the cylinder of the float vehicle, the correction is strongly dependent on the azimuthal direction of the sun relative to the optics package. For L_u , we used only measurements for which the radiometer was on the sunny side of the float. Unfortunately, we failed to calibrate the compasses after assembly for the floats deployed near BOUSSOLE and Hawaii. Based on the calibration of the compasses for the Atlantic deployment, the heading uncertainty in the other floats may be as large as 30° for some azimuths, so it is possible that relative azimuth varied $\pm 120^\circ$ (rather than $\pm 90^\circ$) in these floats.

We computed $L_u(z_b)$ as the mean of all the shade-corrected radiance measurements that passed the tilt and heading elimination criteria. Using mean or median to compute $L_u(z_b)$ gave similar results.

3) EXTRAPOLATION TO AND PROJECTION THROUGH SEA SURFACE

The $L_u(z_b)$ was extrapolated to the sea surface assuming a constant diffuse attenuation coefficient and using the exponential relationship

$$L_u(0-) = L_u(z_b)e^{-z_b K_L}, \quad (2)$$

where $L_u(0-)$ is the upwelling radiance immediately below the sea surface and K_L is the diffuse attenuation coefficient for the top bin. This upwelling radiance was then projected through the sea surface as, following Austin (1974) and Gordon and Clark (1981),

$$L_w = L_u(0-) \frac{1-\rho}{n_w^2}, \quad (3)$$

where ρ is the internal Fresnel reflectance and n_w is the index of refraction of seawater relative to air (Quan and Fry 1995). Fresnel reflectance was computed (Mobley 1994) as

$$\rho = \left(\frac{n_w - 1}{n_w + 1} \right)^2. \quad (4)$$

e. Computing in situ remote sensing reflectance

We computed remote sensing reflectance R_{rs} , using in situ estimates of L_w and modeled estimates of downwelling irradiance at the sea surface E_s , determined using the clear-sky model of Frouin et al. (1989), similar to Bailey and Werdell (2006). Using modeled values for E_s was necessary because of the poor quality of in situ observations.

Remote sensing reflectance was computed iteratively as

$$R_{rs} = \tilde{R}_{rs} F, \quad (5)$$

where

$$\tilde{R}_{rs} = \frac{L_w}{E_s} \quad (6)$$

is a first estimate, and F is a bidirectional reflectance correction computed using instrument geometry, solar geometry, and an estimate of chlorophyll-*a* concentration as inputs (Morel et al. 2002). This correction adjusts all R_{rs} estimates to a common condition of an overhead sun and nadir-viewing geometry. The initial estimate \tilde{R}_{rs} was used to compute F using a blue-green reflectance band ratio to estimate chlorophyll (O'Reilly et al. 1998).

f. Satellite observations

We acquired level 2 satellite R_{rs} estimates for MODIS *Aqua* and VIIRS using the operational NASA Ocean Biology Processing Group validation infrastructure (<http://seabass.gsfc.nasa.gov/seabasscgi/search.cgi>). MODIS *Aqua* and VIIRS data were processed using their R2013.1 configurations from September 2013

and June 2013, respectively. Satellite data processing and quality assurance followed Bailey and Werdell (2006). Specifically, satellite values were the filtered mean (via the interquartile range) of all unflagged pixels in a 5×5 box centered on the in situ target. The maximum allowable time difference between float and satellite observations was defined as ± 3 h. Satellite values were excluded if they failed the other criteria thresholds defined by Bailey and Werdell, which included testing for extreme variation between pixels, excessive solar zenith or satellite zenith angles, minimum numbers of unflagged pixels, and if multiple satellite files matched the in situ data, only the closest in time was used.

g. Statistics for analysis and quality control

Because this study is effectively validating the float observations, we take the satellite observations as the “truth.” Our primary metric is the ratio of float to satellite estimates of R_{rs} , which we label G :

$$G = \frac{R_f}{R_s}, \tag{7}$$

where R_f is the float estimate of R_{rs} and R_s is the satellite estimate of R_{rs} . The mean ratio over the full deployment is \bar{G} :

$$\bar{G} = \frac{1}{N} \sum_{i=1}^N G_i, \tag{8}$$

where N is the number of profiles, i is an index variable, and G_i refers to the results from individual pairs of satellite and float observations (matchups). In addition to the mean and median, we compute the standard deviation (σ_G), standard error (σ_G/\sqrt{N}), and kurtosis of G_i following their conventional definitions (e.g., Bendat and Piersol 2000).

We also compute the median over all profiles as

$$\tilde{G} = \text{median}(G_i), \tag{9}$$

and we compute two ranges directly from the distributions of G_i . The interquartile range S_{50} contains 50% of the observed values of G :

$$S_{50} = G_{0.75N} - G_{0.25N}, \tag{10}$$

and S_{95H} is the half-width of the range of the distribution that contains 95% of the observed values of G_i :

$$S_{95} = G_{0.975N} - G_{0.025N} \tag{11}$$

$$S_{95H} = \frac{S_{95}}{2}. \tag{12}$$

In both of these definitions G_i is sorted in ascending order and the subscripts are rounded to the nearest integer. For S_{50} , 25% of the observations are smaller than the minimum value that defines the range and 25% of the observations are larger than the maximum value that defines the range. For S_{95} , 2.5% of the observations are smaller than the minimum value that defines the range and 2.5% of the observations are larger than the maximum value that defines the range. In a perfect sample of a Gaussian distribution, $2\sigma_G$ and S_{95H} would be nearly equal.

We compute the mean absolute relative difference (MARD) as

$$\text{MARD} = \frac{1}{N} \sum_{i=1}^N \left| \frac{R_{fi} - R_{si}}{R_{si}} \right| = \frac{1}{N} \sum_{i=1}^N |G_i - 1| \tag{13}$$

and the median absolute relative difference (EARD) as

$$\text{EARD} = \text{median}(|G_i - 1|), \tag{14}$$

where R_{fi} and R_{si} are individual estimates of R_{rs} by the floats and satellites, respectively. We also make direct comparisons of float and satellite R_{rs} values by using a reduced major axis regression to estimate coefficients in the model

$$R_{fi} = a_0 + a_1 R_{si}, \tag{15}$$

computing the root-mean-square difference (RMSD),

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (R_{fi} - R_{si})^2}, \tag{16}$$

and the squared Pearson’s correlation coefficient, r^2 . For reference, we compare the dimensional statistics to the mean of the satellite estimates of R_{rs} ,

$$\bar{R}_{si} = \frac{1}{N} \sum_{i=1}^N R_{si}. \tag{17}$$

To examine the variability within individual profiles, we compute normalized differences of a generic quantity from a reference or modeled value as

$$f_A = \frac{A_x - A_r}{A_r}, \tag{18}$$

where A is a quantity of interest, f_A is the normalized difference, the subscript r refers to a reference or modeled value, and the subscript x refers to a value that is measured or computed. The quantities used for A_x and A_r are made clear for each specific usage. Because A_r is constant, many statistics computed using f_A are equivalent to statistics computed using A_x alone and then

TABLE 3. MODIS *Aqua* statistics. Statistics are defined in section 2g. values of a_0 , RMSD, and $\overline{R_{si}}$ have been multiplied by 100 for readability. Units of λ are nm, and units of a_0 , RMSD, and $\overline{R_{si}}$ are sr^{-1} .

λ	\overline{G}	\tilde{G}	$\frac{2\sigma_G}{\sqrt{N}}$	σ_G	S_{95H}	kurtosis	S_{50}	MAPD	EAPD	r^2	a_1	$a_0 \times 100$	RMSD $\times 100$	$\overline{R_{si}} \times 100$
412	1.033	1.006	0.034	0.138	0.251	7.481	0.135	0.095	0.057	0.829	0.940	0.086	0.116	1.134
443	0.940	0.936	0.025	0.103	0.218	4.346	0.113	0.099	0.099	0.784	0.819	0.106	0.112	0.955
488	0.993	0.982	0.018	0.072	0.150	3.895	0.076	0.056	0.052	0.739	0.794	0.124	0.043	0.641
555	1.123	1.085	0.045	0.180	0.275	9.989	0.155	0.140	0.089	0.013	0.063	0.151	0.024	0.146

taking differences from and/or normalizing by A_r . If A_r represents the mean of a population of individual measurements, A_x , then the coefficient of variation of the population is the standard deviation of f_A :

$$\text{CV}_A = \text{std}(f_A). \quad (19)$$

h. Quality control of float measurements

Quality control was performed for the float observations using $L_u(z_b)$ and the estimates of K_L and \widehat{L}_u in each of the 3-m bins. For each criterion we examined all wavelengths. If the criterion failed for any wavelength, then the profile was eliminated from the dataset.

Quality control on K_L required

$$K_L > 0, \quad (20)$$

$$K_L < 0.2 \text{ m}^{-1}, \quad (21)$$

$$\left| \frac{K_{L4} - K_{L3}}{0.5(K_{L4} + K_{L3})} \right| < \frac{2}{3}, \quad (22)$$

where K_{L4} is K_L in the top (fourth) bin, centered at a nominal depth of -3 m, and K_{L3} is K_L in the next bin from the surface (third bin), centered at a nominal depth of -6 m. The first two criteria [(20), (21)] were applied to all wavelengths and all bins. If any value of K_L violated the criterion, then the profile was discarded. The third criterion [(22)] assesses variability in estimates of K_L . The threshold for (21) was chosen following Morel and Maritorena (2001) with a chlorophyll concentration of 2 mg m^{-3} (much larger than those observed in this study). The threshold for (22) was chosen for consistency with observed variability in K_L estimates.

For upwelling radiance we required that estimates of L_u increase vertically such that

$$L_u(z_b) > \widehat{L}_{u4} > \widehat{L}_{u3} > \widehat{L}_{u2} > \widehat{L}_{u1}, \quad (23)$$

where the subscripts refer to bins centered at -3 , -6 , -9 , and -12 m, respectively. Because each bin is evaluated separately, this criterion is distinct from (20).

We also required minimal variability of observations of L_u during ascent, quantified as

$$\langle \text{CV}_{L_u} \rangle < 0.05, \quad (24)$$

where CV_{L_u} is defined using (19) and (18), and the angle brackets denote a mean. For (24), $L_u(z)$ refers to the noisy observations and the reference values, $\widehat{L}_u(z)$, are best-fit estimates at the observation depths. The mean is over all wavelengths and all bins.

Finally, we required that the projection of \widehat{L}_u from the top bin to z_b using (1) be within 10% of the measured value of $L_u(z_b)$, expressed as

$$\left| \frac{L_u(z_b) - \widehat{L}_u(z_b)}{L_u(z_b)} \right| < 0.1, \quad (25)$$

where $\widehat{L}_u(z_b)$ is the extrapolation from the top bin. For this comparison, $L_u(z_b)$ was not corrected for shading, but it was restricted to observations from only the sunny side of the float. This criterion is included as a holistic test of the consistency of the radiance estimates and the accuracy of the K_L estimate.

3. Results

Of our 1181 profiles, 1088 occurred within 3 hours of an overpass of the *Aqua* satellite. Of these 1088 matchups, 126 had satellite data that passed satellite quality control (QC) criteria, and 230 had in situ data that passed float QC criteria. Only 65 profiles (6.0% of matchups) passed both satellite and float QC evaluations and were useful for validation against MODIS *Aqua* (Table 3). For VIIRS, 439 profiles occurred within 3 hours of an overpass of *Suomi-NPP*. Of these matchups, 42 passed satellite QC, 113 passed float QC, and only 15 (3.4%) passed both sets of QC criteria and were useful for validation against VIIRS (Table 4). For comparison, out of 1450 MOBY observations, Franz et al. (2007) reported 150 matchups ($\sim 10\%$) that passed the more stringent quality control criteria used for vicarious calibration.

We examined in detail the float quality control criteria that failed most commonly at times of good MODIS *Aqua* observations. Of the 126 profiles with good MODIS *Aqua* observations, 61 profiles failed float quality control. Eleven of these profiles had zero samples because of

TABLE 4. VIIRS statistics. Symbols as in Table 3.

λ	\bar{G}	\tilde{G}	$\frac{2\sigma_G}{\sqrt{N}}$	σ_G	S_{95H}	Kurtosis	S_{50}	MAPD	EAPD	r^2	a_1	$a_0 \times 100$	RMSD $\times 100$	$\bar{R}_i \times 100$
412	1.030	1.017	0.054	0.105	0.215	5.353	0.098	0.077	0.063	0.741	1.207	-0.200	0.097	1.127
443	0.990	0.990	0.056	0.108	0.216	4.772	0.091	0.074	0.042	0.699	1.063	-0.071	0.081	0.926
488	0.986	0.979	0.041	0.080	0.149	4.169	0.088	0.060	0.047	0.593	0.866	0.077	0.047	0.662
555	0.994	0.968	0.059	0.115	0.195	3.234	0.157	0.089	0.071	0.009	-0.062	0.172	0.018	0.164

technical problems or zero samples with low-tilt and unshaded orientations. Of the remaining 50 failed profiles, 20 failed only one QC criterion, 18 failed two QC criteria, and the remaining 12 profiles failed 3–6 criteria. The most commonly failed criterion was the requirement of minimal difference in K_L between the top two bins [(22)]. This accounted for 37 rejected profiles, of which 12 passed all other criteria. The next most common failure criterion was the requirement that $L_u(z_b)$ be within 10% of the radiance projected upward from the top bin [(25)], which accounted for 21 rejected profiles, of which 5 passed all other criteria. The requirement that $L_u(z_b)$ be larger than the top bin estimate of L_u [first inequality of (23)] accounted for 10 rejected profiles and for the remaining 3 unique rejections. For this dataset (but perhaps not for future datasets), if we had used only these three quality

control criteria [(22), (25), and first inequality of (23)], we would have arrived at the same final dataset as we did with the full set of criteria.

The observed G_i are relatively consistent in time (Fig. 4), although there is an increase in magnitude and variability of G_i about one year after float deployments, particularly at 412 and 555 nm. The distributions of G_i for both satellites at all four wavelengths have Gaussian-like properties (Fig. 5). For all sets of measurements except the comparison to MODIS *Aqua* at 555 nm, Kolmogorov–Smirnov tests fail to reject the hypothesis that the G_i have Gaussian distributions. The standard deviation gives a reasonable estimate of the range of observations, with $2\sigma_G$ being similar to S_{95H} at all but 555 nm.

Statistics for G are given in Tables 3 and 4, but we note that the sample sizes are small, particularly for VIIRS

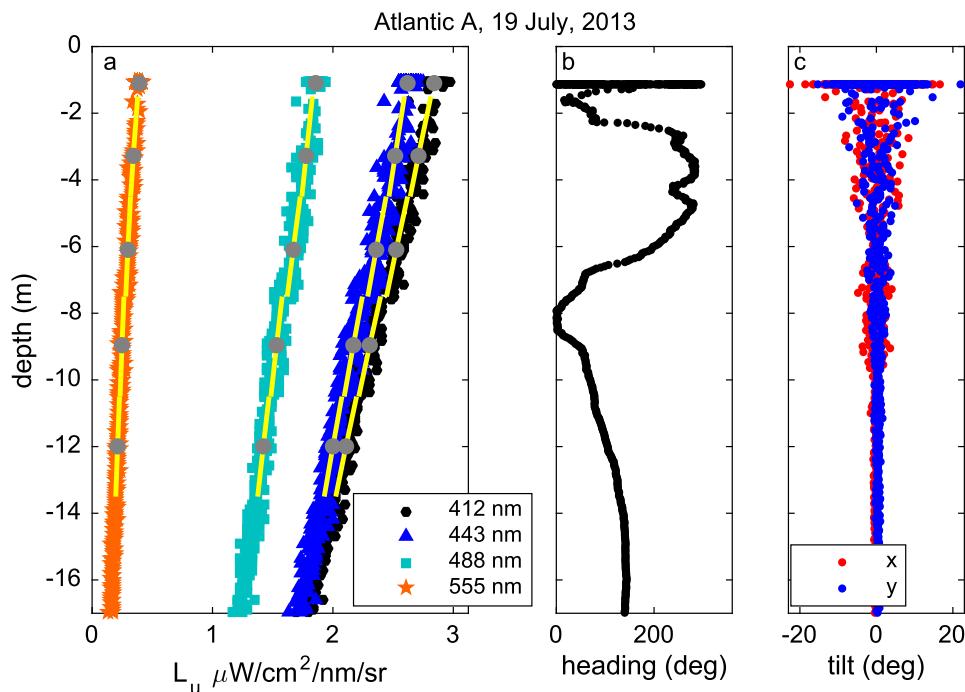


FIG. 3. Example of typical shallow observations from a profile of the Atlantic A float. (a) Upwelling radiance; color symbols show individual observations in different wavelengths. Yellow lines show results of least squares fits of (1) to observations in each bin, and gray dots show the best-fit estimate of radiance $\hat{L}_u(z_m)$ in each bin. Bin boundaries are shown by the breaks between the line segments. (b) Heading (clockwise from north) during ascent. (c) Tilt in the x and y directions during ascent. For tilt and heading, all values are shown. For L_u only values with tilt less than 5° are shown.

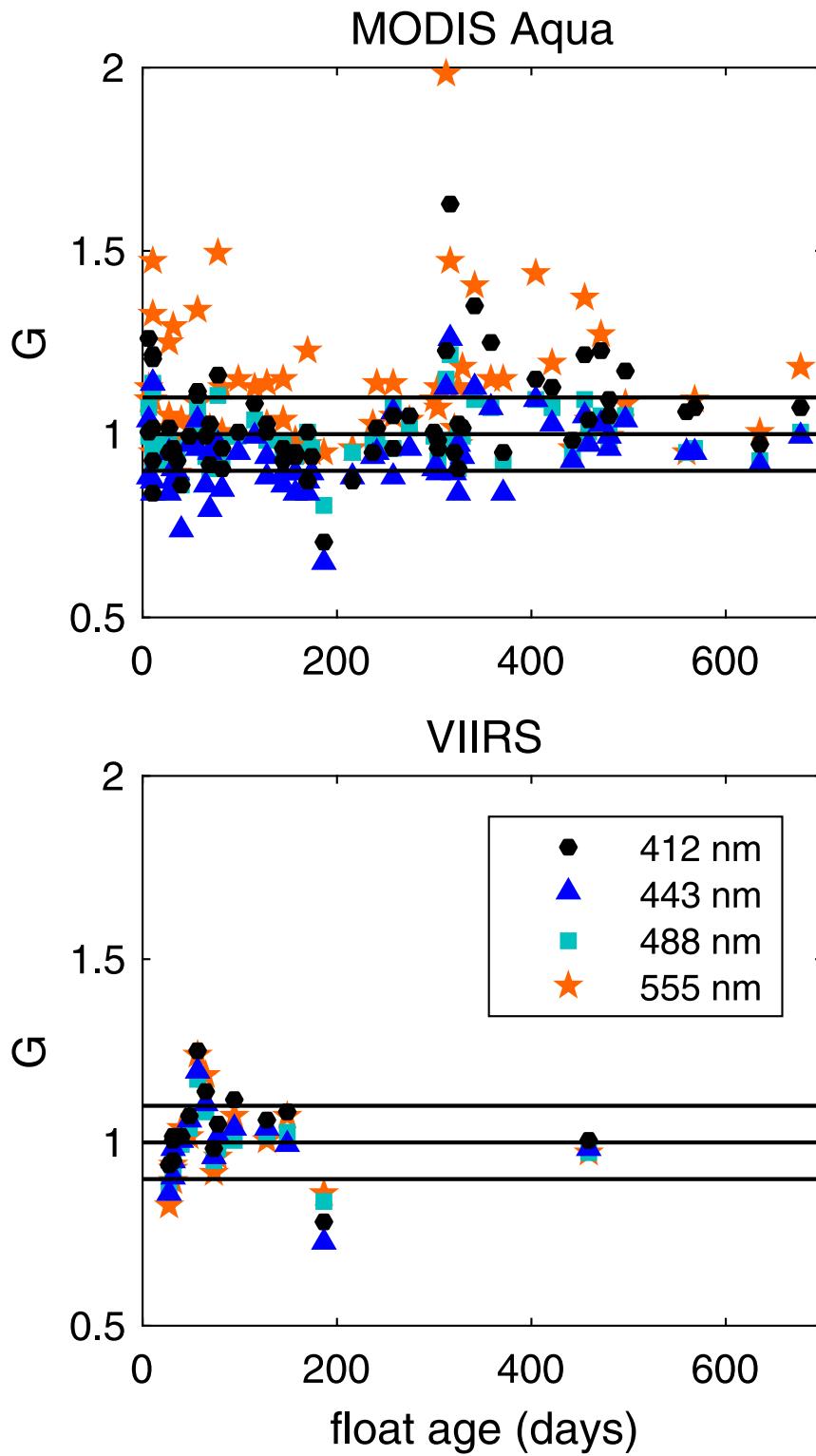


FIG. 4. The G_i vs time since float deployment. Symbols (shapes and colors) denote different wavelengths. Horizontal lines show constant values of G equal to 0.9, 1, and 1.1.

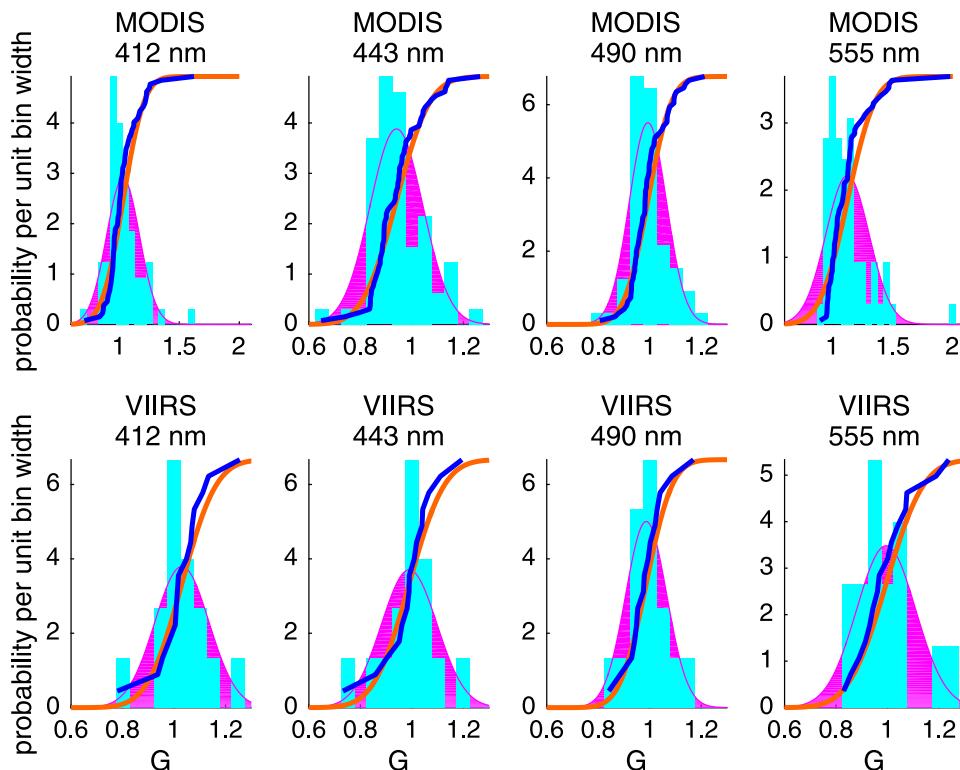


FIG. 5. Distribution of G_i , compared to a Gaussian distribution. Cyan histogram shows relative abundance of observed G_i , normalized so that its integral is 1 (bin width: 0.05). Magenta background and curve show Gaussian probability density function with the same mean and standard deviation as observations. Blue line shows cumulative distribution of observations. Orange line shows cumulative distribution of a Gaussian. The cumulative distribution functions have been rescaled by the axis height. Note different x-axis scale for MODIS 412- and 555-nm bands.

matchups. Median and mean G are similar, differing by only a few percent in the worst cases. Because of the larger number of good matchups, the standard errors of \bar{G} estimates is smaller for MODIS *Aqua* than for VIIRS, even in cases of larger σ_G . In two cases (443 and 555 nm for MODIS *Aqua*), differences of \bar{G} from 1 are statistically significant, suggesting the possibility of biased errors in either the float or satellite estimates of R_{rs} .

The G_i for each float are similar and show similar variability (Fig. 6). Only three floats were operational for more than one year, and one of those (Hawaii A) had limited observations because a malfunction in the E_d sensor affected sampling of the remaining optical sensors. The other floats with long records, Hawaii B and Atlantic A, show the elevated G_i that are apparent in Fig. 4. In both floats the G_i are high in the first third of the year, but the elevation occurs earlier in Hawaii B than in Atlantic A. For Hawaii B and Atlantic A, the largest G_i occur at solar zenith angles between about 30° and 50° , but the relationship between G_i and zenith angle is not consistent across all floats (Fig. 7). For Atlantic B there is little variation in G_i with zenith angle,

but that float did not have observations during the times when the most variation in Atlantic A occurred. The elevated G_i occur in all wavelengths but are most noticeable in the 412- and 555-nm bands (Fig. 8). Variability in G_i does not appear to be well correlated with deployment region or diffuse attenuation coefficient.

In addition to G_i , we also made direct comparisons of R_{rs} (Fig. 9). For the three shorter wavelengths, we find similar variability across the range of observations of R_{rs} , and the comparison gives moderate to high r^2 (Tables 3 and 4). For 555 nm, however, we find that the range of variability of the in situ R_{rs} estimates is much smaller than the range of variability of the satellite estimates. Values of r^2 for 555 nm are correspondingly low.

4. Analysis of method and uncertainty

Uncertainty in R_{rs} estimates is introduced from many sources. Some uncertainties have direct effects on estimates of K_L , and some have direct effects on estimates of $L_u(z_b)$. Both of these affect estimates of L_w . Additional uncertainty is introduced in the computation of

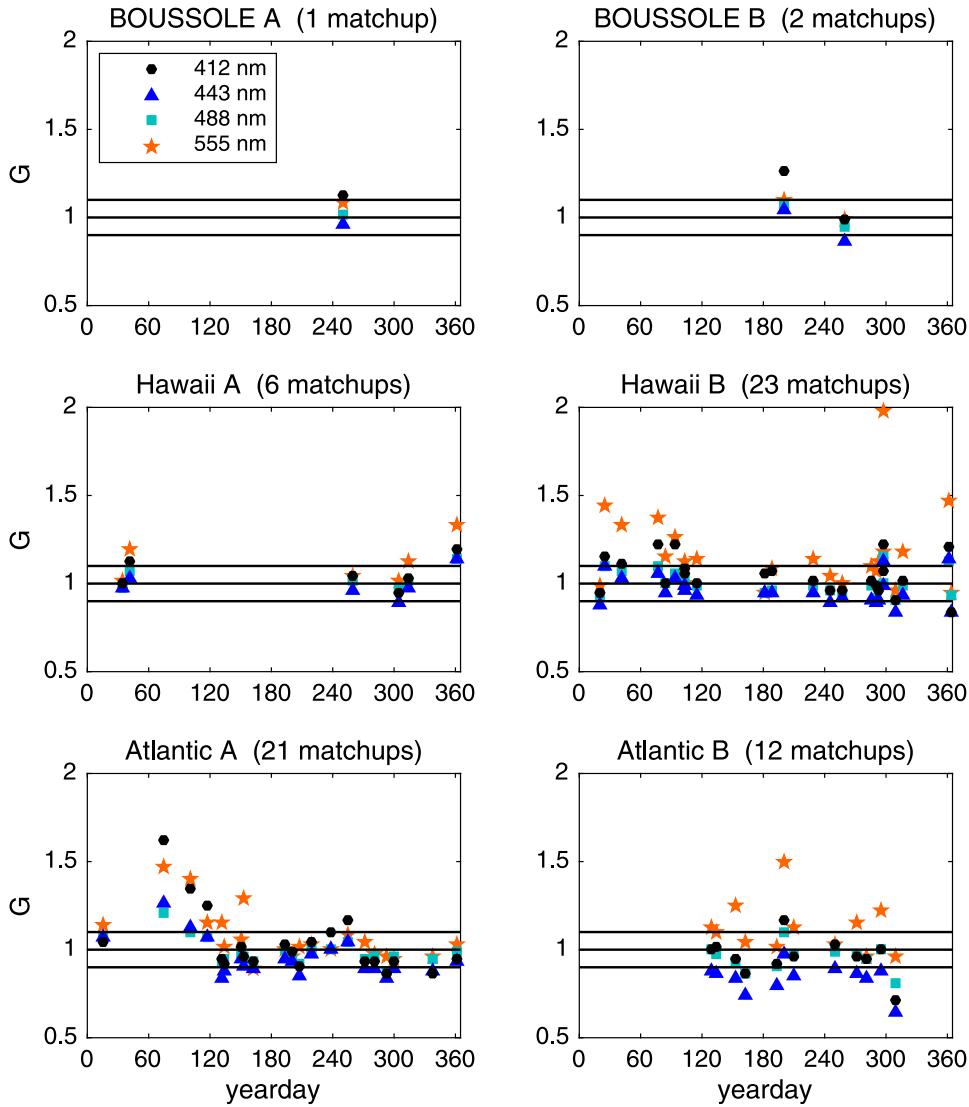


FIG. 6. The G_i vs yearday are shown separately for each float. MODIS *Aqua* only. Yearday 1.5 is noon on 1 Jan, and yeardays are repeated for floats that were deployed for more than one year. Symbols denote different wavelengths as in Fig. 4. Lines show values of 0.9, 1, and 1.1.

R_{FS} from L_w and E_s . In the following section, we discuss sensor calibration (briefly), individual sources of uncertainty, and possible direct effects of these uncertainties on quantities derived from each measurement. We then discuss a Monte Carlo model that incorporates most of the likely uncertainty sources to determine an integrated uncertainty in the estimates of L_w .

The sensors used in this study respond to radiance according to

$$L_{uCAL} = c_1(L_{uRAW} - c_0), \quad (26)$$

where L_{uCAL} is a calibrated value in physical units and L_{uRAW} is a value in uncalibrated instrument counts. The

calibration constants c_1 and c_0 are determined prior to deployment. Using the c_0 determined at the factory or a value of c_0 determined from measurements at the park depth of 1000 m led to differences in near-surface L_{uCAL} typically less than 0.05%. This suggests that the near-surface radiances are large enough that values of L_{uCAL} are relatively insensitive to uncertainty in c_0 . Previous assessment of the radiometer calibration procedure determined that calibration uncertainty in radiance estimates is 2%–3% (Hooker et al. 2002; Voss et al. 2010).

a. Uncertainty in K_L

If K_L is constant with depth, then estimates of K_L are independent of the calibration coefficient c_1 . This

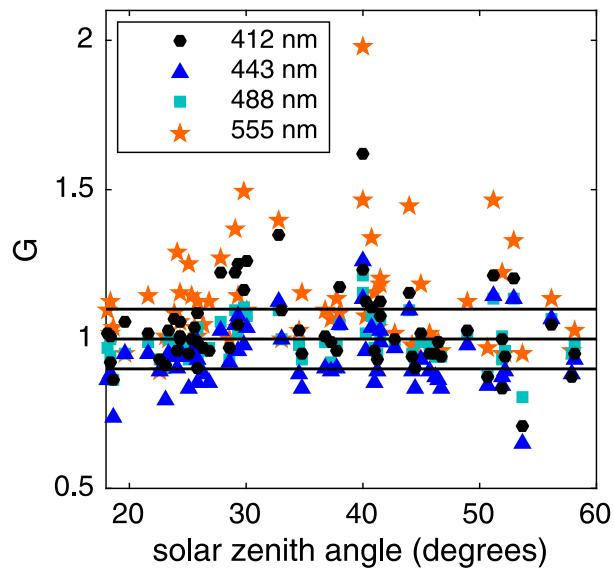


FIG. 7. The G_i vs float solar zenith angle. MODIS *Aqua* only. Lines show values of 0.9, 1, and 1.1.

is because K_L is estimated from (1) using relative magnitudes of L_u . Changes in the c_1 due to electronic variation, biofouling, or other sources will not affect estimates of K_L . Here we examine the effect of sample resolution, shading, and bin size on K_L estimates.

Zibordi et al. (2004) examined the effects of sample resolution on estimates of several radiometric quantities, including L_u , downwelling irradiance (E_d), and diffuse attenuation coefficient for downwelling radiance (K_d). Several environmental, sampling, and analysis differences cause challenges in extrapolating uncertainties from their results to our measurements, but their work offers some guidance. In the coastal waters studied by Zibordi et al. (2004), spatial resolution of one sample every 4–6 cm is more than sufficient to give accuracies of bin estimates of L_u within $\sim 1\%$. Although Zibordi et al. (2004) do not give results for accuracy of K_L estimates, they found that the relative uncertainty in K_d was roughly twice as large as the uncertainty in E_d . Extrapolating this proportionality to K_L and L_u suggests that for samples spaced every 4–6 cm, uncertainty in estimates of K_L may be of order 2% in coastal waters. Given higher uncertainty in clear water, we expect that the uncertainty in K_L related to the spatial resolution of our sampling is on the order of 2%–5%. As discussed below, overall uncertainty in K_L estimates is likely larger.

Shading varies as the float heading changes during ascent. To quantify the effects of shading on the estimates of K_L , we compare estimates made using all data during ascent to estimates made using only data when the radiometer was on the same side of the float as the sun. We find that estimates of K_L are similar from both

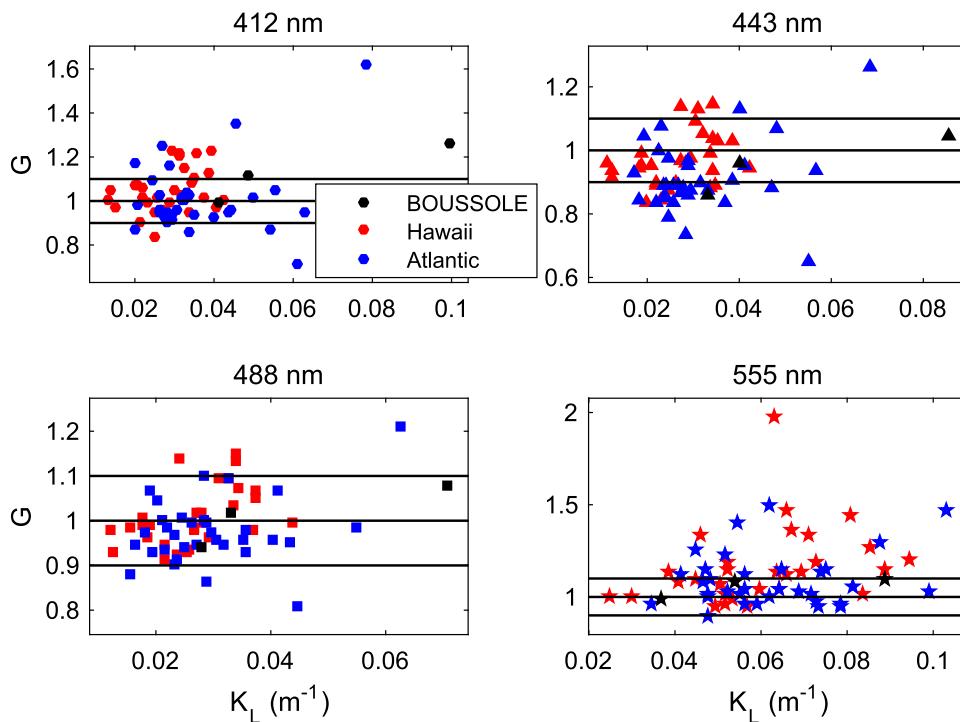


FIG. 8. The G_i vs K_L are shown separately for each wavelength. MODIS *Aqua* only. Symbol colors denote float deployment region. Lines show values of 0.9, 1, and 1.1.

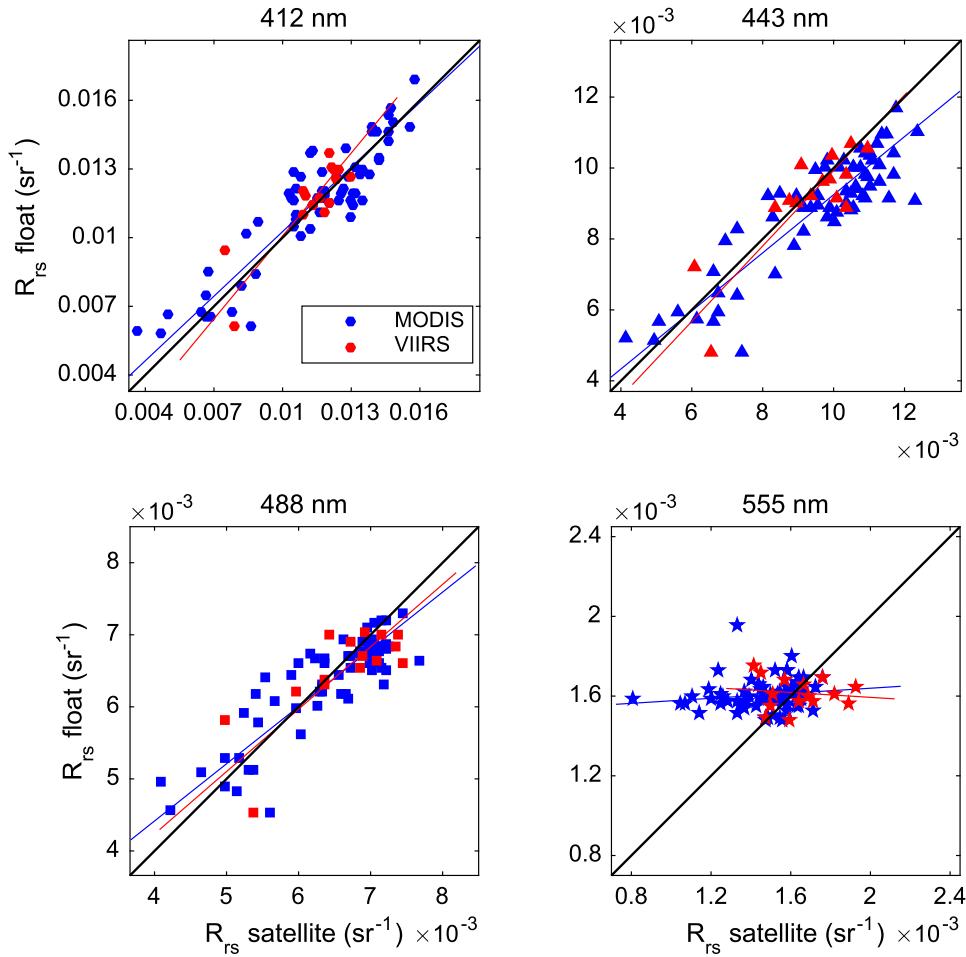


FIG. 9. Scatterplots of R_{rs} . Vertical axes are float estimates. Horizontal axes are satellite estimates. Blue is MODIS *Aqua*. Red is VIIRS. Black lines are 1:1, and thin colored lines show the least squares major axis fits to the observations.

sets of allowed azimuths (Fig. 10), with variability in their agreement. We computed $\sigma_{f_{KL}}$ and $\langle f_{KL} \rangle$ using the measurements from the sunny side of the float as the reference quantity, with the angle brackets denoting a mean over all profiles. We found $\sigma_{f_{KL}}$ of 23%–30% and $\langle f_{KL} \rangle$ of negative 4%–6%. Standard errors are all of order 4%–5%, so the bias introduced by using all azimuths rather than sunny-side-only measurements is likely not statistically significant. Using measurements only from the sunny side reduced the number of profiles passing quality control by about 45%.

To understand some the effects caused by our choices of bin size, we performed our analysis using different values of bin size. We used only profiles that passed the quality control checks for both the float and satellite data, and we quantify the variability in K_L estimates by examining f_{KL} , defined using (18). All K_L values in this analysis are for the top bin, and we omitted the Hawaii

A float because of its sparse data. Here the statistics $\sigma_{f_{KL}}$ and $\langle f_{KL} \rangle$ compare bin sizes of 2, 4, and 5 m to reference values of 3-m bins. These bin sizes are all smaller than vertical attenuation lengths $1/K_L$, which for most profiles are larger than 10 m (and generally larger than 20 m for the shorter wavelengths). Variability in K_L estimates using different bin sizes is larger than for the shading comparison above, with $\sigma_{f_{KL}}$ of order 20%–30% (Fig. 11). Means of normalized differences over all profiles, $\langle f_{KL} \rangle$, are positive for all bin sizes compared to the 3-m bins. For the 4- and 5-m bins, $\langle f_{KL} \rangle$ is slightly larger than twice the standard error, suggesting that the choice of bin size may cause a bias in estimates of K_L .

The values of $\sigma_{f_{KL}}$ for the shading and bin size estimates suggest that the confidence interval for our K_L estimates is somewhat large. To determine the effect of error in K_L on estimates of L_w , we used (2) to project L_u upward to the sea surface from -1.12 -m depth. For

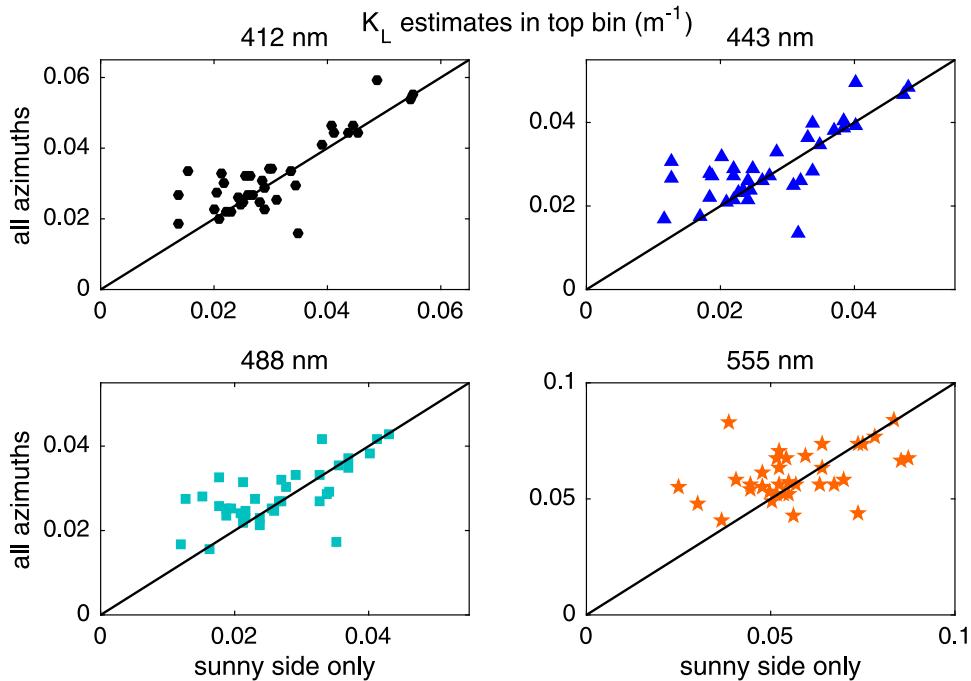


FIG. 10. Comparison of estimates of diffuse attenuation coefficient in top bin using measurements from all azimuths (vertical axes) and measurements only when the radiometer was on the sunny side of the float. Black line is 1:1.

this analysis we use three representative values of K_L ($0.03, 0.05,$ and 0.1 m^{-1}) and vary them by $\pm 5\%$, 10% , 20% , and 40% . For an arbitrary $L_w(z_b)$, we compute a simulated “true” value of L_w with the representative values of K_L and compare it to “error” values of L_w that are computed with the values of K_L varied by $\pm 5\%$, 10% , 20% , and 40% . The ratios $L_{w,error}/L_{w,true}$ are displayed in

Table 5. We find that errors in L_w are generally smaller than 2.5% , except in the cases of large error in K_L or large values of K_L itself. This suggests that for visible wavelengths in most open ocean conditions ($K_L \ll 1/|z_b|$), highly accurate measurements of K_L are not essential for reasonable accuracy of L_w estimates, as long as the errors in K_L are random. Error in K_L could be a larger source of

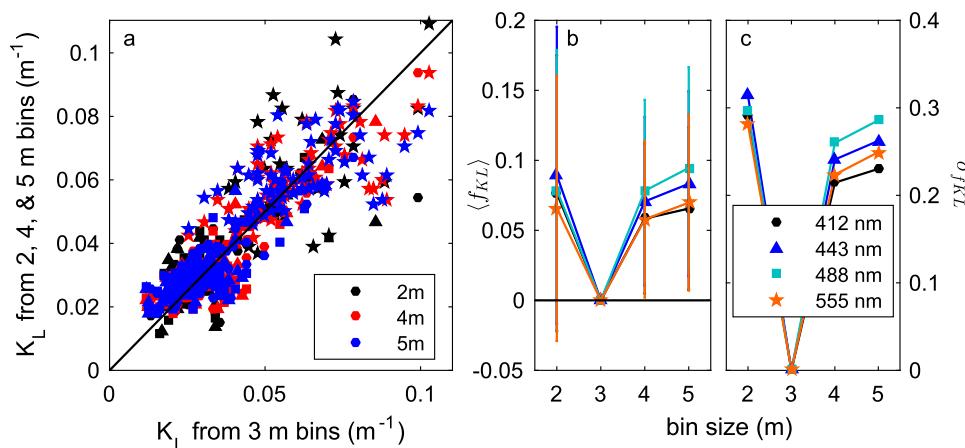


FIG. 11. Comparisons of K_L estimates using different bin sizes. (a) Term K_L computed using alternative bin sizes compared to K_L computed using 3-m bins. Black line is 1:1. (b) Mean of f_{KL} , indicating the relative bias. Error bars are twice the standard error of f_{KL} . (c) Coefficient of variation of K_L estimates (standard deviation of f_{KL}). Symbol shapes denote wavelength as in previous figures. In (a), colors indicate the bin size. In the other panels colors indicate wavelength as in previous figures.

TABLE 5. Errors in L_w caused by inaccurate K_L . Left column gives “true” K_L . Other columns give $L_{w,error}/L_{w,true}$ computed using inaccurate values K_L , with the error given in the top row.

K_L (m^{-1})	−40%	−20%	−10%	−5%	+5%	+10%	+20%	+40%
0.030	0.987	0.993	0.997	0.998	1.002	1.003	1.007	1.014
0.050	0.978	0.989	0.994	0.997	1.003	1.006	1.011	1.023
0.100	0.956	0.978	0.989	0.994	1.006	1.011	1.023	1.046

error in L_w for wavelengths at which K_L approaches $1/|z_b|$. Any bias in K_L estimates will lead to bias in L_w estimates, including possible bias from Raman scattering, which is likely more important at the longer wavelengths.

b. Uncertainty in $L_u(z_b)$

Measurements of $L_u(z_b)$ during the buoy phase of the profile are sensitive to the calibration constant c_1 [(26)] and any effects of biofouling on this value. The measurements are also sensitive to environmental conditions that include the effects of wave focusing, sensor tilt, and self-shading. Following (2) and (3), errors in $L_u(z_b)$ lead to errors in L_w of the same relative magnitude.

Because we did not retrieve the floats after deployment, we have no way to directly assess changes in c_1 over time, either from changes in the sensor or from biofouling. The G_i estimates appear stable through time (Fig. 4), but their variability is large enough that we cannot use the stability of G_i to determine with certainty whether c_1 meet the requirements for long-term stability for vicarious calibration (Mueller et al. 2003a; GCOS 2011; Zibordi et al. 2015).

Based on the work of Stramska and Dickey (1998) and Wei et al. (2014), we expect that wave-induced light fluctuations at our measurement depth of $z_b = -1.12$ m occur with a peak frequency of 0.3–1 Hz or higher. In addition, the effect of light flashes on L_u is likely to be smaller than the effect of flashes on the more commonly studied E_d (Stramska and Dickey 1998; Zibordi et al. 2004; D’Alimonte et al. 2010). Because our measurements integrate 0.933 s out of every second, they likely capture the most light flashes associated with wave focusing. This likely allows accurate $L_u(z_b)$ estimates with a relatively small number of samples.

Our estimates of $L_u(z_b)$ are, unfortunately, limited by small sample sizes. In 60 out of 65 profiles, the float made 10 or more low-tilt measurements in its buoy phase, but only 32 of the profiles had 10 or more buoy phase measurements at both low-tilt and favorable azimuthal orientation. The BOUSSOLE profiles had longer surface intervals than the other floats, which allowed us to characterize some of the variability in the near-surface samples, and we examine nine profiles from these floats (with relaxed QC criteria that allow more variability in the ascent phase data). The buoy phases of these profiles had between 30 and 361 samples over 5–10 min. Coefficients

of variation for the buoy phase radiance measurements in each profile are between 1% and 6%. We estimated $L_u(z_b)$ using subsets of measurements from these BOUSSOLE profiles. For subsets with 10 and 5 measurements, 92% and 88% of the subsets, respectively, had means of $L_u(z_b)$ within 2% of the estimates from the full surface dataset for the given profile.

To analyze the uncertainties associated with our shading correction, we compared our shading correction with results from Simulation Optique (SimulO), an IOP-driven Monte Carlo simulation (Leymarie et al. 2010; <http://omtab.obs-vlfr.fr/SimulO>). We ran SimulO using a chlorophyll-based bio-optical model (Morel and Maritorena 2001) and a representative open ocean chlorophyll concentration of 0.1 mg m^{-3} and no contribution from skylight. For the solar zenith angles in our study and relative solar azimuths of $\pm 120^\circ$ (where 0° is defined as the sun being on the same side of the float as the radiometer), we found that the two shading models had similar shapes and magnitudes, with mean shading corrections of $<1\%$ (Fig. 12). However, the heading uncertainty leads to uncertainty in the corrected L_u of a few percent in the worst cases (large relative azimuth and small solar zenith angle). At 412- and 555-nm wavelengths, the Leathers et al. (2004) model predicted a shading larger on average than SimulO by less than 0.1%, although for larger chlorophyll concentrations this possible bias increases (not shown).

Taken together, these analyses suggest that the overall uncertainty in $L_u(z_b)$ estimates due to environmental and measurement variability is on the order of 2%–3%, similar to the estimate by Zibordi and Voss (2014) for generic in-water measurements. Uncertainty in the calibration (also likely 2%–3%; Hooker et al. 2002; Voss et al. 2010) will raise the total uncertainty of $L_u(z_b)$ to roughly 4%.

c. Other sources of uncertainty in R_{rs}

Our estimates of R_{rs} have other sources of uncertainty. Here we examine the possible effects of measuring or modeling downwelling irradiance and the bidirectional reflectance correction.

1) DOWNWELLING IRRADIANCE

By using model estimates of E_s , we eliminate one source of potential discrepancy between float and

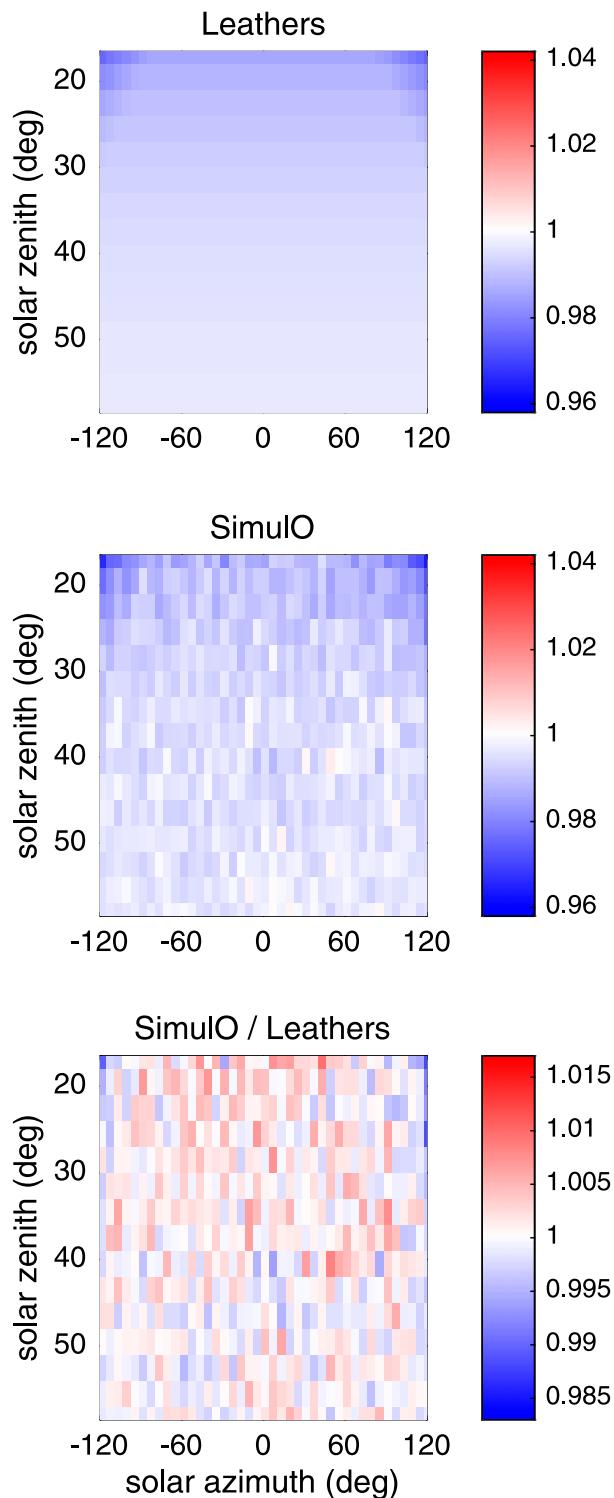


FIG. 12. Comparisons of Leathers and Simulo shading corrections at 555 nm, $[CHL] = 0.1 \text{ mg m}^{-3}$, only for cases where the optics package is on the sunny side of the float. (top),(middle) The colors are the ratio $L_{u,\text{shaded}}/L_{u,\text{actual}}$, where $L_{u,\text{shaded}}$ is computed with either the Leathers or Simulo model. Shading should produce values smaller than 1. Much of the variability in Simulo values, including values larger than 1, are likely the result of Monte Carlo noise and suggest uncertainty of 1% or less in individual shading estimates from Simulo. (bottom) The ratio of the top two panels.

satellite R_{rs} estimates. However, the model we used for E_s (Frouin et al. 1989) is not exactly the same as that used in the satellite atmospheric correction algorithm (Gordon and Wang 1994; Franz et al. 2007). Both models are based on assumptions about atmospheric properties, but the satellite processing uses near-infrared measurements to determine atmospheric aerosol composition, whereas the Frouin et al. (1989) model relies on modeled aerosol composition. A comparison of E_s predictions suggests that using these two different atmospheric models has small but nonzero effects on estimates of G_i . We compare them using the ratio R_E of the satellite estimate to the Frouin estimate of E_s ,

$$R_E = \frac{E_s(\text{satellite})}{E_s(\text{Frouin})}, \quad (27)$$

for the 65 observations that passed satellite and float quality control (Fig. 13c). For each wavelength R_E has both means and standard deviations of $\sim 1\%$ – 2% . The means of R_E are further than two standard errors from a value of one, suggesting that there may be a small but statistically significant bias in the Frouin estimates compared to the satellite estimates.

2) BIDIRECTIONAL REFLECTANCE CORRECTION

The Morel et al. (2002) normalization uses viewing geometry and chlorophyll concentration ($[Chl]$) as inputs. Assuming that the viewing geometries can be exactly calculated, uncertainties associated with this normalization emerge from the estimation of $[Chl]$, which is determined using a reflectance band ratio (O'Reilly et al. 1998, 2000). We estimated uncertainty in the correction factor F by estimating variable values of F using a simplified Monte Carlo simulation and the in situ geometry. For each profile we introduced variability into the $[Chl]$ input and computed F with these variables $[Chl]$. Repeated iterations gave standard deviations in F that converged to less than 1.5% for most profiles at all wavelengths. We chose distributions of $[Chl]$ that varied around each profile's estimate with a median absolute percent difference of about 26% as in Bailey and Werdell (2006). This empirically determined value for variability accounts in an integrated way for uncertainty in R_{rs} and uncertainty in determining $[Chl]$ from R_{rs} .

d. Global uncertainties using Monte Carlo simulations

We examine the combined effects of measurement variability on our estimates of L_w using Monte Carlo simulations of float measurements during the ascent and buoy phases. The simulations begin with a noise-free

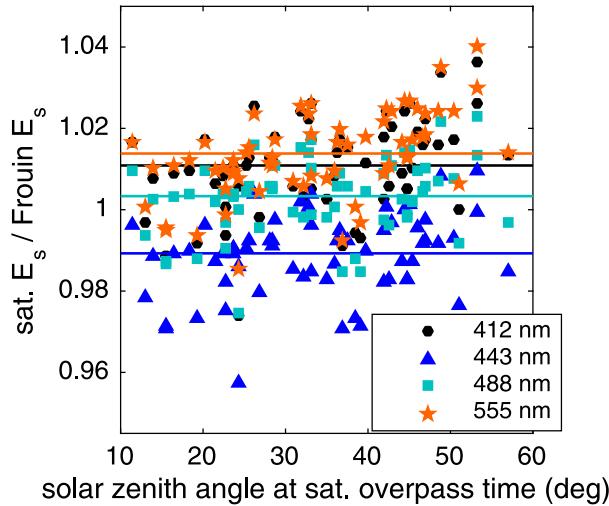


FIG. 13. Comparison of E_s computed during satellite data processing to the Frouin model of E_s for the same times and locations. Lines show mean ratios.

reference profile $L_{u,\text{true}}$ that represents the truth. In 5000 iterations we construct noisy profiles that represent observations. To construct the noisy profiles, we subsample the reference profile with 5-cm resolution during the ascent phase, use 10 samples in the buoy phase, and add several sources of variability. These include error in the observation depth (during the ascent phase), possible bias in the shading correction (during buoy phase), effects of tilting, effects of wave focusing, variability in sensor response, and unbiased errors in the shading correction. Then, using the same methods as are described in section 2, we compute L_w for each noisy profile and compare it to the water-leaving radiance for the reference profile, $L_{w,\text{true}}$. The reference profile uses parameters that are representative of values in this study: $L_{w,\text{true}} = 1 \text{ mW cm}^{-2} \text{ nm}^{-1} \text{ sr}^{-1}$ and diffuse attenuation coefficient $K_{L,\text{true}} = 0.03 \text{ m}^{-1}$. The simulated radiance measurements had Gaussian variability about the profile of $L_{u,\text{true}}$, and a coefficient of variation of $\text{CV}_{L_u} = 0.04$ (our observations of CV_{L_u} were generally between 0.015 and 0.05).

The results of the Monte Carlo simulations show that unbiased uncertainty in the observations gives unbiased estimates of L_w (Fig. 14) and $L_u(z_b)$ (not shown). As might be expected, bias in the shading correction leads to bias in radiance estimates of the same relative size. We found the normalized standard deviation of the $L_u(z_b)$ estimates, $\sigma_{L_u}/L_u(z_b)$, to be $\sim 1.4\%$, consistent with the 2%–3% uncertainty due to environmental/measurement variation estimated from the observations (using a 2σ confidence interval). The variability of simulated L_w is larger than the variability in simulated

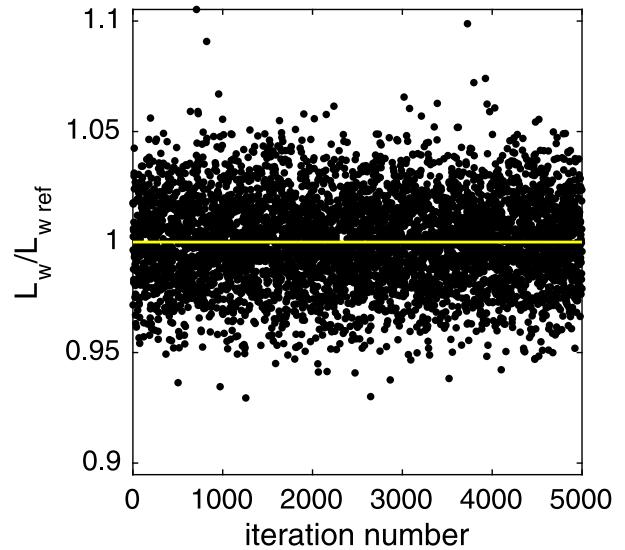


FIG. 14. Estimates of L_w normalized by reference value from Monte Carlo simulations. Yellow line shows a reference value of 1.

$L_u(z_b)$ (as expected), but smaller than the variability of observational estimates of G , with $\sigma_{L_w}/\overline{L_w}$, of $\sim 2\%$ (Monte Carlo simulations) and σ_G between 7% and 19% (observations). Increasing the variability of the noisy profiles in the simulations does have some effect on the simulated output, but not enough to explain the discrepancy between σ_{L_w} and σ_G . Increasing the noise in the profiles by a factor of 3–4 increases the variability in estimates of L_w by a similar amount. Increasing the noise in the buoy phase measurement has a larger effect than increasing the noise in the ascent phase measurement because the ascent data are used only in estimating K_L . These results suggest one or more of the following: 1) uncertainty in observed L_w is larger than predicted by the Monte Carlo simulations [likely due to uncertainty in extrapolation rather than $L_u(z_b)$]; 2) radiometer calibration uncertainty is larger than 2%–3%; 3) uncertainty in converting observed L_w to R_{rs} is not negligible (due to estimating E_s or the bidirectional reflectance correction); and 4) uncertainty in the satellite estimates of R_{rs} contributes to uncertainty in estimates of G .

Because the Monte Carlo simulations compute estimates of K_L in each profile, we can compare these results to the uncertainties estimated for those quantities in section 4a. The mean of all estimated values is unbiased, with $\langle K_L/K_{L,\text{true}} \rangle \approx 1.002$. The coefficient of variation of simulated K_L estimates is about 0.3 (not shown), which is similar to, but slightly smaller than, the variability estimated by computing K_L from observations using different bin sizes (as shown in Fig. 11d). This gives us some confidence in our estimates of variability in K_L .

5. Discussion

a. Quality of possible float-based validation

At most wavelengths for MODIS and VIIRS, values of \bar{G} and \hat{G} differ from 1 by a few percent (Tables 3 and 4). Except for comparisons with MODIS at 443 and 555 nm, \bar{G} cannot be distinguished statistically from 1 using twice the standard error as a confidence interval. Thus, within observational precision, these mean estimates of R_{rs} match the satellite estimates. For MODIS, \bar{G} is less than 1 by 6% at 443 nm and greater than 1 by 12% at 555 nm. For both satellites, the 555-nm comparisons are not as good as the comparisons at other wavelengths, with low r^2 , relatively high σ_G , and a larger range of values in the satellite estimates than the float measurements. This disagreement is more notable for MODIS than for VIIRS. In addition, some values of G_i are larger at zenith angles above about 25°, which is most notable in MODIS *Aqua* in the 412 and 555 bands.

We compared the quality of float data to that of MOBY by examining R_{rs} matchups between satellites and MOBY. These MOBY observations were, following Bailey and Werdell (2006), extracted from the SeaWiFS Bio-Optical Archive and Storage System (SeaBASS) database (<http://seabass.gsfc.nasa.gov/seabasscgi/search.cgi>). The MOBY data are available as a preliminary product and include the observations used for vicarious calibration and additional observations. The quality control and data processing are similar but slightly different for these data as for data used in vicarious calibration (<http://seabass.gsfc.nasa.gov/wiki/article.cgi?article=MOBY>). The quality control thresholds on satellite data are the same as used for our float comparisons, and they are less stringent than the quality control criteria used for vicarious calibration.

Direct comparison of MOBY to satellite estimates of R_{rs} show that the float observations perform similarly to this expanded set of MOBY observations at all wavelengths (cf. Figs. 15 and 9). The bulk statistics (\bar{G} , EARD, and RMSD) comparing in situ observations with MODIS *Aqua* are similar in quality for floats and MOBY (Fig. 16), although the wavelength dependence of mismatches is different. The variability of G in the MOBY comparisons is likely a product of the looser quality control in the validation dataset than in the data used for calibration.

We have been unable to determine the cause of the disagreement between in situ and satellite R_{rs} estimates at 555 nm. We also examined MOBY and MODIS *Aqua* R_{rs} at 547 nm, for which the satellite sensor is better suited to ocean color (the 555-nm band is optimized for terrestrial observations). The comparison at 547 nm is similar to the comparison at 555 nm. That the

in situ–satellite disagreement exists in both float and MOBY comparisons indicates that difficulty in R_{rs} estimates for green wavelengths is not simply a problem with the floats. It is likely due to measurement difficulties or atmospheric correction errors. One possible in-water cause is Raman scattering, which may be particularly problematic at green wavelengths in oligotrophic conditions (Waters 1995; Westberry et al. 2013).

The general agreement between float and satellite estimates of R_{rs} suggests that the floats are likely to be useful platforms for validation of radiometric satellite data products. The disagreements described above suggest that more work is needed to fully understand the behavior of some wavelength bands. A more ambitious goal would be to use the floats for vicarious calibration of ocean color satellites, but it is beyond the scope of this analysis to test the capability of the floats to perform vicarious calibration. We interpret the statistical similarity of G from floats and MOBY as suggesting that floats may be capable of returning data with quality as high as that for MOBY, but more work is needed to determine whether the data from floats have sufficient stability to enable their use for vicarious calibration.

b. Recommendations

The floats that we used were modified versions of traditional APEX floats and were not fully optimized for radiometer measurements. Several improvements could be made so that the next generation of autonomous radiometry floats achieves more high-quality matchups and gives more precise (and possibly more accurate) estimates of R_{rs} . Improvements could be made to the vehicle design, instrumentation, sampling strategy, and analysis.

Starting with physical design, minimizing shading is extremely important. A float with two L_u sensors on arms that extend less than 1 m in opposite directions from the float body is a design that has been developed and tested by other researchers using a PROVOR float (ProVal project; E. Leymarie 2016, personal communication). With this configuration, one radiometer is always on the sunny side of the float and the shading is confined largely to the effect of the radiometer itself. Dual radiometers can also provide an estimate of measurement uncertainty and redundancy in the event of failure of one radiometer. A dual radiometer design must, however, minimize the effects on physical stability of the float.

We did not use our E_s measurements in this study, in part because our irradiance sensors were optimized for in-water measurements. For in situ measurements of E_s , we recommend using a dedicated in-air sensor. We did get important information from the failure of the E_d

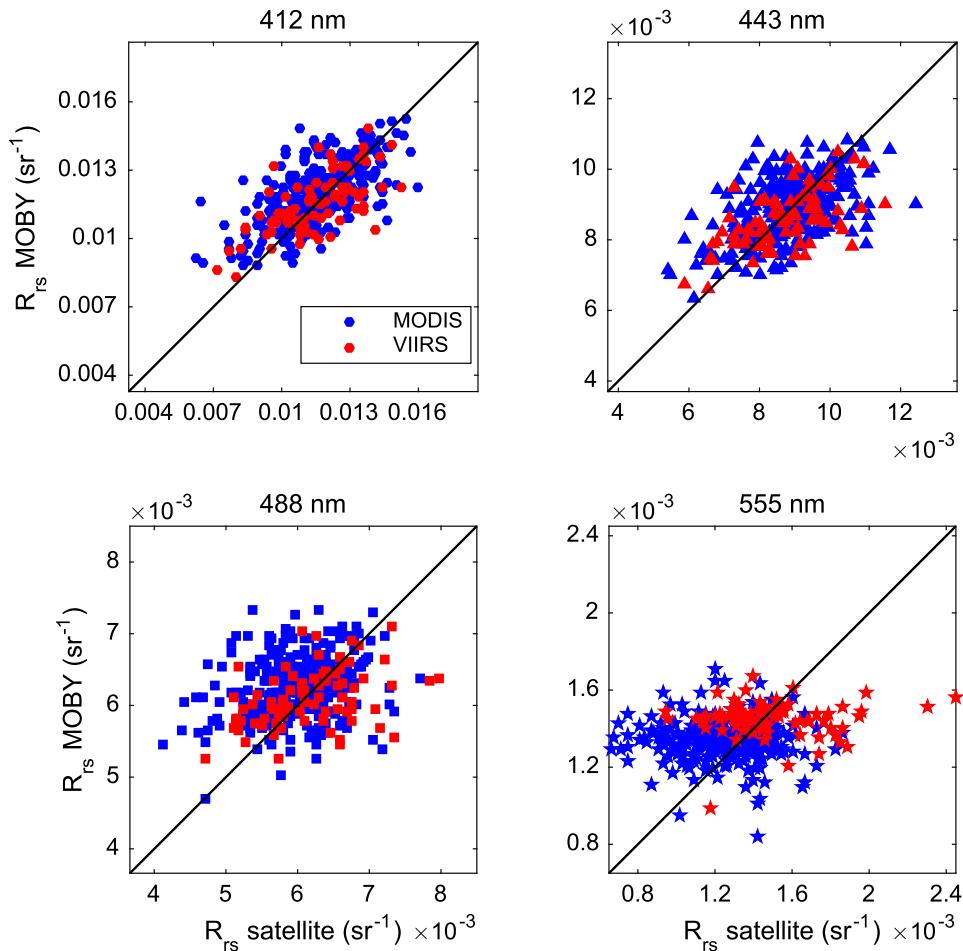


FIG. 15. Scatterplots of R_{rs} . Vertical axes are MOBY estimates. Horizontal axes are satellite estimates. Blue is MODIS *Aqua*. Red is VIIRS. Shapes correspond to wavelength as in previous figures. Black lines are 1:1.

sensor on the Hawaii A float. This failure caused the sampling computer to wait long times between samples of all instruments, which minimized the number of L_u measurements. This reinforces the importance of designing and thoroughly testing autonomous floats with complex sensor suites to respond gracefully to individual sensor failures.

The effects of tilting could be reduced by higher-frequency sampling of radiometric quantities and float orientation. Ideally, these would average over short times and would be at rates faster than the natural oscillation frequency of the float. Higher-frequency measurements will lead to increased data volume, so onboard averaging (to perhaps 1 Hz) may become necessary.

We had a limited number of surface samples (<10) for many profiles because of restrictions on tilt and azimuthal orientation. Future efforts should ensure that sufficient measurements are made to allow rejecting unfavorable samples without causing the number of available samples to be too small. To avoid unnecessary

time at the surface in poor conditions, onboard decision-making (based on E_s) could assess whether conditions are favorable for good L_w measurements. If so, then extended buoy phase measurements could be undertaken.

The L_u sensors could be located higher on the float body so that they are closer to the surface when taking their near-surface measurements. This would help reduce extrapolation errors due to poor estimates of K_L . However, having L_u measurements too close to the surface could increase the risk of the radiometers breaching the surface or lead to contamination by bubbles. Optimizing these trade-offs will require further study.

We note that parking the float at deep depths (~ 1000 m) seemed to be effective at minimizing fouling of the L_u sensors and most of the E_d sensors. Our floats profiled every other day, and each profile lasted several hours from the start to the float's return to park. If future floats profile daily, they will spend a relatively smaller length of time at park, possibly reducing the antifouling effects of the deep parking.

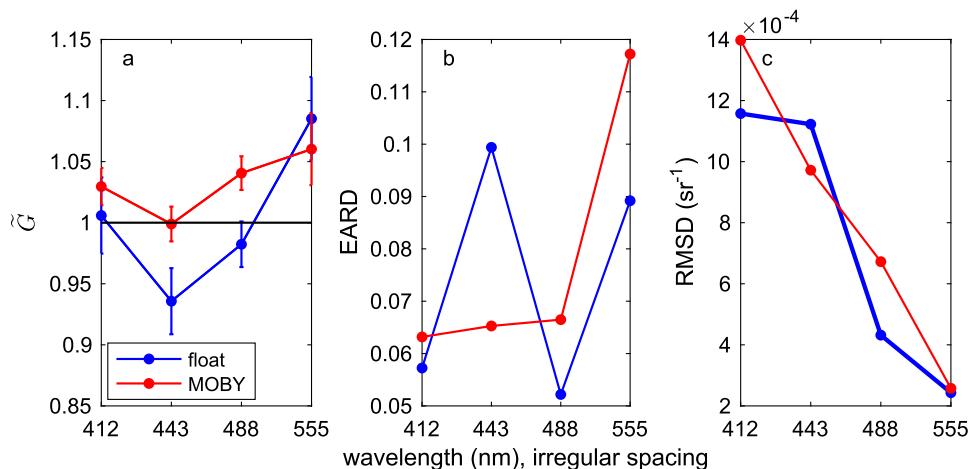


FIG. 16. Median statistics of G for MODIS *Aqua* compared to floats and to MOBY. (a) Median G (\bar{G} , 9), with error bars showing $\pm S_{95H}/\sqrt{N}$. (b) MARD (EARD: 14). (c) RMSD of 16. The number of samples for the MOBY data is 291.

Finally, if autonomous floats are to be used for vicarious calibration activities, hyperspectral sensors and retrieval of the floats to assess changes in calibration or fouling are recommended if possible. This is especially important for the 0.5% stability recommended for long-term records for use in climate studies (GCOS 2011; Zibordi et al. 2015).

6. Conclusions

This study shows that autonomous floats can be used for in situ validation of satellite estimates of remote sensing reflectance in the ocean. We made estimates of remote sensing reflectance using water-leaving radiance estimated from in situ observations and downwelling irradiance determined from a clear-sky model, and we compared these with satellite estimates of R_{rs} . The comparison showed a variability of several percent. Our comparisons have similar variability to satellite-in situ comparisons made using validation data from MOBY. Both floats and MOBY found poor agreement between in situ and MODIS *Aqua* estimates of R_{rs} in the 555-nm band.

We examined sources and magnitudes of uncertainty in our estimates of L_w using both observations and a Monte Carlo model. This analysis suggested that we were able to estimate water-leaving radiance with a precision of $\sim 5\%$ (including uncertainty due to environmental/measurement variability and calibration). As well as we could determine, our analytical procedure gave estimates of L_w that were unbiased. However, statistically significant average differences were found in the comparison of float and MODIS *Aqua* estimates of R_{rs} at two wavelengths. Understanding the

cause of this bias is necessary for use of these platforms for ocean color product validation.

Acknowledgments. The success of this endeavor owes much to many people. At Laboratoire d'Océanographie de Villefranche-sur-Mer we thank David Antoine, Herve Claustre, Emilie Diamond, Yann Hello, Antoine Poteau, and other members of the BOUSSOLE group for their considerable assistance with early testing and deployment of these floats in the Mediterranean; and Edouard Leymarie, for providing access to his SimulO software. For other deployments we appreciate the help of the MOBY operations team (Hawaii floats) and Ben Van Mooy (Atlantic floats). We thank Ronnie van Dommelen at Satlantic for the detailed discussion of radiometer calibration, Bill Woodward and Seth Ornstein of CLS America for their work on communications and telemetry, and Bob Fleming at UMaine for help with deployments and with handling of incoming data. Becca Conneely and Anastasia Rodzianko, then at Skidmore College, helped in early phases of the data analysis. Giuseppe Zibordi and two anonymous reviewers gave many helpful comments and criticisms that greatly improved this work. The work was funded by NASA and the National Oceanographic Partnership Program under Grant NNX09AP51G to the University of Maine.

REFERENCES

- Antoine, D., P. Guevel, J.-F. Diesté, G. Bécu, F. Louis, A. J. Scott, and P. Bardey, 2008: The "BOUSSOLE" buoy—A new transparent-to-swell taut mooring dedicated to marine optics: Design, tests, and performance at sea. *J. Atmos. Oceanic Technol.*, **25**, 968–989, doi:10.1175/2007JTECH0563.1.

- Austin, R., 1974: The remote sensing of spectral radiance from below the ocean surface. *Optical Aspects of Oceanography*, N. Jerlov and E. Steemann-Nielsen, Eds., Academic Press, 317–344.
- Bailey, S. W., and P. J. Werdell, 2006: A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sens. Environ.*, **102**, 12–23, doi:[10.1016/j.rse.2006.01.015](https://doi.org/10.1016/j.rse.2006.01.015).
- Bendat, J. S., and A. G. Piersol, 2000: *Random Data: Analysis and Measurement Procedures*. 3rd ed. John Wiley and Sons, 294 pp.
- Bishop, J. K. B., T. J. Wood, R. E. Davis, and J. T. Sherman, 2004: Robotic observations of enhanced carbon biomass and export at 55°S during SOFeX. *Science*, **304**, 417–420, doi:[10.1126/science.1087717](https://doi.org/10.1126/science.1087717).
- Boss, E., D. Swift, L. Taylor, P. Brickley, R. Zaneveld, S. Riser, M. Perry, and P. Strutton, 2008: Observations of pigment and particle distributions in the western North Atlantic from an autonomous float and ocean color satellite. *Limnol. Oceanogr.*, **53**, 2112–2122, doi:[10.4319/lo.2008.53.5_part_2.2112](https://doi.org/10.4319/lo.2008.53.5_part_2.2112).
- Clark, D. K., H. R. Gordon, K. J. Voss, Y. Ge, W. Broenkow, and C. Trees, 1997: Validation of atmospheric correction over the oceans. *J. Geophys. Res.*, **102**, 17 209–17 217, doi:[10.1029/96JD03345](https://doi.org/10.1029/96JD03345).
- , and Coauthors, 2003: MOBY, a radiometric buoy for performance monitoring and vicarious calibration of satellite ocean color sensors: Measurement and data analysis protocols. Special topics in ocean optics protocols and appendices, J. L. Mueller, G. S. Fargion, and C. R. McClain, Eds., Vol. VI, Ocean optics protocols for satellite ocean color sensor validation, revision 4, NASA Tech. Memo. NASA/TM-2003-211621/Rev4-Vol.VI, 3–34.
- D'Alimonte, D., G. Zibordi, T. Kajiyama, and J. C. Cunha, 2010: Monte Carlo code for high spatial resolution ocean color simulations. *Appl. Opt.*, **49**, 4936–4950, doi:[10.1364/AO.49.004936](https://doi.org/10.1364/AO.49.004936).
- Davis, R. E., J. T. Sherman, and J. Dufor, 2001: Profiling ALACEs and other advances in autonomous subsurface floats. *J. Atmos. Oceanic Technol.*, **18**, 982–993, doi:[10.1175/1520-0426\(2001\)018<0982:PAOAI>2.0.CO;2](https://doi.org/10.1175/1520-0426(2001)018<0982:PAOAI>2.0.CO;2).
- Estapa, M., K. Buesseler, E. Boss, and G. Gerbi, 2013: Autonomous, high-resolution observations of particle flux in the oligotrophic ocean. *Biogeosciences*, **10**, 5517–5531, doi:[10.5194/bg-10-5517-2013](https://doi.org/10.5194/bg-10-5517-2013).
- Franz, B. A., S. W. Bailey, P. J. Werdell, and C. R. McClain, 2007: Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. *Appl. Opt.*, **46**, 5068–5082, doi:[10.1364/AO.46.005068](https://doi.org/10.1364/AO.46.005068).
- Frouin, R., D. W. Lingner, C. Gautier, K. S. Baker, and R. C. Smith, 1989: A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance at the ocean surface. *J. Geophys. Res.*, **94**, 9731–9742, doi:[10.1029/JC094iC07p09731](https://doi.org/10.1029/JC094iC07p09731).
- GCOS, 2011: Systematic observation requirements for satellite-based data products for climate: 2011 update; Supplemental details to the satellite-based component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update),” WMO GCOS-154, 128 pp.
- Gordon, H., and D. Clark, 1981: Clear water radiances for atmospheric correction of coastal zone color scanner imagery. *Appl. Opt.*, **20**, 4175–4180, doi:[10.1364/AO.20.004175](https://doi.org/10.1364/AO.20.004175).
- , and M. Wang, 1994: Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Appl. Opt.*, **33**, 443–452, doi:[10.1364/AO.33.000443](https://doi.org/10.1364/AO.33.000443).
- Hooker, S., S. McLean, J. Sherman, M. Small, G. Lazin, G. Zibordi, and J. Brown, 2002: The Seventh SeaWiFS Intercalibration Round-Robin Experiment (SIRREX-7), March 1999. S. B. Hooker, and E. R. Firestone, Eds., SeaWiFS Postlaunch Tech. Rep. Series, Vol. 17, NASA Tech. Memo. 2002-206892, 69 pp.
- , C. McClain, and A. Mannino, 2007: NASA strategic planning document: A comprehensive plan for the long-term calibration and validation of oceanic biogeochemical satellite data. NASA Special Publ. NASA/SP-2007-214152, 31 pp.
- IOCCG, 2011: Bio-optical sensors on Argo floats. H. Claustre, Ed., International Ocean-Colour Coordinating Group Rep. 11, 89 pp.
- Leathers, R. A., T. J. Downes, and C. D. Mobley, 2004: Self-shading correction for oceanographic upwelling radiometers. *Opt. Express*, **12**, 4709–4718, doi:[10.1364/OPEX.12.004709](https://doi.org/10.1364/OPEX.12.004709).
- Leymarie, E., D. Doxaron, and M. Babin, 2010: Uncertainties associated to measurements of inherent optical properties in natural waters. *Appl. Opt.*, **49**, 5415–5436, doi:[10.1364/AO.49.005415](https://doi.org/10.1364/AO.49.005415).
- Mitchell, B. G., M. Kahru, and J. Sherman, 2000: Autonomous temperature-irradiance profiler resolves the spring bloom in the Sea of Japan. *Proc. 2000 Ocean Optics XV*, Monte Carlo, Monaco, Oceanographic Society, 9 pp.
- Mobley, C. D., 1994: *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, 592 pp.
- Morel, A., and S. Maritorena, 2001: Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.*, **106**, 7163–7180, doi:[10.1029/2000JC000319](https://doi.org/10.1029/2000JC000319).
- , D. Antoine, and B. Gentili, 2002: Bidirectional reflectance of oceanic waters: Accounting for Raman emission and varying particle scattering phase function. *Appl. Opt.*, **41**, 6289–6306, doi:[10.1364/AO.41.006289](https://doi.org/10.1364/AO.41.006289).
- Mueller, J. L., and Coauthors, 2003a: Instrument specifications, characterization and calibration. Vol. 2, Ocean optics protocols for satellite ocean color sensor validation, revision 4, NASA Tech. Memo. NASA/TM-2003-21621/Rev-VolIII, 57 pp.
- , R. Austin, A. Morel, G. Fargion, and C. McClain, 2003b: Introduction, background and conventions. Vol. I, Ocean optics protocols for satellite ocean color sensor validation, revision 4, NASA Tech. Memo. NASA/TM-2003-21621/Rev-Vol I, 50 pp.
- O'Reilly, J. E., S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. R. McClain, 1998: Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.*, **103**, 24 937–24 953, doi:[10.1029/98JC02160](https://doi.org/10.1029/98JC02160).
- , and Coauthors, 2000: SeaWiFS postlaunch calibration and validation analyses. J. E. Hooker and E. R. Firestone, Eds., SeaWiFS Postlaunch Tech. Rep. Series, Vol. 11, Part 3, NASA Tech. Memo. 2000-206892, 49 pp.
- Organelli, E., and Coauthors, 2016: A novel near-real-time quality-control procedure for radiometric profiles measured by Bio-Argo floats: Protocols and performances. *J. Atmos. Oceanic Technol.*, **33**, 937–951, doi:[10.1175/JTECH-D-15-0193.1](https://doi.org/10.1175/JTECH-D-15-0193.1).
- Quan, X., and E. S. Fry, 1995: Empirical equation for the index of refraction of seawater. *Appl. Opt.*, **34**, 3477–3480, doi:[10.1364/AO.34.003477](https://doi.org/10.1364/AO.34.003477).
- Roemmich, D., and Coauthors, 2009: The Argo Program: Observing the global oceans with profiling floats. *Oceanography*, **22**, 34–43, doi:[10.5670/oceanog.2009.36](https://doi.org/10.5670/oceanog.2009.36).

- Stramska, M., and T. D. Dickey, 1998: Short-term variability of the underwater light field in the oligotrophic ocean in response to surface waves and clouds. *Deep-Sea Res. I*, **45**, 1393–1410, doi:[10.1016/S0967-0637\(98\)00020-X](https://doi.org/10.1016/S0967-0637(98)00020-X).
- Voss, K. J., and Coauthors, 2010: An example crossover experiment for testing new vicarious calibration techniques for satellite ocean color radiometry. *J. Atmos. Oceanic Technol.*, **27**, 1747–1759, doi:[10.1175/2010JTECHO737.1](https://doi.org/10.1175/2010JTECHO737.1).
- Waters, K. J., 1995: Effects of Raman scattering on the water-leaving radiance. *J. Geophys. Res.*, **100**, 13 151–13 161, doi:[10.1029/95JC00530](https://doi.org/10.1029/95JC00530).
- Wei, J., M. R. Lewis, R. V. Dommelen, C. J. Zappa, and M. S. Twardowski, 2014: Wave-induced light field fluctuations in measured irradiance depth profiles: A wavelet analysis. *J. Geophys. Res. Oceans*, **119**, 1344–1364, doi:[10.1002/2013JC009572](https://doi.org/10.1002/2013JC009572).
- Werdell, P. J., S. W. Bailey, B. A. Franz, A. Mor el, and C. R. McClain, 2007: On-orbit vicarious calibration of ocean color sensors using an ocean surface reflectance model. *Appl. Opt.*, **46**, 5649–5666, doi:[10.1364/AO.46.005649](https://doi.org/10.1364/AO.46.005649).
- Westberry, T. K., E. Boss, and Z. Lee, 2013: Influence of Raman scattering on ocean color inversion models. *Appl. Opt.*, **52**, 5552–5561, doi:[10.1364/AO.52.005552](https://doi.org/10.1364/AO.52.005552).
- Xing, X., A. Morel, H. Claustre, F. D’Ortenzio, and A. Poteau, 2012: Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption retrieval. *J. Geophys. Res.*, **117**, C04022, doi:[10.1029/2011JC007632](https://doi.org/10.1029/2011JC007632).
- , H. Claustre, H. Wang, A. Poteau, and F. D’Ortenzio, 2014: Seasonal dynamics in colored dissolved organic matter in the Mediterranean Sea: Patterns and drivers. *Deep-Sea Res. I*, **83**, 93–101, doi:[10.1016/j.dsr.2013.09.008](https://doi.org/10.1016/j.dsr.2013.09.008).
- Zibordi, G., and K. J. Voss, 2014: In situ optical radiometry in the visible and near infrared. *Optical Radiometry for Ocean Climate Measurements*, G. Zibordi, C. J. Donlon, and A. C. Parr, Eds., Experimental Methods in the Physical Sciences, Vol. 47, Academic Press, 247–304, doi:[10.1016/B978-0-12-417011-7.00010-6](https://doi.org/10.1016/B978-0-12-417011-7.00010-6).
- , D. D’Alimonte, and J.-F. Berthon, 2004: An evaluation of depth resolution requirements for optical profiling in coastal waters. *J. Atmos. Oceanic Technol.*, **21**, 1059–1073, doi:[10.1175/1520-0426\(2004\)021<1059:AEODRR>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<1059:AEODRR>2.0.CO;2).
- , and Coauthors, 2009: AERONET-OC: A network for the validation of ocean color primary products. *J. Atmos. Oceanic Technol.*, **26**, 1634–1651, doi:[10.1175/2009JTECHO654.1](https://doi.org/10.1175/2009JTECHO654.1).
- , and Coauthors, 2015: System vicarious calibration for ocean color climate change applications: Requirements for in situ data. *Remote Sens. Environ.*, **159**, 361–369, doi:[10.1016/j.rse.2014.12.015](https://doi.org/10.1016/j.rse.2014.12.015).