

Radiometric field inter-comparison of fiducial reference measurements using an open source community processor

GAVIN H. TILSTONE,^{1,*} THOMAS M. JORDAN,¹ DIRK AURIN,² AGNIESZKA BIAŁEK,³ ALEXIS DERU,⁴ ASHLEY RAMSAY,³ MARTIN HIERONYMI,⁵ GIORGIO DALL'OLMO,⁶ MARTIN LIGI,⁷ CHARLES KOVACH,⁸ ILMAR ANSKO,⁷ MICHAEL ONDRUSEK,⁸ VIKTOR VABSON,⁷ GIUSEPPE ZIBORDI,^{9,10} JUAN IGNACIO GOSSN,¹¹ EWA KWIATKOWSKA,¹¹ AND RIHO VENDT⁷

¹Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

²MSU GESTAR-II, NASA Goddard Space Flight Center, Code 616, USA

³Climate and Earth Observation Group, National Physical Laboratory, Teddington TW11 0LW, UK

⁴ACRI-ST, Sophia-Antipolis, 06904 Valbonne, France

⁵Department of Optical Oceanography, Institute of Carbon Cycles, Helmholtz-Zentrum Hereon, 21502 Geesthacht, Germany

⁶Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, 34010 Sgonico, Trieste, Italy

⁷Tartu Observatory, University of Tartu, Tõravere, Estonia

⁸NOAA/NESDIS Center for Satellite Applications and Research (STAR), College Park, Maryland, USA

⁹National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD, USA

¹⁰Southeastern Universities Research Association, Washington, DC 20005, USA

¹¹EUMETSAT, Eumetsat-Allee 1, 64295 Darmstadt, Germany

*ghti@pml.ac.uk

Abstract: A radiometric field intercomparison was conducted at the Acqua Alta Oceanographic Tower (AAOT) in the Adriatic Sea from 14 to 21 July 2022 to assess differences in the accuracy of above-water radiometer systems (Sea-Bird HyperSAS, pySAS, TriOS-RAMSES) processed using an open-source community processor (HyperCP). Class-based and sensor-specific characteristics of the radiometers were used to determine the quantities $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$, and their associated uncertainties. Using sensor-specific characteristics, the differences among systems were 2% for $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and 2.5% for $R_{rs}(\lambda)$ and the uncertainties were 1.5%, 2%, 1.5% and 5%, respectively. The differences were higher when using class-based characteristics. $L_{wn}(\lambda)$ values were also compared to the above-water AERONET-OC SeaPRiSM and in-water HyperPro II. For SeaPRiSM, the differences and uncertainties were <5% over blue and green bands when using Mobley [App. Opt. 38, 7442 (1999)] sea surface reflectance factors (ρ) and no NIR correction. For HyperPro II the differences were larger but were reduced when $R_{rs}(\lambda)$ from the above-water systems was computed using ρ from Zhang et al. [Opt. Express 25, 1 (2017)] with similarity (sim) spectrum (spec) NIR correction. HyperCP using sensor-specific characteristics is highly recommended to reduce the associated uncertainties and to produce the highest quality data for satellite OC validation.

Published by Optica Publishing Group under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Radiometric inter-comparison exercises are necessary to ensure measurement consistency among optical systems that are used to validate satellite radiometric data products. The ocean color

(OC) community has a long history of instrument inter-comparison exercises for OC missions. Following the launch of SeaWiFS, NASA conducted SIRREX [1,2], an inter-calibration program that established common protocols, which led to a decrease in absolute radiometric calibration differences from 8% to 1%. In SIRREX-5, an inter-comparison was conducted with instruments from 8 laboratories. The discrepancies between $R_{rs}(\lambda)$ determined using various methods and diverse instruments were greater than the desired uncertainties for the field measurements [2]. This was followed by SIMRIC-1 and -2 [3,4], the purpose of which was to investigate consistency among the calibration procedures and protocols used by 10 different calibration laboratories. The agreement between calibration coefficients determined for a single radiometer (SeaWiFS Transfer Radiometer SXR-II) were within the combined uncertainties for all but two of the laboratories. Following the launch of MERIS in 2002, an inter-comparison off Plymouth, UK, of above- and in-water radiometers were within 1 to 2%, except for one sensor that exhibited cosine collector degradation [5]. A second field inter-comparison of above-water and in-water measurements for MERIS, conducted off south-west Portugal in 2003 showed differences within 5%. A successive inter-comparison included above-water and in-water radiometers and was conducted under near ideal deployment conditions at the AAOT in the northern Adriatic Sea in 2010 [6]. Differences among water-leaving radiance (L_w), downwelling irradiance (E_d) and remote sensing reflectance (R_{rs}) were within 1% and 6% for the above-water systems and 9% for the in-water systems. A radiometer inter-comparison exercise in support of Sentinel-3 OLCI (FRM4SOC 1) was also conducted at the AAOT in 2018, in which above-water systems, and in-water profilers, were compared against SeaPRISM [7]. The differences in normalized water-leaving radiance (L_{wn}) were 4.8% across the visible bands for two above-water systems and 11.8% for the two in-water systems due to temporal and spatial differences in sampling and instrument self-shading for one system. This study highlighted the impact of differences in processing, that rely on diverse data reduction procedures, which can lead to error compensations that make it difficult to assess the results based on differences alone. This further showed the importance of uncertainty budgets in future studies and spurred the FRM4SOC Phase 2 Radiometric Field inter-comparison (FICE-22) at the Aqua Alta Oceanographic Tower in July 2022. The objectives of this new inter-comparison were to assess differences in the accuracy of above-water radiometer systems (Sea-Bird HyperSAS and TriOS-RAMSES) processed using the same open-source community data processor (HyperCP). The HyperCP is a toolkit designed to provide automated processing of above-water hyperspectral radiometric data using state-of-the-art methods and protocols for quality procedures that also includes uncertainty propagation. Though the radiometric systems are hyperspectral, for the inter-comparison they were matched to OLCI bands. The same processor was also coded to produce end-to-end uncertainties, which are reported with $k = 1$ coverage factor. It was run using class-based and sensor-specific radiometer characteristics. Class-based radiometric quantities and uncertainties are computed using mean instrument class characteristics derived from ~40 individual radiometers. The sensor-specific processing chain differs from the class-based one by using the individual instrument characteristics to compute the radiometric quantities and uncertainties. The outputs from class-based and sensor-specific were compared and recommendations on reducing the uncertainties, based on these, are given.

2. Materials and measurements

The radiometric quantities, $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$, $R_{rs}(\lambda)$ and $L_{wn}(\lambda)$ from five above-water systems; one SeaBird-HyperSAS, one pySAS, two TriOS-RAMSES and AERONET-OC SeaPRISM and one in-water SeaBird HyperPro II, were inter-compared. Sections 2.1 to 2.3 describe how these quantities are determined.

2.1. Determination of water-leaving radiance: above-water

Illumination in above-water radiometry during clear sky conditions is largely determined by the sun zenith angle θ_0 , and to a lesser extent by atmospheric properties. The water-leaving radiance L_w is computed by removing glint perturbations from L_t as follows:

$$L_w(\theta, \theta_0, \Delta\phi, \lambda) = L_t(\theta, \theta_0, \Delta\phi, \lambda) - \rho(\theta, \theta_0, \Delta\phi, U_{10})L_{sky}(\theta', \theta_0, \Delta\phi, \lambda), \quad (1)$$

where $L_t(\theta, \theta_0, \Delta\phi, \lambda)$ is the total radiance from the sea, $L_{sky}(\theta', \theta_0, \Delta\phi, \lambda)$ the sky radiance, and $\rho(\theta, \theta_0, \Delta\phi, U_{10})$ is the sea surface reflectance factor typically expressed as a function of the sun-sensor geometry, with θ and θ_0 being the viewing and sun zenith angles, $\Delta\phi$ the relative azimuth with respect to the sun, and U_{10} the wind speed at 10 m above the sea surface [8]. For this inter-comparison, θ was set to 40° implying $\theta' = 180^\circ - \theta$ and $\Delta\phi$ set to either 90° or 135° .

2.2. Determination of water-leaving radiance: in-water

The sub-surface upwelling radiance determined with nadir view, $L_u(0^-, \lambda)$, is transmitted above water to obtain the water-leaving radiance:

$$L_w(0, \theta_0, \Delta\phi, \lambda) = L_u(0^-, \lambda) \cdot \frac{(1 - \rho')}{n^2}, \quad (2)$$

where 0^- indicates just below the surface, ρ' is the water-air interface sea surface reflectance, n is the refractive index of water.

2.3. Determination of remote sensing reflectance and normalized water-leaving radiance for above-water measurements

The remote sensing reflectance $R_{rs}(\theta, \theta_0, \Delta\phi, \lambda)$ is defined as the ratio of the L_w to the above-water downwelling irradiance $E_d(0^+, \lambda)$, where 0^+ indicates just above the surface and is computed as:

$$R_{rs}(\theta, \theta_0, \Delta\phi, \lambda) = \frac{L_w(\theta, \theta_0, \Delta\phi, \lambda)}{E_d(0^+, \lambda)}. \quad (3)$$

The exact normalized water-leaving radiance L_{wn} is computed as per Morel et al. (2002):

$$L_{wn}(\lambda) = R_{rs}(\theta, \theta_0, \Delta\phi, \lambda)BRDF(\theta, \theta_0, \Delta\phi, \lambda, chl)F_0(\lambda), \quad (4)$$

where $F_0(\lambda)$ is the extraterrestrial solar irradiance [9] and chl is the total chlorophyll concentration. $BRDF(\theta, \theta_0, \Delta\phi, \lambda, chl)$ indicates bi-directional reflectance factor

$$BRDF(\theta, \theta_0, \Delta\phi, \lambda, chl) = \mathfrak{R}_0(U_{10}) \frac{f_0(\lambda, U_{10}, chl)}{Q_0(\lambda, U_{10}, chl)} \left[\mathfrak{R}(\theta, U_{10}) \frac{f(\theta_0, \lambda, U_{10}, chl)}{Q(\theta, \theta_0, \Delta\phi, \lambda, U_{10}, chl)} \right]^{-1} \quad (5)$$

where, \mathfrak{R} accounts for combined reflection/refraction effects by the sea surface, Q is the Q -factor indicating the ratio of upward irradiance to upwelling radiance just below the surface, and f relates the irradiance reflectance to the water inherent optical properties conveniently expressed by chl . Look-up-tables of $\frac{f}{Q}$ and \mathfrak{R} values were taken from Morel et al. [10] for specific wavelengths. The quantities $f_0(\lambda, U_{10}, chl)$, $Q_0(\lambda, U_{10}, chl)$ and $\mathfrak{R}_0(U_{10})$ indicate values of $f(\theta_0, \lambda, U_{10}, chl)$, $Q(\theta, \theta_0, \Delta\phi, \lambda, U_{10}, chl)$ and $\mathfrak{R}(\theta, U_{10})$ determined for $\theta_0 = 0$, respectively. The BRDF factor was computed as a function of chl estimated from OC4 M, which resulted in values ranging from 0.1 to 0.8 mg m^{-3} . BRDF corrections were determined for the ± 90 and ± 135 degrees relative azimuths between sensor and sun, and the sensor viewing angle of 40 degrees assuming $U_{10} = 2$ m/s.

2.4. Field intercomparison

The field intercomparison was conducted from 14 to 21 July 2022 at the AAOT in the Gulf of Venice, Italy, at 45.31°N, 12.51°E. The AAOT is a fixed oceanographic tower allowing for the measurement of ocean properties under stable conditions. The platform has a long history of optical measurements, which often provides near ideal conditions to validate OC missions under clear skies, low wind speed and calm sea state [11]. Two AERONET-OC SeaPRISM instruments are deployed at the western corner of the AAOT (Fig. 1). A CE-318 9-channel radiometer (called Venice) has been active since April 2002, and a CE-318 T 12-channel radiometer (named AAOT) has been operated since October 2017 [12]. For the inter-comparison, four above-water, two RAMSES [7,13], one HyperSAS [14], one automated HyperSAS (pySAS) [15] and one in-water HyperPro II [16] were deployed.

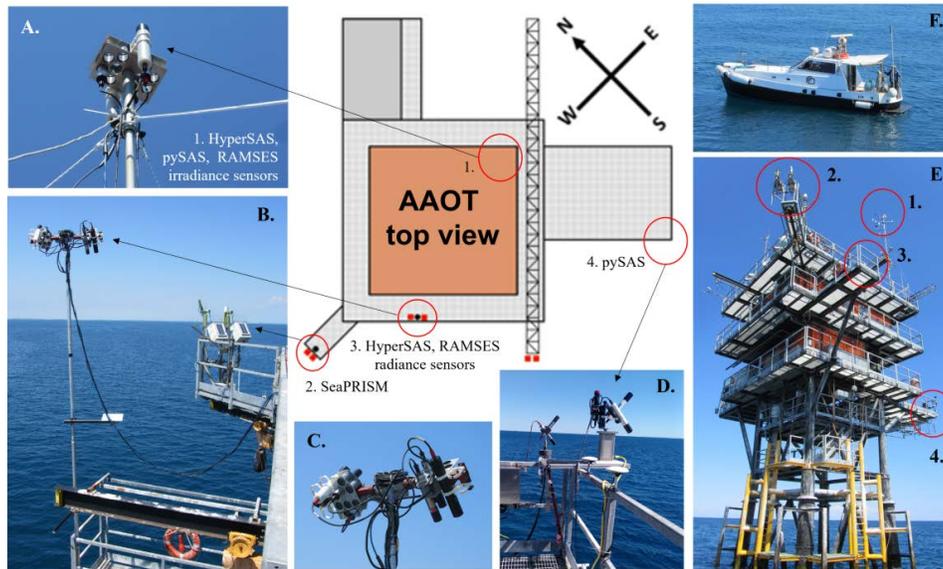


Fig. 1. Locations of the radiometric measurement systems on the AAOT for the field inter-comparison. (A.) HyperSAS, pySAS, RAMSES-A & -B irradiance sensors located on a Fireco mast on the top level of the AAOT, (B.) SeaPRISM and HyperSAS, RAMSES-A & -B radiance sensors located side-by-side (C.) close up of HyperSAS, RAMSES-A, -B radiance sensors, (D.) pySAS system, (E.) view of the AAOT showing locations of radiance sensors, (F.) deployment of Hyperpro II from *RV Litus*,

To rule out any difference arising from independent absolute radiometric calibrations, all sensors were calibrated at the University of Tartu (UT) [17] applying the same methods and standards within ~1 month of the campaign. The instruments were then shipped directly to the AAOT. Processed $L_{sky}(\lambda)$, $L_t(\lambda)$, $E_d(\lambda)$ and $R_{rs}(\lambda)$ data with application of the OLCI's spectral response functions to obtain radiometric quantities equivalent to those corresponding to the OLCI channels centered at 400, 412, 443, 490, 510, 560, 620, 665, 674, 681 nm, were produced along with a UTC timestamp, the make, model, serial number of the instrument and integration time setting used during the acquisition.

2.5. Above-water systems

2.5.1. TriOS-RAMSES

For above-water measurements RAMSES-A and -B, both comprising three TriOS (Mess- und Datentechnik GmbH, Germany) radiometers were deployed by UT and Hereon (Table 1). The

systems consisted of two RAMSES ARC-VIS hyperspectral radiance sensors for measuring $L_{sky}(\lambda)$ and $L_t(\lambda)$ respectively, and one RAMSES ACC-VIS irradiance sensor for measuring $E_d(\lambda)$. Measurements were made over the spectral range of 350–950 nm with a spectral resolution of about 10 nm, a spectral sampling of approximately 3.3 nm and a wavelength accuracy of 0.3 nm, respectively. The nominal full angle field-of-view (FOV) of the radiance sensors is 7°. The sensors are based on the Carl Zeiss Monolithic Miniature Spectrometer (MMS 1) incorporating a 256 channel silicon photodiode array. Integration time varied from 4 ms to 8 s, which is automatically adjusted based on light intensity to prevent saturation. The data stream from all three instruments is integrated by an IPS-104 power supply and interface unit and logged on a PC via a RS232 connection. For the deployment and processing of data, both UT and Hereon followed published satellite validation protocols [18].

Table 1. Field intercomparison measurement systems, sensors and institutes used in the analysis.

	Method (ID)	Radiometers	Institute
1	Above-water (RAMSES-A)	TriOS	University of Tartu, Estonia
2	Above-water (RAMSES-B)	TriOS	Hereon, Germany
3	Above-water (HyperSAS)	Seabird	Plymouth Marine Lab, UK
4	Above-water (pySAS)	Seabird	NASA, USA
5	Above-water (SeaPRISM)	SeaPRISM	Joint Research Centre, Italy
6	In-water HyperPro II	Seabird	NOAA, USA

2.5.2. Sea-Bird HyperSAS

PML deployed a measurement system consisting of three hyperspectral Sea-Bird (Washington, USA; formerly SATLANTIC Inc.) radiometers, two measuring $L_{sky}(\lambda)$ and $L_t(\lambda)$ and one measuring $E_d(\lambda)$, that are routinely deployed on the Atlantic Meridional Transect [14]. The sensors measure over the 350–900 nm wavelength range with a spectral sampling of approximately 3.3 nm and a spectral resolution of about 10 nm. Integration time can also vary from 4 ms to 8 s and is automatically adjusted to the light intensity. The data stream from all three instruments is integrated by an interface unit and logged on a PC via a RS232 connection. The radiance sensors have a FOV of 6°. HyperSAS radiometers are equipped with a shutter that closes periodically to record dark values. The $E_d(\lambda)$, $L_t(\lambda)$, and $L_{sky}(\lambda)$ data were first dark corrected by interpolating the dark values in time to match the light measurements for each sensor. Then, the dark values were subtracted from the spectral light measurements of each radiometer. $E_d(\lambda)$, $L_t(\lambda)$ and $L_{sky}(\lambda)$ data were then interpolated to a common set of wavelengths every 3.3 nm from 353–898 nm.

2.5.3. pySAS system

The pySAS system comprises Sea-Bird HyperSAS radiometers (as described in Section 2.7.2) operated on a custom robotic sun-tracking mount connected to a Raspberry Pi 3B + computer for control and data logging. The system autonomously adjusts the relative azimuth to within preset limits to avoid platform perturbations, while continuously logging data from the sensors and controller. Data from all instruments, including directional GPS, tilt sensor, sun tracker, and the three radiometers are merged into hour-long, Sea-Bird-styled files using python software prepSAS (<https://github.com/OceanOptics/pySAS/tree/master/repSAS>).

2.5.4. SeaPRISM AERONET-OC

The SeaWiFS Photometer Revision for Incident Surface Measurements (SeaPRiSM) is a CE-318 T sun-photometer (CIMEL, Paris, France) that has the capability to perform autonomous above-water measurements with a FOV of 1.2° at a number of 10 nm wide spectral bands. The CE-318

T instrument is a new version of the previous CE-318, which had fewer spectral bands centred on those of MODIS-Aqua. CE-318 T replicated measurement sequences (typically 3) are made every 30 min. $L_w(\lambda)$ is determined using spectral bands centred at 400, 412, 443, 490, 510, 560, 620, 667, 779, 865, 1020 nm, which largely match those of the Sentinel-3 OLCI in the visible spectral region. These measurements are: 1.) the direct sun irradiance $E_s(\lambda)$ acquired to determine the aerosol optical depth $\tau_a(\lambda)$ used for the theoretical computation of $E_d(\lambda)$, and; 2.) a sequence of 11 sea-radiance measurements for determining $L_t(\lambda)$ and of three sky radiance measurements for determining $L_{sky}(\lambda)$. These measurements are sequentially repeated for each λ with $\Delta\phi = 90^\circ$ and $\theta = 40^\circ$. The larger number of sea measurements, when compared to sky measurements, are required because of the higher environmental variability (mostly produced by wave perturbations) affecting the sea measurements during clear skies. Quality flags are applied at the different processing levels to remove poor data. These include checking for cloud contamination, high variance of the sequential measurements applied for the determination of $L_t(\lambda)$ and $L_{sky}(\lambda)$, elevated differences between pre- and post- deployment calibrations of the SeaPRiSM system, and spectral inconsistencies affecting $L_{wn}(\lambda)$ [12]. Fully quality controlled data are accessible at Level 2.0 (https://aeronet.gsfc.nasa.gov/cgi-bin/draw_map_display_seaprisim_v3).

2.6. In-water system

2.6.1. HyperPro II deployment method

The Seabird HyperPro II comprises a HyperOCR radiometer that measures the upwelling radiance $L_u(z, \lambda)$ and a HyperOCI irradiance sensor that measures the downward irradiance $E_d(z, \lambda)$ in the water column at depths z . The HyperPro II also includes an HyperOCI irradiance sensor to measure the above-water downward irradiance $E_d(0^+, \lambda)$, which is used as the reference during data reduction. Both HyperOCR and HyperOCI have a 10-nm spectral resolution over the 350 to 900 nm spectral range. The FOV of the in-water radiance sensor is 8.5° . The HyperOCRs benefit from dark signal corrections performed using a shutter operated every 5 regular measurement scans. The HyperPro II was deployed from the in-shore boat *RV Litus* and the above-water reference irradiance sensor was located on a pole mounted above the wheel house. At each coincident above-water station cast, one HyperPro II profile station was produced. HyperPro profile data were collected with a deployment speed as low as 15 cm s^{-1} . The multicast technique allowed to continuously log data while the instrument was profiled up and down in the water column up to 5 times. This led to the collection of a sufficient number of readings per meter to ensure that depth resolution requirements were met for accurate determination of sub-surface radiometric values [19].

2.7. Radiometer setup and experimental design

The irradiance sensors from HyperSAS, pySAS and RAMSES systems were located on the same telescopic mast (Fireco, Italy) (Fig. 1(a)) at the eastern corner of the AAOT, 21 m above the sea surface (Fig. 1(d)). The radiance sensors from three of the manually operated measurement systems (RAMSES-A, RAMSES-B, HyperSAS) were located on the same purpose-built frame on the western corner of the AAOT at a height of 17 m from the sea surface, alongside the SeaPRISM systems (Fig. 1(b)–(d)). The L_{sky} and L_t sensors had identical viewing zenith angles of 40° . A sundial was located mid-way down the mast of the frame to determine the correct $\Delta\phi$. The deployment frame was adjusted for each measurement sequence so that $\Delta\phi = 135^\circ$, which are typically used to reduce sun glint [8]. If this was not possible, $\Delta\phi = 90^\circ$ was used. The pySAS was located on the south-western corner of the tower on a lower level (Fig. 1(d)). For the manually operated systems, measurements were made at 20 min intervals, from 08:00 to 13:00 UTC, over a discrete measurement period of 5 min (called “cast”), with all instruments having a synchronized start time so that the data collected were directly comparable. Data from the pySAS were extracted over the same measurement period. In-water HyperPro II (Fig. 1(e))

measurements were also coordinated to these times, though with a temporal delay that is inherent with the practicalities of deployment.

2.8. Convolution of hyperspectral measurements to the Sentinel-3 OLCI bands

Hyperspectral measurements of $L_t(\lambda)$, $L_{sky}(\lambda)$, $E_d(\lambda)$ and $R_{rs}(\lambda)$ were converted into equivalent Sentinel-3 (S-3) Ocean and Land Colour Instrument (OLCI) bands by applying the OLCI spectral response functions [20] as in the following example for $L_t(\lambda)$:

$$L_t(\lambda_{i,OLCI}) = \frac{\int S_{i,OLCI}(\lambda)L_t(\lambda)d\lambda}{\int S_{i,OLCI}(\lambda)d\lambda}, \quad (6)$$

where $L_t(\lambda_{i,OLCI})$ and $S_{i,OLCI}(\lambda)$ are L_t and the OLCI Spectral Response Function (SRF) for the i^{th} OLCI channel, respectively. The AAOT AERONET-OC SeaPRISM system measures $L_t(\lambda)$ and $L_{sky}(\lambda)$ directly at the OLCI bands, while $E_d(\lambda)$ is estimated using a solar irradiance model (see Section 2.9.1).

2.9. Data processing and computation of uncertainties using HyperCP

HyperCP was developed through an open science, open-source collaboration effort established to facilitate community development of a data processing toolkit for *in situ* above-water radiometry supporting aquatic biogeochemistry applications, algorithm development and satellite validation. HyperCP has a strong heritage in NASA's HyperInSPACE and University of Victoria's PySciDON software processing packages, both of which originally supported Sea-Bird HyperSAS radiometry data for archiving to NASA SeaBASS. HyperCP adds the capability of processing RAMSES data and computing either class-based or sensor-specific uncertainties for both HyperSAS and RAMSES systems following the Guide for Uncertainty Measurement (GUM) [21]. HyperCP processes datasets with a common baseline using standardized parameters. The main objective of HyperCP is to provide the community with the means of processing above-water radiometric data to Fiducial Reference Measurement (FRMs) standards, which requires the determination of uncertainties.

HyperCP has a range of processing solutions and options that can be utilized. For this inter-comparison the data were processed, as follows: Firstly, before applying radiometric calibration, the data were verified to meet the relative azimuth angles between instrument and sun, i.e., 90° and 135° . Spectral measurements were synchronized across radiometers to interpolate them to time and 3.3 nm spectral sampling to account for small inter-radiometer differences. Quality control filters were then applied to remove data exhibiting $L_w(\lambda)$ higher in the NIR [780 to 850 nm] than in the UV [350 to 400 nm], wind speeds $>5 \text{ m s}^{-1}$ and solar zenith angles outside the 20° - 60° range. Any remaining spectral outlier was filtered by removing spectra (normalized to their maxima) within each file that fall outside of 5.0, 8.0, and 3.0 standard deviations over the spectral range from 400 nm to 700 nm for $E_d(\lambda)$, $L_t(\lambda)$ and $L_{sky}(\lambda)$, respectively. Acquisitions were then binned to 5 min ensembles, from which the brightest 90% (as measured by the L_t sensor) were discarded to reduce the effect of capillary wave solar reflection (glitter). Finally, $R_{rs}(\lambda)$ and $L_{wn}(\lambda)$ were computed. Spectra in which $R_{rs}(\lambda)$ is negative between 380 nm and 700 nm, were also removed. The hyperspectral data were then convolved to S-3 OLCI wavebands.

For this intercomparison, the data were processed using the Mobley [8] ρ with no residual NIR correction (hereafter referred to as M99NN). The data were also processed using the Zhang et al. [22] ρ factors, with the similarity spectrum residual NIR correction (hereafter called Z17SS) determined over the spectral range from 750 to 800 nm. For the computation of $L_{wn}(\lambda)$, the BRDF correction described in Section 2.3 was applied.

2.9.1. HyperCP class-based data processing

Class-based uncertainties are computed using the individual radiometric calibration files that, in addition to calibration coefficients, include the uncertainty of the calibration coefficients assigned for each sensor. To account for instrument characteristics, the uncertainty contributions for an instrument class were determined by UT based on characterization of 37 individual radiometers, which included 20 RAMSES and 16 hyperOCR. Eight uncertainty contributions were accounted for:

- i. Absolute radiometric calibration, from the calibration certificate.
- ii. Absolute calibration stability, estimated as $\sim 1\%$.
- iii. Detector non-linearity, estimated as $\sim 2\%$.
- iv. Spectral straylight sensitivity, which varies spectrally and per class of instrument.
- v. Temperature sensitivity, which varies spectrally and per class of instrument.
- vi. Polarization sensitivity for radiance radiometers, which varies spectrally and per class of instrument.
- vii. Cosine response for irradiance radiometers, estimated to be 3.5% for RAMSES and 2% for HyperSAS. These estimates neglect any dependence on the sun zenith angle.
- viii. Environmental variability, estimated from the data statistics of the quality controlled individual measurements contributing to each measurement cast.

The contribution from each uncertainty is propagated through the various processing steps leading to the determination of $R_{rs}(\lambda)$ following the GUM methodology. Uncertainties are propagated using a Monte Carlo method [21], where each component of the measurement equation has assigned a probability distribution function (PDF) that expresses its uncertainty. Then, the measurement equation is run many times by randomly drawing its input values from the PDFs of the input components, thus leading to a PDF of the output value representing its uncertainty. The correlations between input components are defined and the information is stored in correlation matrixes. For the practical implementation of Monte Carlo uncertainty propagation in HyperCP the “Propagation of UNcertainties in Python” (pumpy) package included in the CoMet Toolkit [23], is used.

2.9.2. HyperCP sensor-specific data processing

The sensor-specific processing chain differs from the class-based one. In this regime, each instrument has an extended calibration file which contains, in addition to calibration coefficients with their uncertainty, the data from the laboratory calibration acquired at two different integration times. These are used to calculate sensor specific nonlinearity correction following Vabson et al. [17]. In addition, each instrument has a dedicated file with correction coefficients for:

- Spectral stray light,
- Temperature sensitivity,
- Angular cosine response for irradiance sensors.

Thus, corrections due to the above non-ideal performances of each radiometer are applied in the processing chain and only the residual uncertainty on corrections are propagated. In this way the instrument related uncertainty contributors are significantly reduced. The polarization uncertainty is handled in the same manner in both workflows as well as all remaining uncertainty components that are not directly related to the instruments. These include the uncertainties in ρ factors and BRDF corrections.

2.10. Individual processors

2.10.1. TriOS-RAMSES data processing

Proprietary MSDA XE software was used to acquire the data from both RAMSES-A and -B systems acquired every 10 s for the duration of each 5 min cast, and then applying calibration coefficients determined before the campaign by UT. For RAMSES-A, data were merged using bespoke Python software. The irradiance sensor had GPS time, location, tilt and heading. No corrections were applied, but spectra with missing or saturated values were removed from the database. For RAMSES-B, all 60 spectra per cast were used for averaging and determine the standard deviation. No flagging was applied and visual quality control confirmed expected natural variability for clear sky conditions. All sensor data were interpolated to a common wavelength basis with 2.5 nm intervals. For RAMSES-B, to account for the uncertainties from sky glint correction, three ρ factor, from Mobley [1], Ruddick et al. [24] and sea state-related roughness modifications from Hieronimi [25], were used and the results were averaged.

2.10.2. Sea-Bird HyperSAS data processing

HyperSAS data processing follows Lin et al. [14]. In brief, data were first extracted from the raw instrument files and the pre-campaign calibration coefficients were applied. The standard processing used no NIR correction.

2.10.3. pySAS system data processing

PySAS data were processed in HyperCP as described in section 2.9 with a slightly different parameterization. Specifically, the surface reflectance correction of Z17SS was applied and $R_{rs}(\lambda)$ and $L_{wn}(\lambda)$ were further corrected by applying the similarity spectrum approach of [24] to remove residual NIR reflectance.

2.10.4. HyperPro II data processing

Multi-cast data saved into a single file were processed using Seabird ProSoft Software version 8.1.6 applying straylight and thermal corrections, and successively removing measurements affected by high tilt. A least squared regression fit was applied to the log-transformed $L_u(z, \lambda)$ data to determine the subsurface upwelling radiance $L_u(0^-, \lambda)$. Spectrally independent extrapolation depths were determined to ensure log-transformed data still exhibit linear decay at each spectral band. $L_w(\lambda)$ and $R_{rs}(\lambda)$ were calculated as described in the sections above. The $R_{rs}(\lambda)$ from individual multicasts were averaged for each station.

2.11. Data quality control

AERONET-OC SeaPRiSM Level 2.0 data were quality controlled (QC) following Zibordi et al. [26]. The mid-point of the cast time +/- 10 minutes from QC SeaPRiSM data were used to filter the data for the inter-comparison. Wind speed was acquired from the AAOT meteorological station. Only casts with wind speeds $< 5 \text{ ms}^{-1}$ and exhibiting clear skies were retained for the inter-comparison.

2.12. Statistical analyses

For all above-water systems, $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ were determined over a 5 min period for each cast. Results are presented as an inter-comparison between RAMSES-A, -B, HyperSAS and pySAS with uncertainties from HyperCP for sensor-specific corrections applied for the instruments non-ideal performances. Results from individual above-water systems were compared to the mean of all sensors. Differences between individual processors (IP) and the HyperCP were also assessed to cross check the performance of HyperCP against previously published processors. For the IP, individual quality control procedures were applied, and mean,

median and standard deviation values were computed. $L_{wn}(\lambda)$ determined from SeaPRISM and in-water HyperPro II data were used as independent reference data to compare against the above-water data.

3. Results

3.1. Variability in $R_{rs}(\lambda)$

Figure 2(a) shows the variation in $R_{rs}(\lambda)$ measured by the pySAS during the campaign. Figure 2(b) shows the stations that passed the QC filtering for the manually operated (HyperSAS, RAMSES-A, -B) and automated systems (pySAS and SeaPRiSM indicated by ‘AAOT’), and illustrates the larger number of measurements (N) for the autonomous systems. To illustrate the quality of the data, the Quality Water Index Polynomial (QWIP) [27] was used (Fig. 2(c)). The QWIP is a metric of color correlated to the spectral shape and represents a polynomial relationship between the location of the weighted peak in $R_{rs}(\lambda)$ (known as the Apparent Visible Wavelength, AVW) [28] and a Normalized Difference Index determined with red and green wavelengths. The technique screens for spectra with incorrect removal of surface reflected skylight or unrealistic shapes, which are indicated by QWIP indices > 0.2 . Figure 2(d) shows matching data from the AAOT meteorological station, and Fig. 2(e) shows the time-series of chl and the absorption coefficient of the sum of detrital and colored dissolved organic matter (a_{dg}) derived from the pySAS $R_{rs}(\lambda)$. A high value of a_{dg} is here indicative of optically complex waters where the in-water signal is related to total suspended material (TSM) and colored dissolved organic material (CDOM) not correlated with chl .

On 15th July, the shape of the $R_{rs}(\lambda)$ is typical when chl is $< 1 \text{ mg m}^{-3}$ and AVW is between 480–490 nm (blue spectra in Fig. 2(a)). From the 16th to 18th July 2022, high wind speeds (which prohibited reaching the AAOT to conduct manually operated measurements) were followed by an increase in total suspended materials (TSM) and chl . This caused an increase in $R_{rs}(\lambda)$ slopes

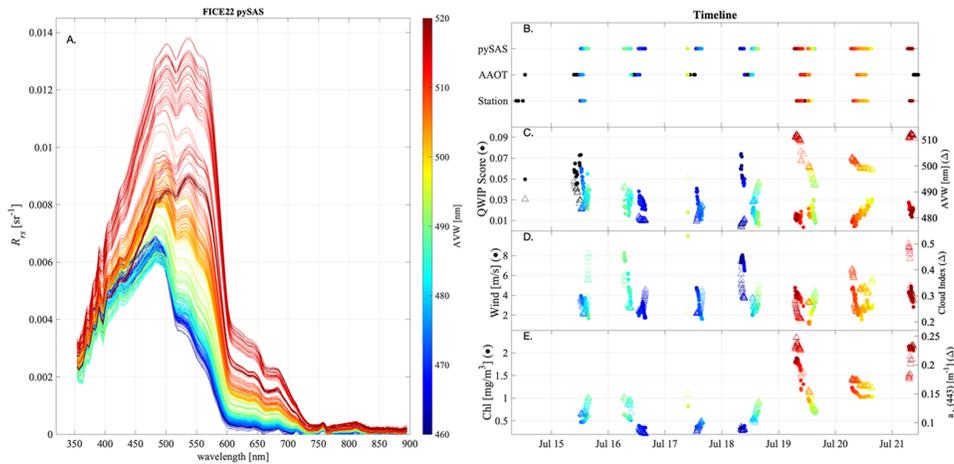


Fig. 2. Variability in (A.) quality controlled $R_{rs}(\lambda)$ from the pySAS data during the campaign from 14 to 22 July 2022, (B.) timeline of data acquisitions, (C.) pySAS-derived reflectance quality parameters QWIP and AVW, (D.) meteorological data from a nearby buoy and (E.) pySAS-derived OC properties chl and a_{dg} (all identified by the symbols • and Δ). Stations in (B.) indicate when synchronized manual radiometric acquisitions were collected. Prior to the pySAS becoming operational on 15 July, the QWIP and AVW are calculated from SeaPRISM (indicated by AAOT and shown in black in (B.)). In all sub-figures, the color indicates AVW.

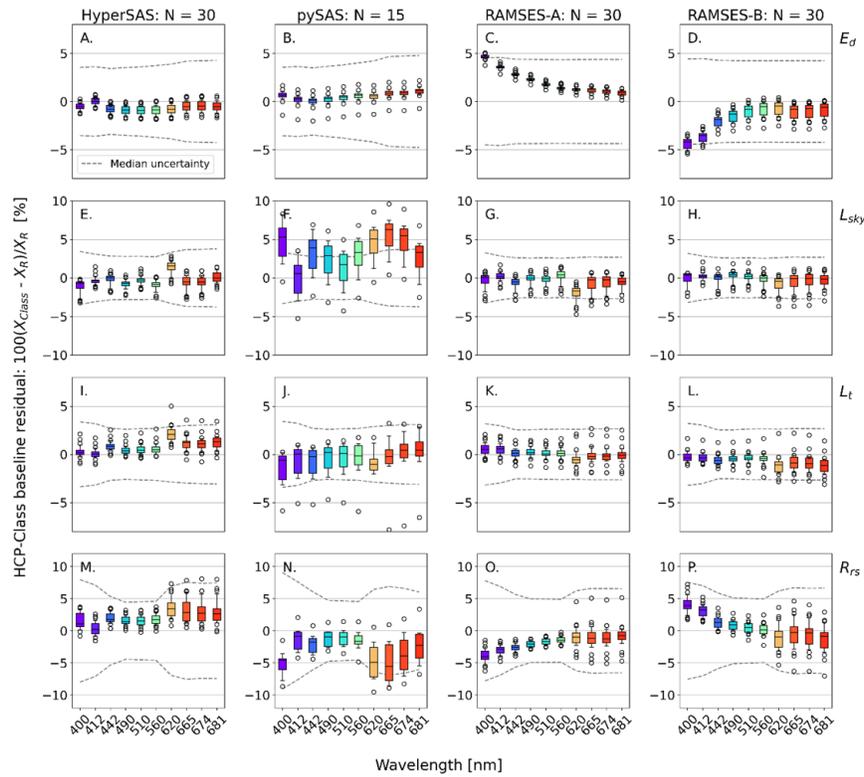


Fig. 3. Percent residuals of $E_d(\lambda)$ (A.-D.), $L_{sky}(\lambda)$ (E.-H.), $L_t(\lambda)$ (I.-L.), and $R_{rs}(\lambda)$ (M.-P.) using HyperCP in class-based mode compared to the mean of all systems for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.), RAMSES-A (C., G., K., L.), and RAMSES-B (D., H., L., P.), respectively. The residuals at each wavelength are calculated for each system as $[(X - X_R)/X_R] * 100$, where X_R is the mean reference value across the available sensor systems for each cast and $X = E_d, L_{sky}, L_t$, and R_{rs} , respectively. The boundary of the box closest to zero indicates the 25th percentile, the solid line within the box is the median, the boundary of the box farthest from zero indicates the 75th percentile, the error bars above and below the box indicate the 90th and 10th percentiles and the points above and below the 90th and 10th percentiles are outliers. The dashed grey lines are the \pm values of the median propagated uncertainties for class-based processing (given in Fig. 4).

from 400 to 490 nm and 550 to 600 nm and a double peak near 490 and 535 nm, indicating a change in the optical water type (i.e., AVW 490–500 nm in Fig. 2(a)). There was an abrupt increase in a_{dg} on 19th July indicating significant influence of TSM and CDOM.

On 19th July, there was the passage of high *chl* water followed by a further increase in *chl* at the end of the campaign (Fig. 2(e)). During this period $R_{rs}(490)$ decreased (red spectra in Fig. 2(a)) with a shift in the peak from 490 nm to 535 nm.

3.2. Intercomparison of $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ between HyperSAS, RAMSES and pySAS processed using HyperCP (M99NN) in class-based and sensor-specific modes

Figure 3 (class-based processing) and Fig. 5 (sensor-specific processing) show the percentage residuals in the radiometric quantities with respect to a reference mean value, which is calculated by averaging across the available sensor systems (three or four) for each cast. Figures 4 and 6

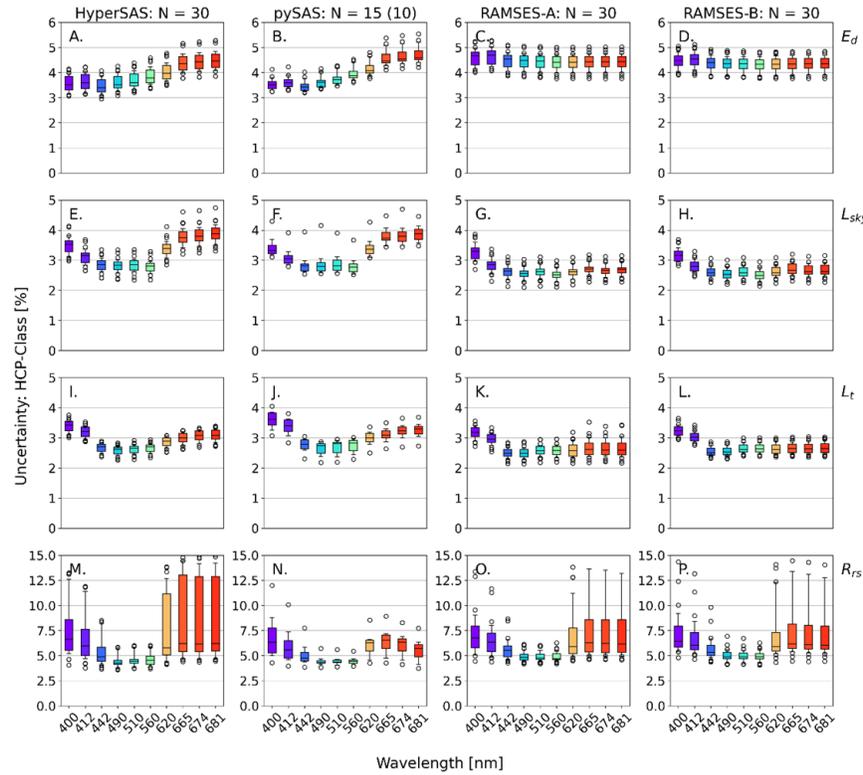


Fig. 4. Percentage propagated uncertainties for $E_d(\lambda)$ (A.-D.), $L_{sky}(\lambda)$ (E.-H.), $L_t(\lambda)$ (I.-L.), and $R_{rs}(\lambda)$ (M.-P.), computed using HyperCP in class-based mode compared to the mean of all systems for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.) RAMSES-A (C., G., K., O) and RAMSES-B (D., H., L., P.).

show the corresponding propagated uncertainties for class-based and sensor-specific processing. The percentage residuals quantify the internal consistency of the field measurements, and are therefore a natural point of comparison for the propagated uncertainties from HyperCP. For the HyperCP class-based and sensor-specific processing, there were $N = 30$ measurements of $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ that passed QC for RAMSES-A, -B and HyperSAS, and $N = 15$ and $N = 17$ measurements respectively, for pySAS (Fig. 3, 4, 5, 6). The differences in N for pySAS were due to a combination of a delay in the system being operational after setup, as well as azimuth restrictions applied to match the azimuth angles of the manually operated systems. A slightly different location of the pySAS on the AAOT compared to the other systems, prevented pySAS performing some the measurement sequences made by the manually operated systems. HyperCP class-based $E_d(\lambda)$ processing for HyperSAS and pySAS exhibited small differences over the visible bands compared to the mean of the four systems (Fig. 3(a),(b)). For RAMSES-A and -B, the differences in $E_d(\lambda)$ were higher: $\sim 2\%$ in the red, and up to 5% in the blue (Fig. 3(c),(d)). For class-based $L_{sky}(\lambda)$ processing, the differences for HyperSAS, RAMSES-A and -B compared to the mean of all systems were generally $< 1\%$ (Fig. 3(e),(g),(h)), except at 620 nm which were $< 2\%$. For pySAS, the differences in $L_{sky}(\lambda)$ were consistently higher ($1\text{--}6\%$) and higher than the mean uncertainties, except at 412 and 681 nm (Fig. 3(f)). For class-based processing of $L_t(\lambda)$, the differences were $< 1.0\%$ and lower than the median uncertainties (Fig. 3(i)–(l)). The range in pySAS $L_t(\lambda)$ compared to the mean was greater than the other systems and the outliers were greater than the median uncertainties (dashed lines in Fig. 3(j)). For class-based $R_{rs}(\lambda)$, the

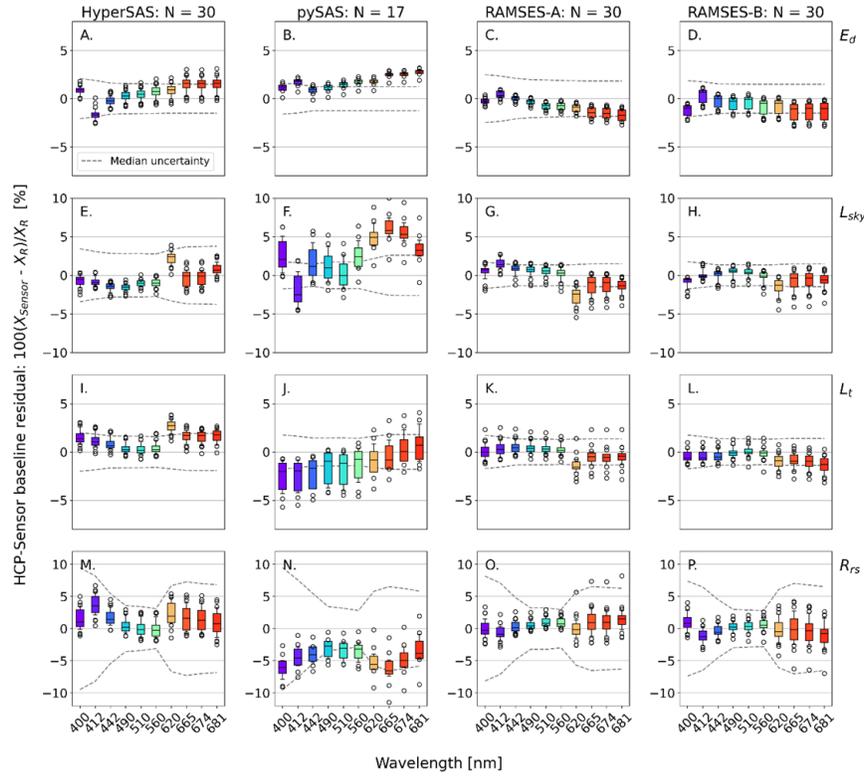


Fig. 5. Percent residuals of $E_d(\lambda)$ (A–D.), $L_{sky}(\lambda)$ (E–H.), $L_t(\lambda)$ (I–L.), and $R_{rs}(\lambda)$ (M–P.) using HyperCP in sensor-specific mode compared to the mean of all systems for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.), RAMSES-A (C., G., K., L.), and RAMSES-B (D., H., L., P.), respectively. The residuals at each wavelength are calculated for each system as $[(X - X_R)/X_R] * 100$, where X_R is the mean reference value across the available sensor systems for each cast and $X = E_d, L_{sky}, L_t,$ and R_{rs} , respectively. The boundary of the box closest to zero indicates the 25th percentile, the solid line within the box is the median, the boundary of the box farthest from zero indicates the 75th percentile, the error bars above and below the box indicate the 90th and 10th percentiles and the points above and below the 90th and 10th percentiles, are outliers. The dashed grey lines are the median propagated uncertainties for sensor-specific processing given in Fig. 6.

differences from the mean of all systems were $<5\%$ (Fig. 3(m)–(p)), except for pySAS at red bands, which were $<7.5\%$ (Fig. 3(n)).

The differences in $E_d(\lambda)$ for all systems processed using class-based HyperCP were within their median uncertainties (dashed line in Fig. 3(a)–(d)), which were 3.5 to 4.5% across all bands for all systems (Fig. 4(a)–(d)). The class-based $L_{sky}(\lambda)$ uncertainties for all systems were between 2.5 and 4% (Fig. 4), which for RAMSES-A and -B exhibited an asymptotic curve from blue to red (Fig. 4(g), (h)). For HyperSAS and pySAS the class-based $L_{sky}(\lambda)$ uncertainties were higher in the red compared to the other systems (Fig. 4(e), (f)). The class-based $L_t(\lambda)$ uncertainties were well constrained for all systems and between 2.5 and 3.5% (Fig. 4). The class-based $R_{rs}(\lambda)$ uncertainties were generally $<7\%$, though the range in red bands for HyperSAS, RAMSES-A and -B were up to $\sim 15\%$, nevertheless the 75th percentile was much less (Fig. 4).

For the HyperCP sensor-specific processing, the differences in $E_d(\lambda)$ for all systems were $<2\%$ compared to the mean, except for HyperSAS and pySAS were slightly higher at 412 and at 681

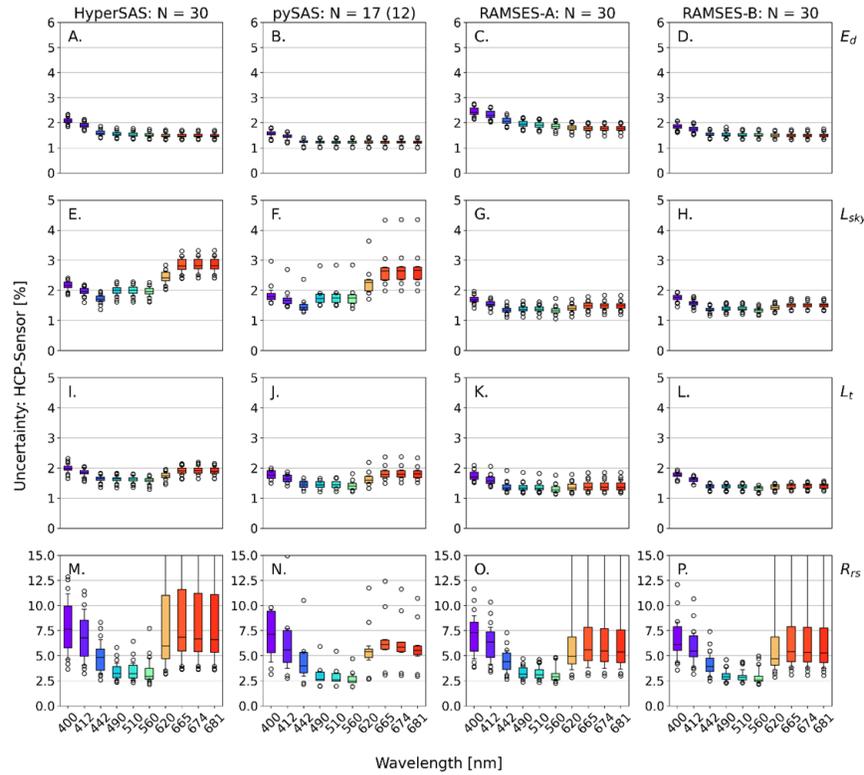


Fig. 6. Percentage propagated uncertainties for $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ computed using HyperCP in sensor-specific mode for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.) RAMSES-A (C., G., K., O) and RAMSES-B (D., H., L., P.).

nm reaching 2.5% (Fig. 5(a)–(d)). For the sensor-specific processing, the difference in $L_{sky}(\lambda)$ for HyperSAS, RAMSES-A and -B were $<2\%$ and within the uncertainties, except for RAMSES-A at 620 nm which was slightly higher (Fig. 5(e)–(h)). For pySAS, the differences from 442 to 560 nm were similar, but at 412 nm were $<3\%$ and at red bands were 4 to 6% and higher than the uncertainties (dashed lines in Fig. 5(b)). For sensor-specific processing of $L_t(\lambda)$, the differences were generally $<1\%$, except for pySAS from 400 to 442 nm which were $<2\%$ (Fig. 5(i)–(l)). The differences in sensor-specific $R_{rs}(\lambda)$ for RAMSES-A and -B compared to the mean of the systems were $<1\%$, for HyperSAS were $<2\%$ and for pySAS were $<6\%$ due to the higher residuals in $L_{sky}(\lambda)$ and $L_t(\lambda)$ (Fig. 5(m)–(p)).

The sensor-specific uncertainties in $E_d(\lambda)$ for the systems were $<2\%$ at all bands, except RAMSES-A at 400 and 412 nm which were slightly higher (Fig. 6(a)–(d)). The sensor-specific $L_{sky}(\lambda)$ uncertainties for all systems and bands were also generally $<2\%$, though HyperSAS and pySAS at red bands were $<3\%$ (Fig. 6(e)–(h)). For $L_t(\lambda)$, sensor-specific uncertainties were between 1 and 2% (Fig. 6(i)–(l)). For $R_{rs}(\lambda)$, the sensor-specific values were $<3\%$ from 490 to 560 nm, and $<7.5\%$ at blue and red bands (Fig. 6(m)–(p)).

Using HyperCP for class-based versus sensor-specific processing of $E_d(\lambda)$, the differences were generally $<5\%$ though there were exceptions to this for some systems at specific bands (Fig. 7(a)–(d)). HyperSAS $E_d(\lambda)$ at 400 nm was just above 5% (Fig. 7(a)), whilst for RAMSES-B $E_d(\lambda)$ at both 400 and 412 nm the differences were 7 and 6%, respectively (Fig. 7(d)). Similarly, the differences between class-based versus sensor-specific processing of $L_{sky}(\lambda)$ using HyperCP were also $<5\%$ for all systems (Fig. 7(e)–(l)). The differences for pySAS and RAMSES-B $L_{sky}(\lambda)$ were notably lower than the other systems, and $<2\%$ (Fig. 7(f), (h)). For $L_t(\lambda)$, the differences

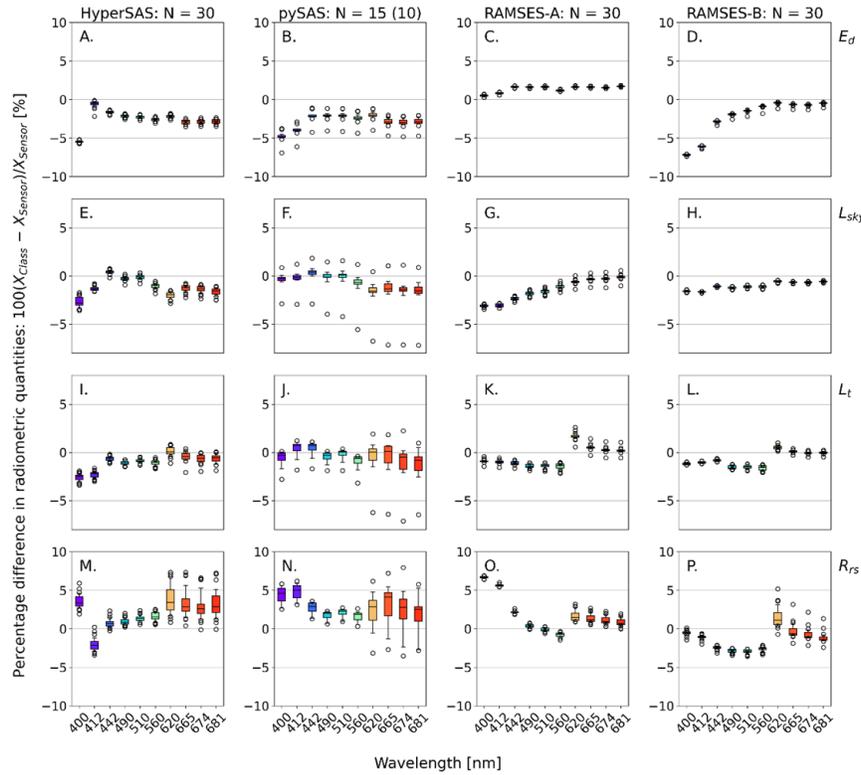


Fig. 7. Differences in $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ processed using HyperCP in class-based and sensor-specific mode for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.), RAMSES-A (C., G., J., O.) and RAMSES-B (D., H., K., P.).

were generally $<2\%$, except for HyperSAS at 400 and 412 nm which were $\sim 2.5\%$ (Fig. 7(i)–(l)). The differences in $R_{rs}(\lambda)$ between class-based and sensor-specific HyperCP processing were generally $<5\%$ across all bands, though for RAMSES-A at 400 and 412 nm the differences were slightly higher (Fig. 7(m)–(p)). The percentage differences in uncertainties between the class-based versus sensor-specific uncertainties for each radiometric quantity was far more evident (Fig. 8). The differences were greatest for $E_d(\lambda)$, which were lower at blue bands varying from 1.5% for HyperSAS and 2.5% for RAMSES-B, and higher for red bands varying from $\sim 3\%$ to 3.5% for HyperSAS and pySAS (Fig. 8(b)). By comparison, the differences in uncertainties for $L_{sky}(\lambda)$ and $L_t(\lambda)$ for all systems were generally $<1.5\%$ across most wavebands (Fig. 8(e)–(l)).

The differences in uncertainties for pySAS for both $L_{sky}(\lambda)$ and $L_t(\lambda)$ were slightly higher and $<2\%$. For $R_{rs}(\lambda)$, the differences between class and sensor-specific uncertainties for all systems were $<2.5\%$ and lower ($<1\%$) in blue and red bands (Fig. 8(m)–(p)).

3.3. Intercomparison of $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ between HyperSAS, RAMSES and pySAS, processed using individual lab processors and HyperCP

The difference between individual lab processors (IP) and HyperCP class-based processing of $E_d(\lambda)$ for all systems were $<1\%$ (Fig. 9(a)–(d)). As expected, due to the similar processing, pySAS was closest to HyperCP and the differences across nearly all bands, except 400 nm, was minimal (Fig. 9(a)). Class-based $L_{sky}(\lambda)$ were within 1% of the IP for all systems, though the range in the individual values for RAMSES-A and -B were greater than for HyperSAS and pySAS (Fig. 9(e)–(h)). The differences between IP and HyperCP class-based $L_t(\lambda)$ were also generally $<1\%$, except for RAMSES-A and RAMSES-B at red bands which were $<4\%$ (Fig. 9(k),(l)), due to differences in spectral outlier removal for pronounced capillary glitter (see Section 2.9). For the

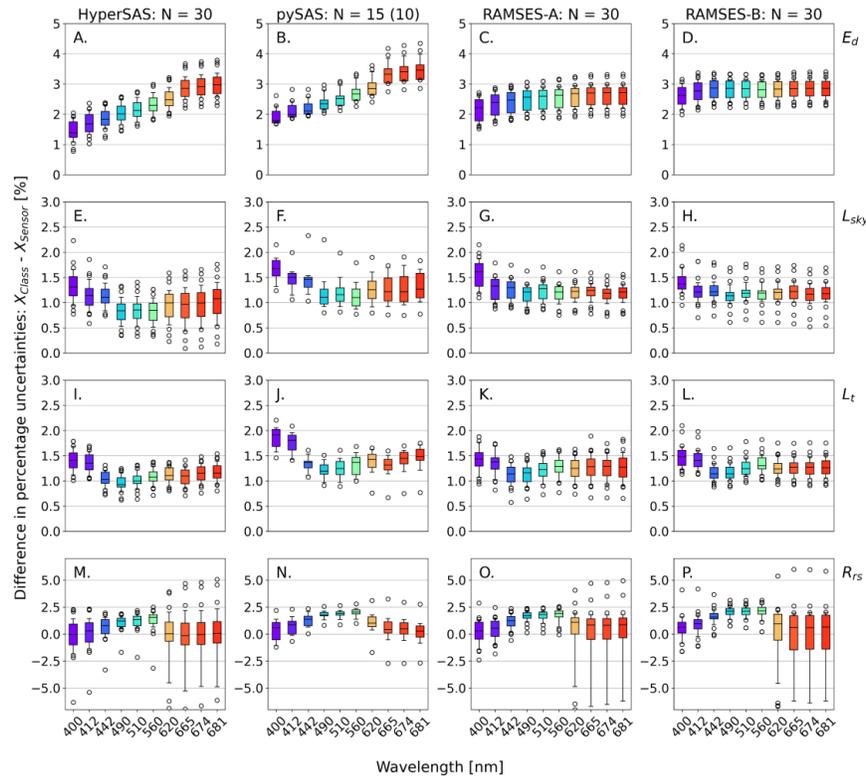


Fig. 8. Percentage differences in uncertainties between the class-based (given in Fig. 4) and sensor-specific (given in Fig. 6) uncertainties for $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ processed using HyperCP in class-based and sensor-specific mode for HyperSAS (A., E., I., M.), pySAS (B., F., I., N.), RAMSES-A (C., G., J., O.) and RAMSES-B (D., H., K., P.).

comparison of $R_{rs}(\lambda)$ computed using class-based HyperCP and IP, the difference for HyperSAS and pySAS were small and typically $<1\%$ (Fig. 6(i)–(l)). The differences for RAMSES-A and -B were higher and especially at red bands were $<7\%$. This is likely to result from differences in the RAMSES IP processing solutions and input parameters (e.g., ρ).

3.4. Comparison of HyperCP (M99NN & Z17SS) $L_{wn}(\lambda)$ for HyperSAS, RAMSES, pySAS against independent measurements

The HyperCP offers different processing options that include glint and NIR corrections for the computation of $L_{wn}(\lambda)$, as described in Section 2.9. For this, $L_{wn}(\lambda)$ from RAMSES, HyperSAS and pySAS processed using no NIR correction with the Mobley [1] correction (M99NN) was compared against SeaPRISM (Fig. 10), and HyperPRO II (Fig. 11). In these figures the residuals are defined with respect to an independent reference system and therefore enable separate assessment of $L_{wn}(\lambda)$. This differs from the calculated residuals for the other radiometric quantities in Figs. 3 and 5, which quantified the internal consistency of the measurements. Results indicate that all systems were within or near to the expected 5% of SeaPRISM at blue and green bands, but greater than this in the red (Fig. 10). The uncertainties in $L_{wn}(\lambda)$ for RAMSES, HyperSAS and pySAS were $<5\%$ from 442 to 560 nm, and $<7\%$ at 412, 620 and 665 nm (Fig. 10). By comparison, using HyperPro II as an independent source, the magnitude and range of the differences in the above-water systems processed using M99NN were generally higher and $<10\%$ (Fig. 11), but for red bands the differences were $>25\%$. The associated uncertainties for RAMSES, HyperSAS and pySAS processed using sensor-specific characteristics and M99NN

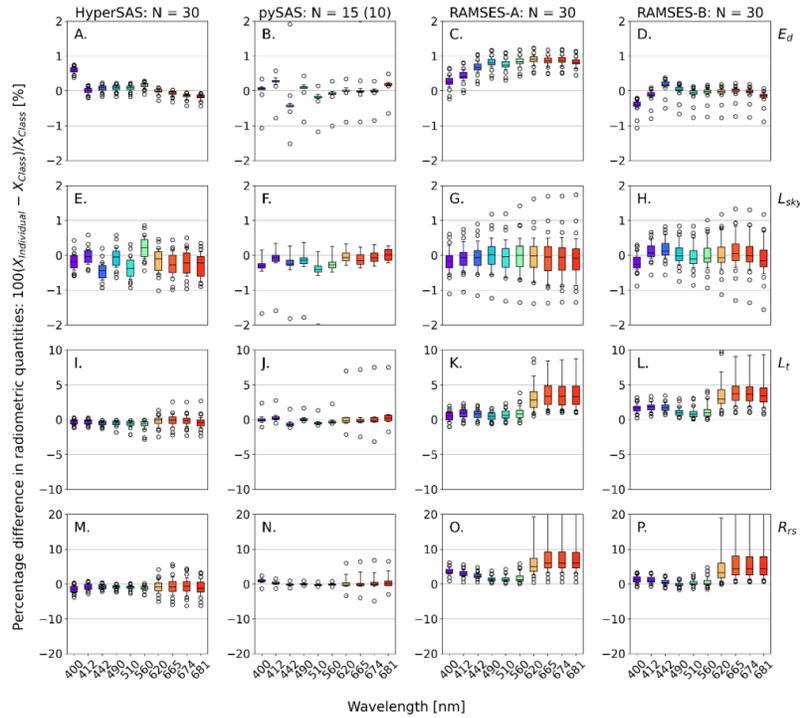


Fig. 9. Differences between individual processors and HyperCP for $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and $R_{rs}(\lambda)$ run in class-based mode for HyperSAS (A., E., I., M.), pySAS (B., F., J., N.) RAMSES-A (C., G., K., O) and RAMSES-B (D., H., L., P.).

were $\leq 5\%$ at 442 to 560 nm, but $< 7.5\%$ at 412, 620 and 665 nm (Fig. 11(e)–(h)). $L_{wn}(\lambda)$ from RAMSES, HyperSAS and pySAS were also computed using Zhang et al. [2], with sim spec NIR correction (Z17SS). Compared against HyperPro II, the differences were lower and generally

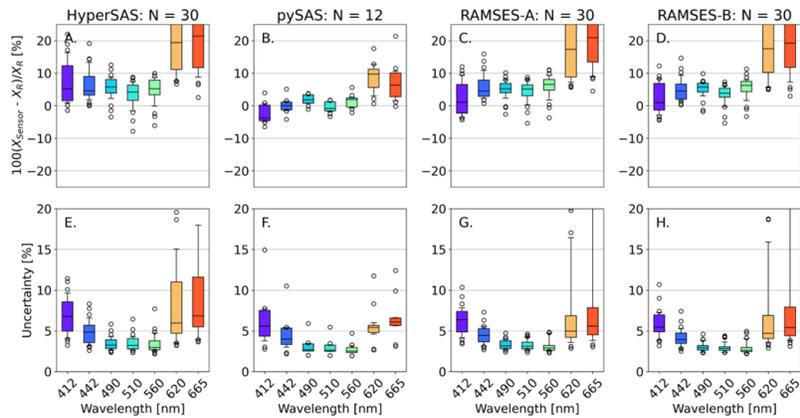


Fig. 10. Percent residuals of $L_{wn}(\lambda)$ for the above-water systems HyperSAS, RAMSES-A, RAMSES-B, pySAS computed with the Mobley (1999) glint correction and no NIR correction in sensor-specific mode against SeaPRiSM (A., B., C., D.). The residuals at each wavelength are calculated from each system as $[(L_{wn} - L_{wn}^R) / L_{wn}^R] * 100$, where L_{wn}^R is the reference value for SeaPRiSM. Percentage uncertainties for $L_{wn}(\lambda)$ computed using HyperCP in sensor-specific mode for HyperSAS, pySAS, RAMSES-A and RAMSES-B compared to SeaPRiSM (E., F., G., H.).

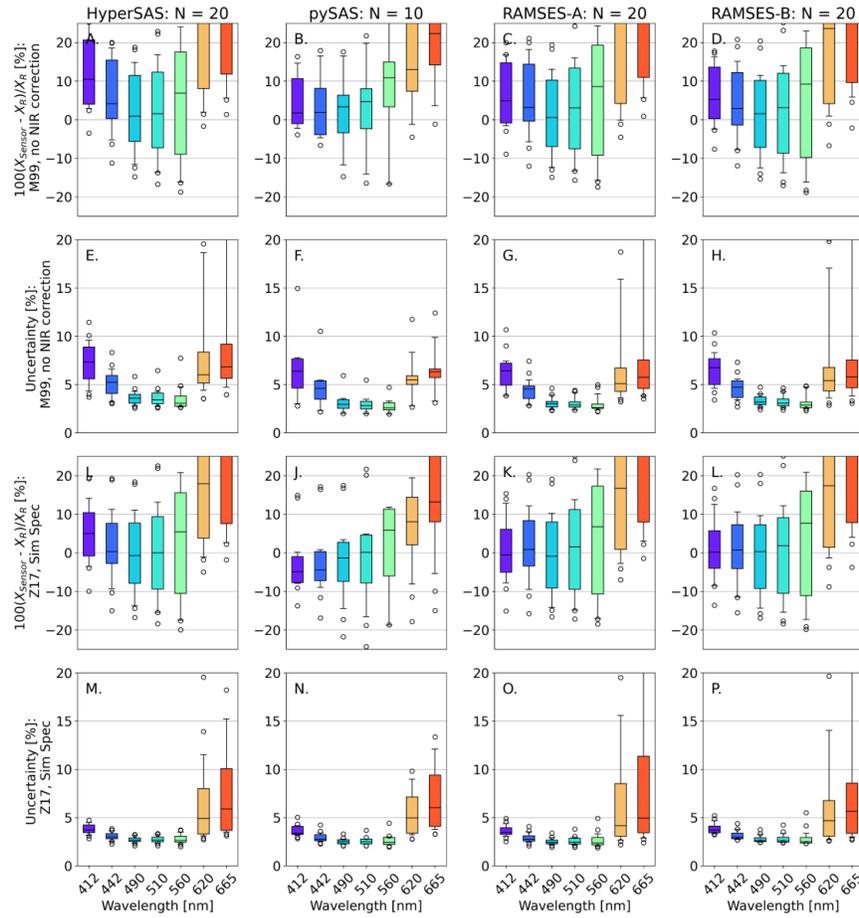


Fig. 11. Percent residuals of $L_{wn}(\lambda)$ for the above- water systems HyperSAS, RAMSES-A, RAMSES-B, pySAS computed with Mobley (1999) glint correction and no NIR correction in sensor-specific mode against HyperPro II (A., B., C., D.) and the percentage uncertainties for $L_{wn}(\lambda)$ (E.,F.,G.,H.). Percent residuals of $L_{wn}(\lambda)$ for each above- water system computed with Zhang et al. (2017) and sim spec NIR correction against HyperPro II (I., J., K., L.) and the percentage uncertainties for $L_{wn}(\lambda)$ (M.,N.,O.,P.). The residuals at each wavelength are calculated from each system as $[(L_{wn} - L_{wn}^R)/L_{wn}^R]*100$, where L_{wn}^R is the reference value for HyperPro II.

<4% for blue and most green bands, but were higher at 560 nm and at red bands (Fig. 11(i)–(l)).

The associated uncertainties were significantly reduced using Z17SS compared to M99NN at blue wavelengths. For example, at 412 nm, median uncertainties were reduced from ~8% to ~4% (Fig. 11(m)–(p)).

4. Discussion

4.1. Need for an open source radiometric community processor

The quantification and reduction of uncertainties for *in situ* measurements is a fundamental requirement for ocean color scientists [29]. An uncertainty threshold of 5% was originally defined as a mission requirement for satellite $L_{wn}(\lambda)$ at blue bands for SeaWiFS (EOSAT/NASA SeaWiFS Working Group, 1987), in order to achieve <35% uncertainty in OC *chl* in oligotrophic waters [30]. Subsequent OC missions adopted this as the target threshold, regardless of the

wavelength. From this, the need to improve the uncertainties in *in situ* optical radiometry data has repeatedly been highlighted [29]. Field inter-comparisons are a necessary step in the traceability chain of FRMs to ensure that radiometric measurements made by a single laboratory converge to independent quality controlled reference measurements. Since the launch of SeaWiFS, a range of radiometric inter-comparisons have been conducted, each of which resulted in a reduction in the differences and uncertainties between radiometric systems due to a consensus on protocols and methods as well as improvements in sensor calibration and characterization [e.g., 31,32]. The use of a wide range of instruments, methods and laboratory practices may only add to the uncertainty of satellite OC products. To improve on differences among laboratories, a number of studies have advocated the use of a single community processor to reduce the differences and uncertainties in radiometric quantities [6,7]. Tilstone et al. [7] showed that for TriOS-RAMSES systems, differences in $R_{rs}(\lambda)$ for individual instruments and processors were 3.5% in the blue and 11% for red bands, which were reduced to ~2% in the blue and 8% in the red if a single processor was used. Of the radiometric inter-comparisons conducted to date, only one, also carried out at the AAOT, reported the uncertainties for each system [6]. The study highlighted that there was <5.5% difference in $E_d(\lambda)$ at blue, green and red wavebands between two RAMSES sensors and the in-water WiSPER, but for $E_d(\lambda)$ specifically, the uncertainties were not reported [6]. Similarly, a comparison of $E_d(\lambda)$ between three RAMSES and two HyperSAS instruments at the same site showed that there was <6% between them, though one system not included in the mean exhibited a systematic bias due to a poor cosine response and $E_d(\lambda)$ uncertainties for all systems were not reported [7].

In the current study, by using the same processor there was <1% difference in $E_d(\lambda)$ between two RAMSES and two HyperSAS sensors (Fig. 5). For $R_{rs}(\lambda)$, Zibordi et al. [6] found that the difference between two RAMSES and the WiSPER system using similar number of QC casts (N = 28), was between 3 and 5% and the associated uncertainties for the RAMSES instruments were 6.3, 3.5 and 4.5% at 443, 555 and 665 m, respectively. The differences arise partly from environmental variability, and also from nuances in processing the radiometric data. Zibordi et al. [11] reported a large variation in inherent optical properties and associated biogeochemical quantities at the AAOT as it is influenced by both the open sea and coastal waters, with occasional significant riverine run-off from regional rivers, wind or tidal mixing. On an annual basis, approximately 60% of the $R_{rs}(\lambda)$ spectra are reported to be case-1 type waters and 40% case-2 type [33]. The $R_{rs}(\lambda)$ spectra observed in this study during the campaign (Fig. 2), captured these changes in the optical water types. The differences in $R_{rs}(\lambda)$ for RAMSES and HyperSAS sensors computed using class-based characteristics were <5% at all bands except for pySAS at the red bands, which was <7.5% (Fig. 3) and these were further reduced using sensor-specific characteristics in the processing (Fig. 5). The differences between class-based and sensor-specific processing were greater for RAMSES-A at blue bands compared to RAMSES-B, which indicates higher instrument related correction were applied for RAMSES-A based on sensor-specific characterization files.

In summary, by using the same common processor and settings, differences between $E_d(\lambda)$, $L_{sky}(\lambda)$ and $L_t(\lambda)$ and the associated uncertainties were reduced compared to previous inter-comparisons to at least <3%, over the entire spectral range (except pySAS in the red) when sensor-specific characteristics were used for the data processing. For $R_{rs}(\lambda)$, the differences were <5% from 442 to 560 nm, which is the target over the blue to green spectral range. The differences in $R_{rs}(\lambda)$ for 400, 412 and 620 to 681 nm were <7.5% (Fig. 6).

4.2. Computation of uncertainties in radiometric measurements

Numerous databases exist that contain radiometric data for the validation of remotely sensed OC data products [34]. As the accuracy of satellite sensors have evolved through improvement in signal to noise ratios of sensors and in atmospheric correction models, there is an on-going

demand for an increase in the accuracy of *in situ* radiometric measurements for satellite OC validation. In SeaBASS however, there are few radiometric data that include uncertainties. The NASA EOS project established uncertainties through detailed absolute radiometric calibration. For NASA PACE validation, radiometric data now has to be submitted with the associated uncertainties. Assessment of uncertainties for the principle radiometric systems deployed on stationary platforms for use in radiometric vicarious calibration and validation of OC satellites are well documented. The Marine Optical BuoY (MOBY), deployed in the Pacific Ocean off Hawaii is the principle *in situ* radiometric capacity used for system vicarious calibration of NASA and NOAA OC satellites. It has a combined uncertainty in upwelling radiance $L_u(\lambda)$ measured in surface waters lower than 3% over most MODIS bands [35], excluding contributions from environmental variability. For BOUSSOLE, which deployed underwater fixed-depth Satlantic OCR-200-series radiometers also targeting the vicarious calibration of satellite ocean color sensors, uncertainties of <4% over the blue and green spectral range and ~5% in the red have been estimated [36]. For AERONET-OC SeaPRISM CIMEL CE-318 9-band sun photometers that are deployed at an array of stable platforms in coastal waters in the global ocean as the principle source of OC satellite validation data, the median relative combined uncertainties are 5% in blue and green bands ~7% for red bands [37]. For the newer SeaPRISM CIMEL CE-318 T model, these are <5% at blue and green bands and 8% in the red [38]. Currently, many radiometric systems used in the field for satellite ocean color validation, uncertainties are not documented. This lack of information is even more pronounced for hyperspectral systems that require more complex and comprehensive radiometric characterization. For example, an average uncertainty of ~10% in $L_{wn}(\lambda)$ in blue bands and ~15% in the red is reported for Seabird HyperSAS and TriOS-RAMSES sensors deployed along an Atlantic Meridional Transect [14,39]. Similarly, for the DALEC system deployed on a meridional transect in the Pacific Ocean off Western Australia, uncertainties of 10 to 15% for $\lambda < 600$ nm are reported based on comparison with coincident Seabird HyperPro II deployments [40].

Uncertainties in $R_{rs}(\lambda)$ arise from: A.) Instrument characterization and absolute calibration; B.) Deviations from recommended data acquisition protocols impacting quality assurance; C.) Environmental perturbations and different measurement conditions, such as optical water types, sky radiance distribution, cloudiness; D.) data reduction and processing. With careful and traceable procedures followed for A.) and B.), to minimize the related uncertainties, still the contributions from environmental variability and data processing methods can be large [29]. Building on former analyses, the need for more frequent comparisons and robust measurement uncertainty propagation has been highlighted [41]. For some classes of instruments, factory correction factors to compute uncertainties do not exist [40], and even calibration coefficients are provided without associated uncertainties. The minimization of uncertainties can thus be achieved through extensive laboratory characterizations, as there can be considerable variation in these between instruments of a particular class [23]. Sensor-specific correction factors can then be applied to calibration coefficients. These characterizations can only be carried out by specific calibration laboratories and are expensive. Białek et al. [41] showed that sensor-specific uncertainties in $E_d(\lambda)$ for a RAMSES sensor at 442 nm were ~7% but this was reduced to 1% when sensor-specific characteristics were utilized. The question remains as to whether class-based characterizations are sufficient to obtain uncertainties within desired ranges.

In this study, using sensor-specific characteristics, uncertainties <2% at OLCI bands in the visible spectrum were determined for $E_d(\lambda)$, <3% for $L_{sky}(\lambda)$, <2% for $L_t(\lambda)$ and between 2.5% and 7.5% for $R_{rs}(\lambda)$ (Fig. 6), in principle satisfying validation requirements for satellite data products (as reviewed above). Using class-based characteristics, the uncertainties in $E_d(\lambda)$ were <5%, <4% for $L_{sky}(\lambda)$, <3.5% for $L_t(\lambda)$ and <7% for $R_{rs}(\lambda)$ (Fig. 4). The reduction in percentage uncertainties from class-based to sensor-specific processing were up to 3% for $E_d(\lambda)$, <2% for

$L_{sky}(\lambda)$ and $L_t(\lambda)$ and $<2.5\%$ for $R_{rs}(\lambda)$ (Fig. 8). The differences between class-based and sensor-specific uncertainties arise from the application of corrections for the non-ideal performance of individual radiometers, determined for each individual instrument (i.e., sensor-specific). The principal differences in sensor characteristics between HyperSAS and RAMSES that contribute to the uncertainties, are considered below.

For $E_d(\lambda)$, the response of the radiometer to individual radiance contributions should theoretically be determined by the cosine of the incident angle. This is rarely the case and departure from the cosine response has to be quantified and corrected for. Ideally this departure would be characterized for each instrument, so that a correction for the effect can be applied, and then only a residual uncertainty in that correction would be propagated [32]. Comparing the characteristics of the sensors used in this study, the HyperSAS cosine collector is larger than that of RAMSES sensors, which suggests that a better cosine response could be heuristically assumed for HyperSAS. Consequently, differences between HyperSAS and RAMSES irradiance sensors and the resulting $E_d(\lambda)$, may arise from differences in the cosine response at blue bands. These are $\sim 5\%$ at 400 and 412% RAMSES and close to 0% for HyperSAS and pySAS (Fig. 3(a)–(d)). If the non-ideal performance of individual instruments are characterized, the related effects are reduced as can be clearly seen from the sensor-specific processing (Fig. 5(a)–(d)). When the cosine response is accounted for, the maximum differences due to this could approach 1% [32]. Azimuthal differences in the cosine response of collectors were not accounted for in this study, which should be a potential area of investigation in future inter-comparisons.

The diffuser material used in the construction of HyperSAS irradiance collectors however, can cause a strong hysteresis in the instrument response associated with changes in temperature, which is difficult to compensate for over ambient environmental conditions and can lead to significant errors [42]. By comparison, the cosine error can account for $\pm 10\%$ of the uncertainty [17], and is larger for RAMSES compared to HyperSAS [42]. Similarly the range in non-linearity of RAMSES is greater than in HyperSAS, which undoubtedly contributes to further measurement errors.

Temperature affects the optics, electronics, and mechanics of a radiometer and therefore perturbs measurements. Temperature can also affect darks counts (i.e., implemented through a shutter in HyperSAS and by shaded elements of the detector array in RAMSES). Additionally, internal operational temperatures and temperature coefficients are required to minimize thermal perturbations in radiometric measurements. HyperSAS has an internal temperature sensor, which allows tracking variations in the internal temperature and correcting for it. Though temperature responsivity was determined for RAMSES, its practical application is prevented by the lack of an internal temperature sensor. The mean temperature coefficients for RAMSES varies from $-0.04 \times 10^{-2} \text{ }^\circ\text{C}^{-1}$ at 400 nm to $0.33 \times 10^{-2} \text{ }^\circ\text{C}^{-1}$ at 800 nm [43]. By comparison, the mean temperature coefficients for HyperSAS varies from $-0.15 \times 10^{-2} \text{ }^\circ\text{C}^{-1}$ to $0.2 \times 10^{-2} \text{ }^\circ\text{C}^{-1}$ over the same temperature range [14]. These differences could lead up to a $\sim 1\%$ difference in $R_{rs}(\lambda)$ between HyperSAS and RAMSES [17,14].

When sensor-specific characteristics are used in processing of the data, differences in these between HyperSAS and RAMSES are minimized. Due to the magnitude of the reduction in uncertainty, it is therefore recommended using HyperCP with sensor-specific correction factors to compute the uncertainties to meet the mission requirement of 5% in blue bands. In this study, using sensor-specific correction factors, the differences in $R_{rs}(\lambda)$ between instruments was $<5\%$ across all bands and the uncertainties were $<5\%$ in the green and $<7\%$ in blue and red bands. A requirement for future studies is to determine which factors contribute most to the uncertainties, so that these can be prioritized in the characterization of individual sensors.

For specific instruments, correction factors are often determined once and applied consistently over the instrument lifetime, but in reality these may vary as the instrument ages [17]. The responsivity drift in RAMSES and HyperSAS over the spectral range from 400 nm to 800 nm, is

approximately -1% per year, though this depends on the use of the radiometers. To maintain them, recalibration after each deployment is advised with a minimum repeat calibration of a year to achieve FRM [42]. What is still unknown and requires further work is how much sensor-specific characteristics change over the lifetime of an instrument and therefore how frequently these need to be determined. Future studies should also focus on characterization of other instruments deployed for OC validation, such as DALEC [40], for which detailed sensor-specific characteristics currently do not exist.

4.3. Intercomparison of $L_{wn}(\lambda)$ using independent above and in-water reference systems

Using a weighted mean value from above-water sensors for the inter-comparison can potentially result in a bias. To reduce this, we used two separate systems as independent data sources for comparisons. The first is the AERONET-OC SeaPRiSM CE-318 T 12 band system and the second, the in-water Sea-Bird HyperPro II. The uncertainties of SeaPRiSM are well documented [37], the AERONET-OC SeaPRiSM systems are recognized to produce high quality radiometric products [12] and the data are therefore widely applied for the validation of OC products [26]. The instrument has been characterized by NASA and NIST for linearity and calibrated for absolute radiance responsivity with uncertainties of 1.88% [44]. A previous inter-comparison between SeaPRiSM $L_{wn}(\lambda)$ against RAMSES and HyperSAS systems showed that both systems tended to underestimate $L_{wn}(\lambda)$, which for RAMSES was $<8.0\%$ at 441 nm, $<6.0\%$ at 551 nm and $<9.5\%$ at 667 nm and for HyperSAS was -1.4 to 5.5% at 441 nm, -4 to -7.5% at 551 nm and $<5.0\%$ at 667 nm [7]. However, there have been biases affecting SeaPRiSM measurements during a specific deployment [12]. In the current analysis, the SeaPRiSM $L_{wn}(\lambda)$ compared to RAMSES, HyperSAS and pySAS values determined using M99NN were within or near to the expected 5% in blue and green bands, but greater than this in the red except for pySAS. These differences may be due to a potential underestimate of SeaPRiSM $L_{wn}(\lambda)$ due to strict filtering criteria of $L_t(\lambda)$ values affected by glint perturbations [12].

HyperPro II (and variants thereof including LoCNESS, microNESS, SeaFall, SPRM) have been providing high quality $L_{wn}(\lambda)$ since the dawn of SeaWiFS [45]. Many inter-comparisons with HyperPro II and above-water systems have been conducted. Of these, using a stable platform from which the super-structure effects have been well characterized, Hooker et al. [31] compared $L_{wn}(\lambda)$ from SeaPRiSM and a HyperPro type instrument (microNESS) at the AAOT under Case-1 and Case-2 conditions, and reported a 9.1% difference with uncertainties of $\sim 3\%$. A series of inter-comparisons between above-water and in-water systems have been made on ships at sea where uncertainty from environmental perturbations and errors by super structure perturbations may affect measurements. On a research cruise through the Mediterranean Sea and off the NW coast of Africa, Hooker and Morel [46] compared $L_{wn}(\lambda)$ between LoCNESS and SeaSAS (a predecessor of HyperSAS). They reported large differences, even after strict QC after which only half of the dataset was used to obtain a 5% difference, mainly because the ships superstructure libedo was so high. Comparing the above-water HyperSAS and RAMSES systems against SeaPRiSM, which is also an above water system, differences are lower than when comparing against HyperPRO II in-water systems. There are inherent differences between in-water and above-water systems. As well as being sensitive to meteorological conditions, above-water radiometry requires the application of glint correction methods, which may not accurately represent the actual measurement conditions. Conversely, in-water profiles of upwelling radiance are not largely affected by sun-glint. In-water measurements of the upwelling radiance can however, be effected by illumination changes over depth, perturbations in the deployment of the sensor due to tilt, self-shading, super-structure or ship shadow, air bubbles and wave-induced effects [47]. In this study, using HyperPro II as an independent source of data, the magnitude and range of the differences were much higher than expected especially for red bands, which

for HyperSAS and RAMSES were $>25\%$. Though the inter-comparison was conducted in near ideal conditions, both differences in environmental perturbations away from the platform may have contributed to these differences. Also, the tilt of the $E_d(\lambda)$ sensor operated on the boat was not accounted for whilst HyperPro II was profiled through the water column, and $E_s(\lambda)$ outliers were removed from the processing. A linear regression of the $E_d(\lambda)$ values collected during the cast, was used to average the data during the multi-cast. The above sources of errors and the low values of the upwelling radiance in the red bands, most likely explain the differences observed.

4.4. Glint correction options for above-water radiometry in HyperCP

A correction to remove the surface-reflected light (arising from both sky and sun glint) from $L_t(\lambda)$, so that $L_{wn}(\lambda)$ and $R_{rs}(\lambda)$ that can be accurately computed, is essential [26]. The M99NN method is widely used for above-water measurement systems, whereby the dependence of the sea surface reflectance ρ on sky conditions, wind speed, solar zenith angle, and viewing geometry was determined. For a viewing direction of 40° from the nadir and 135° from the sun, Mobley [8] derived a ρ value of 0.028 for wind speeds $<5\text{ms}^{-1}$ and when the sky is overcast, is used for all wind speeds. For higher wind speeds, curves were derived for the determination of ρ as a function of solar zenith angle and wind speed. The method assumes that ρ is spectrally constant, but more recently it has been shown that ρ can vary spectrally, which can propagate large errors in computing $R_{rs}(\lambda)$ and $L_{wn}(\lambda)$ [48]. To minimize the impact of sun glint on the derivation of $L_{wn}(\lambda)$ (or $R_{rs}(\lambda)$), Hooker et al. [31] and Zibordi et al. [14] devised a method of filtering out the higher $L_t(\lambda)$ values, which improved the derivation of $L_{wn}(\lambda)$. A number of alternative approaches for correcting surface-reflected light have therefore been proposed. The most widely used assume that the near-infrared (NIR) water-leaving signal is zero, and thus $L_t(\lambda)$ signal in the NIR is entirely due to glint [31]. This is not the case in optically complex waters and a number of alternative, spectrally dependent approaches have been derived. These include a spectral optimization scheme [48] inversion algorithms based on inherent optical properties to correct for glint [49] and a two-step procedure for correcting both unpolarized and polarized glint from direct and diffuse sun light as a function of wind speeds, sun angles and aerosol concentrations [2].

To illustrate the differences between M99NN and an alternative approach that accounts for spectral variation in ρ (Z17SS), HyperSAS, pySAS and RAMSES $L_{wn}(\lambda)$ were computed using both methods and compared against HyperPro II (Fig. 11). The percent residuals in the blue and green using the Zhang et al. [22] method, are reduced and the asymptotic spectral dependency of the residuals disappears, but the differences still remain the same as for the M99NN method at red bands. This further suggests that error compensations may drive inter-comparisons, which often cannot lead to conclusions without extensive uncertainty analysis.

5. Summary

This field intercomparison illustrated that using the open-source HyperCP, the percentage difference of the radiometric quantities compared to individual processors was very low and generally $<1\%$. These results could be, however, affected by error compensations, which are difficult to address without a comprehensive uncertainty analysis. Because of this, the use of HyperCP supporting the correction for biases due to the non-ideal performance of radiometers (as determined through instrument characterizations) and additionally the quantification of the related uncertainties, is a major advance. The differences between above-water systems processed with HyperCP, using sensor-specific radiometric characteristics and M99NN, were 2% for $E_d(\lambda)$, $L_{sky}(\lambda)$, $L_t(\lambda)$ and 2.5% for $R_{rs}(\lambda)$. The associated uncertainties were 1.5%, 2%, 1.5% and 5%, respectively. For class-based radiometric characteristics, differences in the radiometric quantities were $<5\%$, but the uncertainties, especially in $R_{rs}(\lambda)$ were higher and were not sufficient to keep the uncertainties to $<5\%$. The uncertainties in the radiometric quantities could

be appreciably reduced if sensor-specific rather than class-based correction factors were used. When comparing HyperSAS and RAMSES $L_{wn}(\lambda)$ processed using M99NN against SeaPRISM, the differences were <5% over all bands, and the uncertainties were <5% over blue and green bands. Compared to HyperPro II, the magnitude and range of the differences in the above-water systems processed using M99NN were higher and <10% for blue and green bands. For the above water systems processed using Z17SS and compared to HyperPro II, the differences were lower and generally <4% for blue and green bands. Processing data with the HyperCP using sensor-specific characteristics is highly recommended to reduce the associated uncertainties, ensure processing workflow compatibility with standard community consensus and practices, and to produce the highest quality data for satellite OC validation.

Funding. European Organization for the Exploitation of Meteorological Satellites (EUM/CO/21/460002539/JIG); Natural Environment Research Council (NE/Y005589/1); Department for Science, Innovation and Technology (HyperCP); National Aeronautics and Space Administration (80NSSC22M0001); Joint Polar Satellite System (ST13301CQ0050/1332KP22FNEED0042).

Acknowledgments. This research was funded by the grant 'Fiducial Reference Measurements for Satellite Ocean Color (FRM4SOC) Phase-2' (EUMETSAT contract no. EUM/CO/21/460002539/JIG) funded by European Commission (Copernicus Programme) and implemented by EUMETSAT. AB and AR were also funded by the UK Government's Department for Science, Innovation & Technology through the UK's National Measurement System programmes, for the development of HyperCP. CK and MO were funded by the Joint Polar Satellite System (grant number ST13301CQ0050/1332KP22FNEED0042). GZ was supported by NASA through Southeastern Universities Research Association, Goddard Earth Sciences Technology and Research II Award (grant number 80NSSC22M0001). GT and TJ were also supported by The Atlantic Meridional Transect is funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Atlantic Climate and Environment Strategic Science - AtlantiS (grant number NE/Y005589/1). This study contributes to the international IMBeR project and is contribution number 417 of the AMT programme.

We are grateful to the crew of *RV Litus* and to Angela Pomaro for logistics on the AAOT. Hereon thanks Henning Burmester who helped with campaign preparation, logistics and packing.

Disclosures. The authors declare no conflicts of interest.

Data availability. The data results presented in this paper are available in the EUMETSAT/Copernicus Ocean Colour Database (OCDB accessible online) under submission(s) "FICE-22_FRM4SOC-2" and can be obtained from the authors.

References

1. S.B. Hooker, S. McLean, J. Sherman, *et al.*, "The Seventh SeaWiFS Intercalibration Round-Robin Experiment (SIRREX-7)," *NASA Tech. Memo.* NASA/TM-2002-206892/VOL17, (NASA Goddard Space Flight Center, 2002) pp. 78.
2. B.C. Johnson, H.W. Yoon, S.S. Bruce, *et al.*, "The fifth SeaWiFS Intercalibration Round Robin Experiment (SIRREX-5)," July 1996; *NASA Tech. Memo.* 1999-206892, Vol. 7, S.B. Hooker and E.R. Firestone, eds. (NASA Goddard Space Flight Center, 1999) pp. 75.
3. G. Meister, P. Abel, C. McClain, *et al.*, "The first SIMBIOS Radiometric Intercomparison (SIMRIC-1)," April-September 2001, *NASA Tech. Memo.* (2002).
4. G. Meister, P. Abel, K. Carder, *et al.*, "The Second SIMBIOS Radiometric Intercomparison (SIMRIC-2)," March-November 2002, *NASA Tech. Memo.* 210006, 2, 1-65 (2003).
5. G.H. Tilstone, G.F. Moore, K. Sorensen, *et al.*, "Regional Validation of MERIS Chlorophyll Products in North Sea Coastal Waters: MAVT inter-calibration results," in *Proceedings from ENVISAT MAVT Conference* (European Space Agency, 2004).
6. G. Zibordi, K. Ruddick, I. Ansko, *et al.*, "In situ determination of the remote sensing reflectance: An intercomparison," *Ocean Sci.* **8**(4), 567–586 (2012).
7. G. H. Tilstone, G. Dall'Olmo, M. Hieronymi, *et al.*, "Field Intercomparison of Radiometer Measurements for Ocean Colour Validation," *Remote Sens.* **12**(10), 1587 (2020).
8. C. D. Mobley, "Estimation of the remote-sensing reflectance from above-surface measurements," *Appl. Opt.* **38**(36), 7442–7455 (1999).
9. G. Thuillier, L. Floyd, T. N. Woods, *et al.*, "Solar irradiance reference spectra for two solar active levels," *Adv. Space Res.* **34**(2), 256–261 (2004).
10. A. Morel, D. Antoine, and B. Gentili, "Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle phase function," *Appl. Opt.* **41**(30), 6289–6306 (2002).
11. G. Zibordi, J.-F. Berthon, F. Melin, *et al.*, "Validation of satellite ocean color primary products at optically complex coastal sites: Northern Adriatic Sea, Northern Baltic Proper and Gulf of Finland," *Remote Sens. Environ.* **113**(12), 2574–2591 (2009).

12. G. Zibordi, B. N. Holben, M. Talone, *et al.*, “Advances in the ocean color component of the aerosol robotic network (AERONET-OC),” *J. Atmos. Ocean. Tech.* **38**(4), 725–746 (2021).
13. K. Uudeberg, I. Ansko, G. Pöru, *et al.*, “Using Optical Water Types to Monitor Changes in Optically Complex Inland and Coastal Waters,” *Remote Sens.* **11**(19), 2297 (2019).
14. J. Lin, G. Dall’Olmo, G. H. Tilstone, *et al.*, “Derivation of uncertainty budgets for continuous above-water radiometric measurements along an Atlantic Meridional Transect,” *Opt. Express* **30**(25), 45648–45675 (2022).
15. N. Haëntjens, K. Forsythe, B. Denholm, *et al.*, “pySAS: Autonomous solar tracking system for surface water radiometric measurements,” *Oceanography* **35**(2), 55–59 (2022).
16. M. Ondrusek, E. Stengel, V.P. Lance, *et al.*, “Report for Dedicated JPSS VIIRS Ocean Color Calibration/Validation Cruise,” *NOAA/NESDIS Tech. Report*, V.P. Lance, ed. (National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Washington DC, 2015) p. 60.
17. V. Vabson, J. Kuusk, I. Ansko, *et al.*, “Laboratory intercomparison of radiometers used for satellite validation in the 400–900 nm range,” *Remote Sens.* **11**(9), 1101 (2019).
18. IOCCG Protocol Series, “Protocols for Satellite Ocean Colour Data Validation: In Situ Optical Radiometry,” G. Zibordi, K. J. Voss, B. C. Johnson, and J. L. Mueller, eds., IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 3.0, (IOCCG, Dartmouth, NS, Canada, 2019).
19. G. Zibordi, D. D’Alimonte, and J.-F. Berthon, “An evaluation of depth resolution requirements for optical profiling in coastal waters,” *J. Atmos. Ocean. Tech.* **21**(7), 1059–1073 (2004).
20. OLCI spectral response functions. *Tech. Guide*. (2016).
21. BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, “Evaluation of measurement data — Guide to the expression of uncertainty in measurement,” (Joint Committee for Guides in Metrology, 2008) p. 100.
22. X. Zhang, S. He, A. Shabani, *et al.*, “Spectral sea surface reflectance of skylight,” *Opt. Express* **25**(4), A1–A13 (2017).
23. P. De Vis and S. Hunt, “CoMet toolkit: a Community Metrology toolkit for handling uncertainty and error-correlation information in Python,” (National Physical Laboratory, 2024).
24. K. G. Ruddick, V. De Cauwer, Y.-J. Park, *et al.*, “Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters,” *Limnol. Oceanogr.* **51**(2), 1167–1179 (2006).
25. M. Hieronymi, “Polarized reflectance and transmittance distribution functions of the ocean surface,” *Opt. Express* **24**(14), A1045–A1068 (2016).
26. G. Zibordi, D. D’Alimonte, and T. Kajiyama, “Automated Quality Control of AERONET-OC L_{WN} Data,” *J. Atmos. Ocean. Tech.* **39**(12), 1961–1972 (2022).
27. H. M. Dierssen, R. A. Vandermeulen, B. B. Barnes, *et al.*, “QWIP: A Quantitative Metric for Quality Control of Aquatic Reflectance Spectral Shape Using the Apparent Visible Wavelength,” *Front. Remote Sens.* **3**, 869611 (2022).
28. R. A. Vandermeulen, A. Mannino, S. E. Craig, *et al.*, “150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean,” *Remote Sens. Environ.* **247**, 111900 (2020).
29. G. Zibordi and C.J. Donlon, “In Situ Measurement Strategies,” In *Experimental Methods in the Physical Sciences*, G. Zibordi, C.J. Donlon, and A.C. Parr, eds., (Academic Press, 2015) Chapter 5, Volume 47, pp. 527–529.
30. H. R. Gordon and D. K. Clark, “Clear water radiances for atmospheric correction of coastal zone color scanner imagery,” *Appl. Opt.* **20**(24), 4175–e4180 (1981).
31. S. B. Hooker, G. Lazin, G. Zibordi, *et al.*, “An Evaluation of Above- and In-Water Methods for Determining Water-Leaving Radiances,” *J. Atmos. Oceanic Technol.* **19**(4), 486–515 (2002).
32. G. Zibordi and B. Bulgarelli, “Effects of cosine error in irradiance measurements from field ocean color radiometers,” *Appl. Opt.* **46**(22), 5529–5538 (2007).
33. J. F. Berthon and G. Zibordi, “Bio-optical relationships for the northern Adriatic Sea,” *Int. J. Remote. Sens.* **25**(7-8), 1527–1532 (2004).
34. P. J. Werdell and S. W. Bailey, “An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation,” *Remote Sens. Environ.* **98**(1), 122–140 (2005).
35. S. W. Brown, S. J. Flora, M. E. Feinholz, *et al.*, “The marine optical buoy (MOBY) radiometric calibration and uncertainty budget for ocean color satellite sensor vicarious calibration,” Proc. SPIE 6744, Sensors, Systems, and Next-Generation Satellites XI (2007).
36. A. Bialek, V. Vellucci, B. Gentil, *et al.*, “Monte Carlo–Based Quantification of Uncertainties in Determining Ocean Remote Sensing Reflectance from Underwater Fixed-Depth Radiometry Measurements,” *J. Atmos. Oceanic Technol.* **37**(2), 177–196 (2020).
37. M. Gergely and G. Zibordi, “Assessment of AERONET-OC L_{WN} uncertainties,” *Metrologia* **51**(1), 40–47 (2014).
38. I. Cazzaniga and G. Zibordi, “AERONET-OC L_{WN} Uncertainties: Revisited,” *J. Atmos. Oceanic Technol.* **40**(4), 411–425 (2023).
39. K. Alikas, V. Vabson, I. Ansko, *et al.*, “Comparison of Above-Water Seabird and TriOS Radiometers Along an Atlantic Meridional Transect,” *Remote Sens.* **12**(10), 1669 (2020).
40. D. Antoine, M. Slivkoff, W. Klonowski, *et al.*, “Uncertainty assessment of unattended above-water radiometric data collection from research vessels with the Dynamic Above-water Radiance (L) and Irradiance (E) Collector (DALEC),” *Opt. Express* **29**(3), 4607–4631 (2021).
41. A. Bialek, S. Douglas, J. Kuusk, *et al.*, “Example of Monte Carlo Method Uncertainty Evaluation for Above-Water Ocean Colour Radiometry,” *Remote Sens.* **12**(5), 780 (2020).

42. V. Vabson, I. Ansko, K. Duong, *et al.*, “Complete characterization of ocean color radiometers,” *Front. Remote Sens.* **5**, 1320454 (2024).
43. G. Zibordi, M. Talone, and L. Jankowski, “Response to Temperature of a Class of In Situ Hyperspectral Radiometers,” *J. Atmos. Ocean. Technol.* **34**(8), 1795–1805 (2017).
44. B. C. Johnson, G. Zibordi, S. W. Brown, *et al.*, “Characterization and absolute calibration of an AERONET-OC radiometer,” *Appl. Opt.* **60**(12), 3380–3392 (2021).
45. J. L. Mueller and R. W. Austin, “Ocean optics protocols for SeaWiFS validation,” TM-1995-104566, vol. 5 of *SeaWiFS Tech. Rep. Ser.* (NASA Goddard Space Flight Center, 1995) p. 46.
46. S. Hooker and A. Morel, “Platform and Environmental Effects on Above-Water Determinations of Water-Leaving Radiances,” *J. Atmos. Oceanic Technol.* **20**(1), 187–205 (2003).
47. G. Zibordi, J. F. Berthon, and D. D’Alimonte, “An Evaluation of Radiometric Products from Fixed-Depth and Continuous In-Water Profile Data from Moderately Complex Waters,” *J. Atmos. Oceanic Technol.* **26**(1), 91–106 (2009).
48. Z. Lee, Y. H. Ahn, C. Mobley, *et al.*, “Removal of surface-reflected light for the measurement of remote-sensing reflectance from an above-surface platform,” *Opt. Express* **18**(25), 26313–26324 (2010).
49. R. W. Gould, R. A. Arnone, and M. Sydor, “Absorption, scattering, and remote sensing reflectance relationships in coastal waters: Testing a new inversion algorithm,” *J. Coastal Res.* **17**, 328–341 (2001).