remote sensing

MDPI

Article

Laboratory Calibration Comparison of Hyperspectral Ocean Color Radiometers in the Frame of the FRM4SOC Phase 2 Project

Viktor Vabson [1,*], Ilmar Ansko [1], Agnieszka Bialek [2], Michael E. Feinholz [3], Joel Kuusk [1], Ryan Lamb [4], Sabine Marty [5], Michael Ondrusek [6], Clemens Rammeloo [2], Eric Rehm [4], Riho Vendt [1], Kenneth J. Voss [7], Juan Ignacio Gossn [8] and Ewa Kwiatkowska [8]

1 Tartu Observatory, University of Tartu, 61602 Tõravere, Estonia; joel.kuusk@ut.ee (J.K.); riho.vendt@ut.ee (R.V.)
2 National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK
3 Moss Landing Marine Laboratories, San Jose State University, San Jose, CA 95039, USA
4 Sea-Bird Scientific, Philomath, OR 97370, USA; rlamb@seabird.com (R.L.)
5 The Norwegian Institute for Water Research, 0579 Oslo, Norway
6 National Oceanic and Atmospheric Administration, College Park, MD 20740, USA
7 Physics, College of A&S, University of Miami, Miami, FL 33124, USA
8 European Organization for the Exploitation of Meteorological Satellites, 64295 Darmstadt, Germany; juanignacio.gossn@eumetsat.int (J.I.G.)
* Correspondence: viktor.vabson@ut.ee

Highlights

What are the main findings?

- During 2022 and 2023, a special laboratory comparison across six laboratories mapped key problems in spectral irradiance and radiance calibrations of hyperspectral ocean color radiometers.
- The comparison results showed agreement between participants within ±3%. This confirmed the validity of the expanded calibration uncertainties around 2% stated by the participants in most cases.

What is the implication of the main finding?

- The metrological compatibility of the absolute calibration coefficients of ocean color radiometers determined in different laboratories can be improved by addressing the identified problems.
- Consistency in laboratory calibrations, along with the use of recently calibrated field instruments, sets the foundation for agreement in field measurement results within the ocean color domain.

Abstract

Variability across different calibration laboratories can impact the consistency of ocean color data; this study addresses that challenge through a coordinated comparison of spectral irradiance and radiance calibrations. As part of the Fiducial Reference Measurements for Satellite Ocean Color (FRM4SOC) Phase 2 project, the metrological consistency across six international laboratories was tested in the years 2022–2023. Each participant determined the responsivity for four transfer radiometers using their own SI-traceable radiometric standards and calibration procedures. This was among the first laboratory comparisons for Ocean Color Radiometry (OCR) using hyperspectral radiometers. The main objective was to verify that the instrument manufacturers and research laboratories can fulfill the updated International Ocean Color Coordination Group (IOCCG) protocols to perform SI traceable calibrations with an uncertainty of 1% (k = 1) for irradiance and slightly more for

check for updates

Academic Editors: Zhenhai Liu and Gennadi Milinevsky

Received: 25 September 2025
Revised: 31 October 2025
Accepted: 5 November 2025
Published: 12 November 2025

Citation: Vabson, V.; Ansko, I.; Bialek, A.; Feinholz, M.E.; Kuusk, J.; Lamb, R.; Marty, S.; Ondrusek, M.; Rammeloo, C.; Rehm, E.; et al. Laboratory Calibration Comparison of Hyperspectral Ocean Color Radiometers in the Frame of the FRM4SOC Phase 2 Project. *Remote Sens.* **2025**, *17*, 3692. https://doi.org/10.3390/rs17223692

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

radiance. The comparison revealed biases among participants and provided an overview of the calibration capabilities of OCRs. The differences between the participants varied from ±1 . . . 2% up to ±5%. Biases due to different measurement conditions were corrected by the Pilot. Furthermore, biases due to traceability and different conditions revealed several data handling errors. However, after uniform data processing, the metrological compatibility between the participants was reached within ±3%.

Keywords: ocean color; hyperspectral radiometer; calibration; uncertainty; comparison measurement; bias; metrological compatibility

1. Introduction

Fiducial reference measurements (FRM) are reliable ground measurements with uncertainty estimates, used for validation of the satellite remote sensing data products [1]. The Committee on Earth Observation Satellites (CEOS) [2] recommends in the FRM principles that in situ ocean-color radiometers (OCRs) have a documented history of SI-traceable calibrations, including uncertainty budgets [3]. Metrological traceability to the International System of Units [4] is the concept that links all metrological measurements to the SI through a series of calibrations or comparisons. Absolute radiometric calibration of OCRs is needed to convert the raw/field results obtained in arbitrary units to the values of measured quantities in the specific SI units, and with an estimate of measurement uncertainty [5] making results measured with different OCRs comparable. Each step in this traceability chain must have a rigorous documented uncertainty analysis. In metrology, the term "comparison" [6] refers to the process of validating an uncertainty analysis of measurements obtained from reference standards (artifacts) operated by different laboratories. A number of round-robin experiments arranged during the last decades [7–21] for validation of performance, calibration, and characterization of OCR instruments have demonstrated the importance of SI-traceability and metrological compatibility of results—a required condition for FRM. Laboratory comparison measurements are focused on the OCRs' calibration and characterization activities, which form the basis for a successful FRM. For absolute radiometric calibration, the SeaWiFS Intercalibration Round-Robin Experiment (SIRREX) and the Second Intercomparison and Merger for Interdisciplinary Ocean Studies (SIMBIOS) Radiometric Intercomparison (SIMRIC-2) exercises showed that the level of measurement uncertainty can be maintained at the level of 2–3% if the protocols are followed [22]. Recently updated IOCCG protocols [23] specified a need for radiometric intercomparisons and set a more challenging uncertainty target:

> "Instrument manufacturers and a few research laboratories are equipped and staffed to perform these calibrations for the ocean color research community. These facilities should perform frequent intercomparisons [24] to ensure the maintenance of the radiometric traceability to NMI standards. An ambitious goal is to perform calibrations from 350 nm to 900 nm with 1% target uncertainty for irradiance and slightly higher for radiance (the coverage factor k = 1)."

The most recent results from the FRM4SOC irradiance comparison [18] showed an agreement between all working standard lamps included in the comparison within ±1.5%. A comparison of radiance sources (calibrated lamp—reflectance plaque combinations) was performed via the transfer of multispectral radiometers calibrated in each participant's laboratory. Discrepancies between the results were at the level of ±4%—significantly higher than expected. Additional investigation revealed potential sources of disagreement, and when these were corrected, an agreement within 2.5% could be reached.

In this study, the methodology and results of international laboratory comparison measurements of the spectral irradiance and radiance in the framework of the Copernicus FRM4SOC Phase 2 project conducted in 2022–2023 are described. Six participants determined the responsivity calibration coefficients in the wavelength range from 350 nm to 900 nm for four transfer radiometers using their own SI-traceable spectral irradiance and radiance standards, and their usual calibration procedures. This is among the first OCR laboratory comparison exercises where the transfer radiometers are hyperspectral; in the past, the transfer radiometers were multispectral instruments. However, currently, the majority of in situ radiometry is acquired using hyperspectral instruments, and some of the multispectral precursors are not currently in production. Therefore, it was decided to run this comparison with the transfer instruments most often used in situ. The purpose of the comparison was to compare the measurement and data handling process as performed by the participants calibrating the same set of transfer standards, not to require them to adopt the same procedure for performing the tasks. These comparison results closely mirror the agreement between the instruments used for the field measurements within the ocean color community when the instruments are calibrated by the comparison participants.

The transfer instruments were the two most common OCR models used by the ocean color community: two TriOS RAMSES (radiance and irradiance) and, similarly, two Sea-Bird HyperOCR hyperspectral radiometers. The comparison was arranged in two loops. At first, the Pilot laboratory measured the transfer standards, then in sequence they were sent to each participant, and then again to the Pilot for the final measurement. The quantitative results of the comparison are presented in terms of differences between each participant's measurement and the comparison consensus value (CCV). Particular attention in this comparison was given to hyperspectral instrument characteristics that can affect deviations of the calibration coefficients in individual laboratories, including non-linearity, ambient temperature, stray light, and temporal drift. The results are presented as uncorrected and corrected for these specific instrument-related effects.

In the Section 2, the materials and methodology are described in detail. This is followed in Section 3 by the results of the comparison without any additional correction, as well as the instrument-related corrections and their effects on the comparison results. Section 4 describes the uncertainty evaluation, and is followed by a discussion and concluding remarks.

2. Materials and Methods

2.1. Participants, Transfer Standards, Time Schedule

Six laboratories, including the Pilot laboratory, participated in this comparison. An overview of participating laboratories is given in Table 1.

Table 1. List of participants.

Institute/Laboratory	Acronym	Contact Person
University of Tartu, Tartu Observatory (Pilot)	TO	Organization: Viktor Vabson Measurements: Ilmar Ansko
Norsk Institutt for Vannforskning	NIVA	Sabine Marty
Sea-Bird Scientific	SB	Cristina Orrico, Eric Rehm
Moss Landing Marine Laboratories, San Jose State University	MLML	Michael E. Feinholz, Kenneth J. Voss
National Oceanic and Atmospheric Administration	NOAA	Michael Ondrusek
National Physical Laboratory	NPL	Agnieszka Bialek, Clemens Rammeloo

Four hyperspectral radiometers were used as comparison transfer instruments: TriOS GmbH (Germany) RAMSES radiance and irradiance sensors, and Sea-Bird Scientific (USA) HyperOCR radiance and irradiance sensors. The radiometers are listed in Table 2. The serial numbers listed in Table 2 are used throughout the remainder of this text to refer to the comparison transfer radiometers. The detailed specifications of the radiometers used for the calibration comparison can be found in [5,25] which are based on documents supplied by the manufacturers of the OCRs and each of the embedded components/modules, and on the results of the instrument's on-site tests. In summary, these radiometers measure radiation from the atmosphere or water in the spectral range of 350 nm to 1000 nm with a spectral resolution of about 10 nm, a spectral sampling interval of approximately 3.3 nm, and a wavelength accuracy of 0.3 nm. The radiometers contain the Carl Zeiss Monolithic Miniature Spectrometer (MMS 1 module), incorporating a 256-channel silicon photodiode array, proprietary front-end electronics, and optical input elements in a watertight housing. Integration time can be set or adjusted automatically from 4 ms to 8 s, depending on light intensity. The housings are cylindrically symmetric, with an optical input and electrical connector at opposite ends of the cylinder. The wavelength scale is defined in the calibration files provided by the manufacturer.

Table 2. List of comparisons transfer radiometers.

No.	Serial Number	Manufacture Date	Function	Manufacturer	OCRs Family
1	SAM_81B0	2006	Radiance (L)	TriOS GmbH	RAMSES
2	SAM_8598	2018	Irradiance (E)	TriOS GmbH	RAMSES
3	SAT2073	2021	Radiance (L)	Sea-Bird Scientific	HyperOCR
4	SAT2072	2021	Irradiance (E)	Sea-Bird Scientific	HyperOCR

The HyperOCR radiometers are equipped with an internal mechanical shutter. The shutter frequency can be selected in the controlling software. The RAMSES radiometers do not have a shutter but contain black-painted pixels—usually the last 16 pixels of the silicon photodiode array. The average signal of the black-painted pixels represents the temperature-dependent dark signal and is always subtracted. Despite the availability of such black-painted pixels, measuring the real background using an external shutter with the same integration time is preferable. However, subtraction of the average signal from the black-painted pixels is still applied as a tool to reduce temporal variation in the instrument's dark signal. Both the dark signal and the radiometric responsivity have significant dependence on temperature, especially above 20 °C, and on integration time, especially if it is long (>1 s). The HyperOCR radiometers have built-in temperature sensors, with readings of the radiometer's internal temperature recorded by the manufacturer's software. The RAMSES radiometers used during this comparison exercise were modified to have temperature sensors as well. The temperature of the RAMSES radiometers was recorded using the software provided by the Pilot.

The circulation schedule of the transfer instruments and a list of presented results for each OCR is given in Table 3.

The comparison consensus value (CCV) is the arithmetic average of the calibration results of all participants. Specific relative differences between values obtained by each participant and the CCV are reported and analyzed in this study.

Table 3. Circulation schedule and a table of presented results.

Lab	Date	Results Reported by Participants			
		RAMSES		HyperOCR	
		Irradiance E1 (SAM_8598)	Radiance L1 (SAM_81B0)	Irradiance E2 (SAT2072)	Radiance L2 (SAT2073)
TO	January 2022	Y	Y	Y	Y
NIVA	February–March 2022	Y	Y	Y	Y
TO	April 2022	Y	Y	Y	Y
Sea-Bird Sci.	June 2022	N/A	Y	Y	Y
NOAA	August 2022	Y	Y	Y	Y
MLML/MOBY	October 2022	Y	Y	Y	Y
NPL	July 2023	Y	Y	Y	N/A
TO	August 2023	Y	Y	Y	Y

The comparison results are anonymized, with each participating laboratory having its own code: P1 ... P6. Following [26] this code will be confidential. Table 4 lists the measurement standards used by the participants, the measurement distances, and the number of different integration times used during calibration. In this comparison, 1000 W quartz tungsten halogen (QTH) lamps, often called FEL lamps (not an acronym) according to ANSI (American National Standard Institute) designation, were used as irradiance sources. The lamps were used at the standard calibration distance of 500 mm measured from their reference plane. For radiance scale realization, four participants used a FEL lamp and reflectance panel combination, while two participants used integrating sphere sources.

Table 4. Measurement standards, distances, and the number of different integration times.

Participant	Standards		Measurement Distance, mm			Number of Used Int. Times
	Irradiance	Radiance	Irradiance FEL–OCR	Radiance Integrating Sphere—OCR		
P1	FEL	Sphere	500	152.4		1
P2	Two FELs	Sphere	500	140		3–4
				FEL—Panel	Panel—OCR	
P3	Two FELs	FEL + Panel	500	500	N/A	3
P4	Two FELs	FEL + Panel	500	1300	N/A	3–4
P5	FEL	FEL + Panel	500	1000	N/A	1
P6	Two FELs	FEL + Panel	500	500	200; 250	3

2.2. Calibration Procedure

The spectral responsivity of a radiometer is calibrated by measuring a known radiation source aligned at a specified distance from the device under test (DUT). Radiometric calibration procedures are well established and validated [11,27–31], but instrument data processing and the format of the derived calibration coefficients may depend on the specific model of DUT. In the case of irradiance sensors, a reference lamp with known spectral irradiance $E_r(\lambda)$ is used for the calibration of the DUT, while for the case of radiance sensors a lamp-panel setup or an integrating sphere with known spectral radiance $L(\lambda)$ is used. The irradiance/radiance values from the calibration certificate(s) are interpolated to designated wavelengths λ_i of the DUT wavelength scale. The calibration coefficients derived will be stored in a model-specific data set and applied to the DUT measurements carried out later.

The distance between the reference lamp and the measuring head of the DUT radiometer needs to be set to match the reference distance specified in the lamp's calibration certificate. A reference plane for the distance measurement is described in the calibration certificate of the lamp, usually defined by a certain plane of the lamp holder or an alignment

jig placed in the lamp holder. A second reference plane for the distance measurement is usually the front surface of the input optics of the DUT radiometer. However, a radiometer's effective optical reference plane does not necessarily coincide with the front surface of its input optics [32–34], and needs to be determined experimentally. If the realized distance d is different from the standard or reference distance, d_{ref}, of the lamp calibration (usually d_{ref} = 500 mm), the spectral irradiance at a distance d can be determined using the formula

$$E(\lambda,\ d) = E_{\mathrm{r}}(\lambda,\ d_{\mathrm{ref}})\frac{(d_{\mathrm{ref}} + \Delta d)^2}{(d + \Delta d)^2}, \tag{1}$$

where Δd is the combined offset of the lamp and radiometer effective working planes from the specified reference planes.

To reduce the random noise, the number of measurements per radiometer was at least 30 for each integration time used, followed by the same number with the light path between the lamp and the radiometer blocked. The radiometer signal must be corrected for the detector's dark signal, and ambient stray light from the surroundings should also be evaluated and subtracted. Both the dark and external stray light measurements should be carried out with the same integration time used for the reference radiation measurement. For the temperature correction, the internal temperature during the calibration measurements shall be recorded for each DUT radiometer.

For the comparison measurements, the participants used their standard calibration procedures. The comparison protocol requested all raw spectra to be submitted to the Pilot in digital counts together with the used integration times, the timestamps, and the derived calibration coefficients with their respective uncertainties. For each series with the reference spectra, a series with dark and/or ambient stray light spectra was requested. The comparison protocol suggested using at least three different integration times when measuring the same radiation source at the same distance between the source and the radiometer. These data were intended to evaluate and eliminate the non-linear effects of the transfer radiometers. The Pilot analyzed the inter-comparison results, repeated the derivation of the calibration coefficients performed by the participants, applied uniform corrections for different measurement conditions in different laboratories (non-linearity, ambient temperature, stray light, temporal drift), and evaluated the consistency between the participants' results.

2.3. Calibration Coefficients

The measurement task for all participants of the comparison was the determination of responsivity calibration coefficients of the transfer radiometers. Although the raw spectra in digital numbers obtained in calibration measurements were similar for all four transfer radiometers (Figures 1a and 2a), the responsivity coefficients as available in model-specific data sets provided by the manufacturers, together with the radiometer's software, have different definitions (Figures 1b and 2b). Differences in hardware and software make the calibration and characterization procedures dependent on the specific radiometer model, which impedes the harmonization of guidelines and the uniform presentation of the results.

Following [23], the calibration coefficients of the sensor irradiance responsivity $F_E(\lambda)$ are calculated for each wavelength band λ by applying:

$$F_E(\lambda) = \frac{E_{\mathrm{r}}(\lambda)}{DN_{\mathrm{r}}(\lambda) - DN_{\mathrm{amb}}(\lambda)}, \tag{2}$$

where $E_{\mathrm{r}}(\lambda)$ is the certified spectral irradiance of the lamp at the reference distance of d_{ref} = 500 mm, $DN_{\mathrm{r}}(\lambda)$ is the sensor response to the lamp radiation, and $DN_{\mathrm{amb}}(\lambda)$ is the response to ambient light when an occulting device is obstructing the direct path

between the lamp and the sensor. If the ambient light signal is perceptibly larger than the dark signal measured with a completely covered collector, then baffling needs to be improved. The irradiance calibration coefficients $F_E(\lambda)$ defined via (2) can be presented in units of [μW cm^{-2}nm^{-1}].

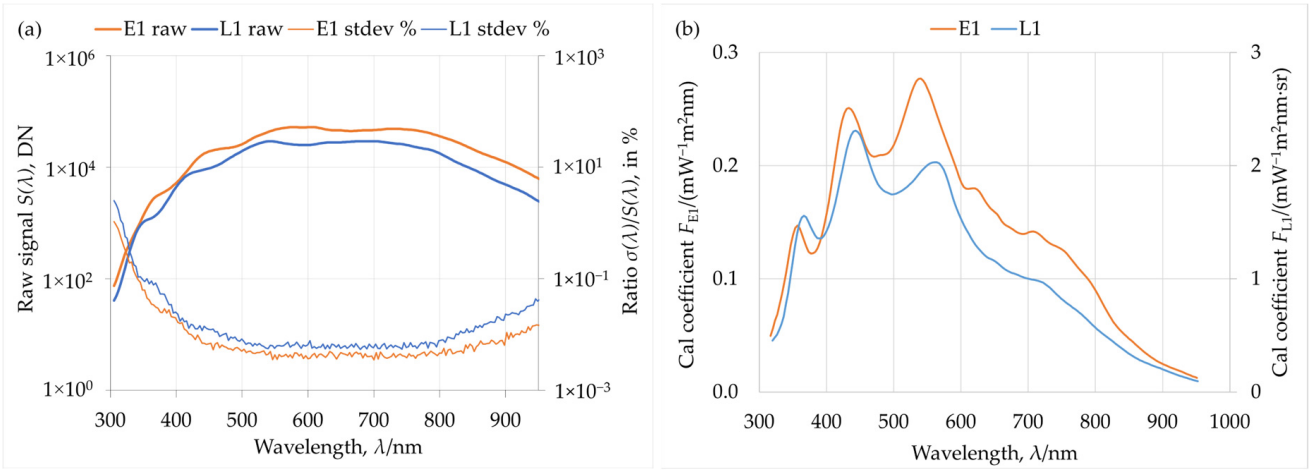


Figure 1. Calibration measurements for the RAMSES irradiance E1 and radiance L1 sensors. (**a**) raw signals in digital numbers and standard deviations; (**b**) responsivity coefficients.

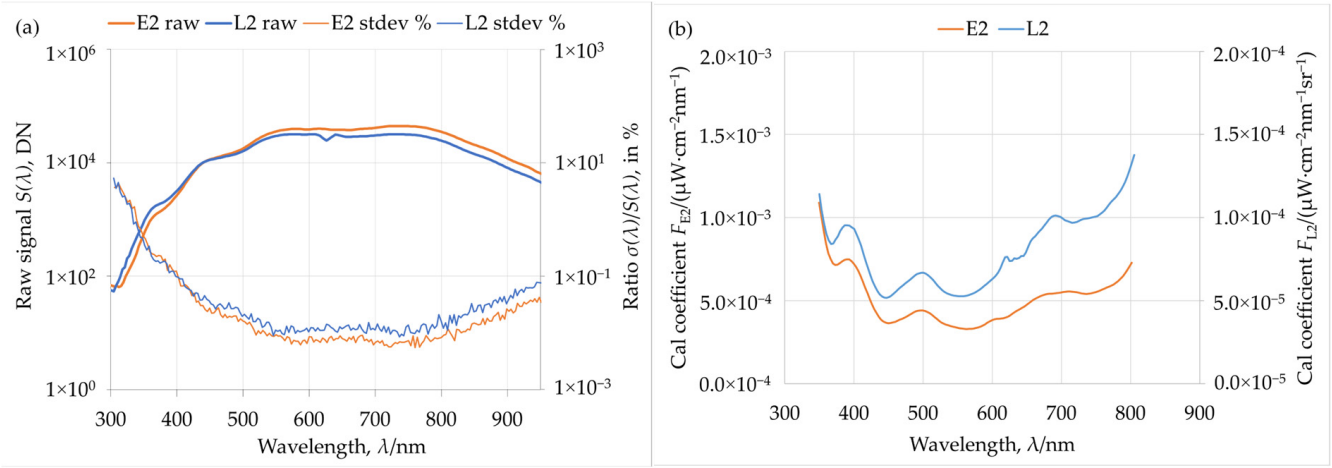


Figure 2. Calibration measurements for the HyperOCR irradiance E2 and radiance L2 sensors. (**a**) raw signals in digital numbers and standard deviations; (**b**) responsivity coefficients.

The irradiance $E(\lambda)$ determined from the measurements $DN(\lambda)$ by using the calibrated DUT sensor can be computed as [23]:

$$E(\lambda) = F_E(\lambda)[DN(\lambda) - DN_{\text{dark}}(\lambda)] \tag{3}$$

Equation (3) is based on $F_E(\lambda)$ obtained from (2) and is valid only if the gain and integration time of the radiometer are the same for both measurements.

2.3.1. Manufacturer's Definition of the Coefficients: TriOS RAMSES

Calibration coefficients of RAMSES irradiance sensors F_{E1} based on the calibration files provided by the manufacturer, the values are calculated as follows:

$$F_{E1}(\lambda) = \frac{S(\lambda)\left(C_{stray}C_{lin}C_T\right)}{E_r(\lambda)(C_1C_2 \ldots C_i \ldots)} \frac{t_{max}}{t_{used}} [\text{mW}^{-1}\text{m}^2\text{nm}] \tag{4}$$

Here $S(\lambda)$ is the raw signal after dark correction and normalization to the range [0, 1]; $E_r(\lambda)$ is the lamp irradiance at the reference distance; $t_{\max}$ = 8192 ms denotes the largest possible integration time; t_{used} is the integration time used for the measurements. The correction factors C_{stray}, C_{lin}, and C_T account for stray light, non-linearity, and temperature effects in the raw signal, while the correction factors C_i account for different effects related to the use of the FEL lamps. These coefficients $F_{E1}(\lambda)$ are inversely proportional to the coefficients calculated from Formula (2). The suggested units are also different: ten times smaller and in inverse proportion if compared to the units suggested in Formula (2).

Calibration coefficients of RAMSES radiance sensors F_{L1} based on the calibration files provided by the manufacturer are calculated as follows:

$$F_{L1}(\lambda) = \frac{S(\lambda)\left(C_{stray}C_{lin}C_T\right)}{L(\lambda)}\frac{t_{max}}{t_{used}}[\text{mW}^{-1}\text{m}^2\text{nm sr}] \quad (5)$$

$L(\lambda)$ is the target radiance taken in the case of the integrating sphere from the calibration certificate, or for the lamp panel setup, calculated as:

$$L(\lambda) = E_r(\lambda)(C_1C_2\ldots C_i\ldots)\frac{R(0^\circ, 45^\circ, \lambda)}{\pi}\frac{(d_{\text{ref}} + \Delta d)^2}{(d + \Delta d)^2} \quad (6)$$

Here $R(0^\circ, 45^\circ, \lambda)$ is the bidirectional reflectance factor of the diffuser plate taken from the calibration certificate (0° illumination, 45° observation) and interpolated to the designated wavelength λ. Calibration measurements for the RAMSES irradiance E1 and radiance L1 sensors are shown in Figure 1. Dark corrected raw signals in digital numbers as an average of 30 repeated measurements and respective relative standard deviations are given in Figure 1a. Calibration coefficients F_{E1} and F_{L1} of the RAMSES radiometers in the manufacturer file format are given in Figure 1b.

2.3.2. Manufacturer's Definition of the Coefficients: Sea-Bird HyperOCR

Calibration coefficients of HyperOCR irradiance sensors $F_{E2}(\lambda)$ based on the calibration files provided by the manufacturer, conform to Equation (2) and are calculated as follows:

$$F_{E2}(\lambda) = \frac{E_{\text{r}}(\lambda)(C_1C_2\ldots C_i\ldots)}{S(\lambda)\left(C_{stray}C_{lin}C_T\right)}[\mu\text{Wcm}^{-2}\text{nm}^{-1}] \quad (7)$$

Here $S(\lambda)$ is a dark-corrected raw signal in digital counts, $E_{\text{r}}(\lambda)$ is the lamp irradiance at the reference distance. Integration time t_{cal} shall always be present in the calibration data set and will be applied by using the manufacturer's software. However, calibration coefficients as defined in (7) may be different among the laboratories, as they will depend on the integration time t_{cal} used for calibration. The calibration coefficients must be adjusted if they are to be used for measurements at different integration times or when comparing between laboratories using different integration times. This can be performed using the ratio of the integration times, t_{meas} divided by t_{cal}.

Calibration coefficients of HyperOCR radiance sensors $F_{L2}(\lambda)$ based on the calibration files provided by the manufacturer are calculated as follows:

$$F_{L2}(\lambda) = \frac{L(\lambda)}{S(\lambda)\left(C_{stray}C_{lin}C_T\right)}[\mu\text{W cm}^{-2}\text{nm}^{-1}\text{sr}^{-1}] \quad (8)$$

Here $S(\lambda)$ is a dark-corrected raw signal in digital counts without any modification; $L(\lambda)$ is the target radiance, either taken from the calibration certificate when an integrating sphere is used, or for the lamp-panel setup, calculated according to Equation (6).

Calibration measurements for the HyperOCR irradiance E2 and radiance L2 sensors are shown in Figure 2. The dark corrected raw signals in digital numbers and their respective relative standard deviations are given in Figure 2a. Calibration coefficients F_{E2} and F_{L2} of the HyperOCR radiometers in the manufacturer's file format are given in Figure 2b.

Integration time, t_{cal}, shall always be present in the calibration data set, and again, calibration coefficients determined by different laboratories using (8) may be different, depending on t_{cal}. Therefore, in the comparison data analysis, the integration times have been accounted for according to (9) and (10).

$$F_{E2}(\lambda) \times \left(\frac{t_{meas}}{t_{cal}}\right) = \frac{E_r(\lambda)}{S(\lambda)} \frac{t_{meas}}{t_{cal}} [\mu W\ m^{-2} nm^{-1}] \quad (9)$$

$$F_{L2}(\lambda) \times \left(\frac{t_{meas}}{t_{cal}}\right) = \frac{L(\lambda)}{S(\lambda)} \frac{t_{meas}}{t_{cal}} [\mu W\ m^{-2} nm^{-1} sr^{-1}] \quad (10)$$

3. Results

3.1. Relative Difference from CCV of Reported Results

In the preliminary analysis, the data reported by the participants are compared, avoiding corrections as much as possible. However, the results are converted to the same units and, in the case of SAT2072 and SAT2073, normalized to a common integration time (Equations (9) and (10)) as was used during initial calibration by the manufacturer. The data submitted by the participant contained, in some cases, data handling errors, which were detected by the Pilot and were corrected after the conversation with the individual participant. The matter will be addressed in the subsequent sections in more detail. The quantitative results of the comparison are presented in terms of differences between each participant's measurement and the comparison consensus value (CCV). The CCV was calculated as the average of all participants. As a consequence, each correction to the participant's measurement will result in a respective change in the relative differences to the CCV of all participants' results. The relative difference from the CCV of reported responsivity coefficients is given in Figure 3. These coefficients are not corrected for drift, non-linearity, and temperature differences, as information about the variability of measurement conditions needed for correcting was not available to participants during reporting. Differences for RAMSES radiance sensor SAM_81B0 are in Figure 3a and for RAMSES irradiance sensor SAM_8598 in Figure 3b. The relative difference from the consensus value of reported responsivity coefficients for the HyperOCR radiance sensor SAT2073 is given in Figure 3c, and for the HyperOCR irradiance sensor SAT2072 in Figure 3d.

Looking at deviations from the CCV, we should always remember that the calibration coefficients of RAMSES (defined in Equations (4) and (5)) are in inverse proportion to the coefficients of HyperOCR ((7) and (8)). Almost all differences between participants stay within $\pm$ 4%. Larger discrepancies are evident in the UV range, while in the visible and IR ranges, only a few results deviate from the group by more than 3%.

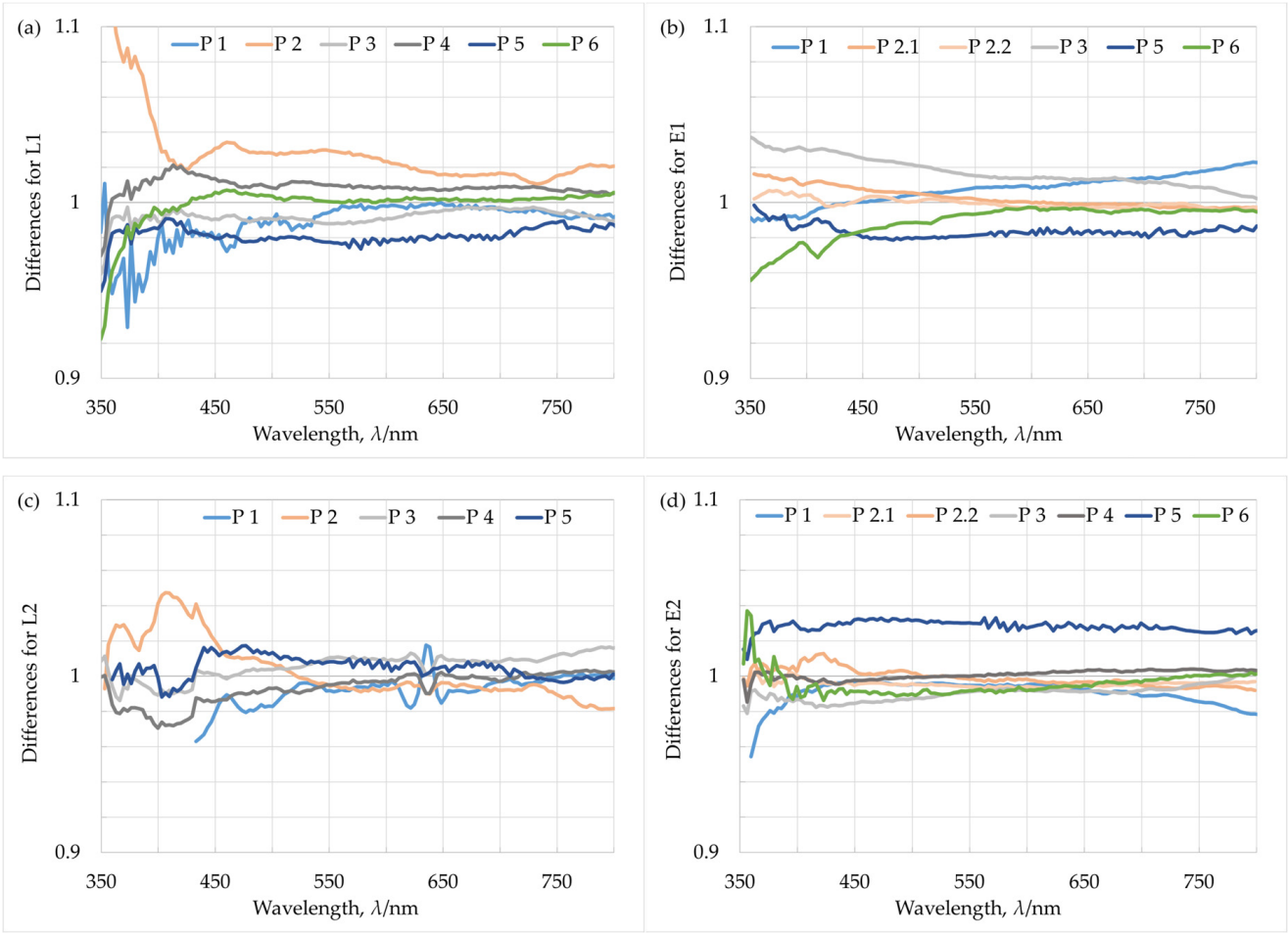


Figure 3. The relative difference from CCV of the uncorrected responsivity coefficients. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for HyperOCR irradiance sensor SAT2072.

3.2. Corrections

3.2.1. Corrections Applied by the Pilot to the Results of All Participants

To demonstrate the consistency of the radiometric scales between the participants, corrections were made for the known systematic effects due to different measurement conditions between participants. For that, the Pilot used the same algorithm to evaluate corrections for the data reported by all participants. The corrections for temporal drift, non-linearity, and temperature differences were evaluated, based on the respective characterization results for the transfer radiometers [25,35] and data about the measurement conditions reported by the participants. Combined corrections were applied to the responsivity coefficients corrected/recalculated by the Pilot. Each correction step altered the responsivity, and hence the consensus value (CCV) and the related combined measurement uncertainty. The stray light correction was applied only to radiance sensors, as differences between the corrections for irradiance sensors between participants were insignificant. Straylight correction factors for radiance calibrations using FEL + panel and/or integrating sphere methods resulted in significantly different straylight correction factors in the UV range. Significant differences in measurement conditions during the comparison are as follows:

1. Non-linearity effects due to differences between the calibration sources used
2. Thermal effects due to differences in ambient temperatures and the radiometer's internal temperatures

3. Drift effects due to different measurement times

Each correction of these effects that is applied by the Pilot involves two factors, as detailed below:

- Non-linearity: The correction is proportional to the participant's raw spectrum in digital numbers $S_{DN}(\lambda)$ corrected for dark and measured with the largest integration time used for the comparison calibration. The proportionality coefficient is the non-linearity factor $\alpha_{\mathrm{nl}}(\lambda)[1/DN]$. For RAMSES, the correction is subtractive:

$$C_{\mathrm{lin}}(\lambda) = 1 - \alpha_{\mathrm{nl}}(\lambda) \cdot S_{DN}(\lambda) \tag{11}$$

- Thermal sensitivity: proportional to the difference between the radiometer's internal temperature T_{lab} measured during calibration in the laboratory and the common reference temperature selected by the Pilot (T_{ref} = 23 °C). The proportionality coefficient is the thermal sensitivity $\alpha_T(\lambda)[1/°\mathrm{C}]$, for RAMSES:

$$C_T(\lambda) = 1 - \alpha_T(\lambda) \cdot (T_{\mathrm{lab}} - T_{\mathrm{ref}}) \tag{12}$$

- Temporal drift: proportional to the difference between the measurement time t_{lab} of a laboratory and a common reference time t_{ref} selected by the Pilot. The proportionality coefficient is the temporal drift $\alpha_t(\lambda)[1/\mathrm{day}]$, for RAMSES:

$$C_t(\lambda) = 1 - \alpha_t(\lambda) \cdot (t_{\mathrm{lab}} - t_{\mathrm{ref}}) \tag{13}$$

Characterization procedures for the correction coefficients $\alpha_{\mathrm{nl}}(\lambda)[1/DN]$, $\alpha_T(\lambda)[1/°\mathrm{C}]$ and $\alpha_t(\lambda)[1/\mathrm{day}]$ present in Equations (11) to (13), as well as their associated uncertainties, are described in [25,35].

The corrections for the HyperOCR sensors are additive (instead of subtractive for RAMSES) because of the different definition of the calibration coefficients:

$$C_{\mathrm{lin}}(\lambda) = 1 + \alpha_{\mathrm{nl}}(\lambda) \cdot S_{DN}(\lambda) \tag{14}$$

$$C_T(\lambda) = 1 + \alpha_T(\lambda) \cdot (T_{\mathrm{lab}} - T_{\mathrm{ref}}) \tag{15}$$

$$C_t(\lambda) = 1 + \alpha_t(\lambda) \cdot (t_{\mathrm{lab}} - t_{\mathrm{ref}}) \tag{16}$$

The combined correction coefficients for both cases were calculated as:

$$C_{\mathrm{comb}}(\lambda) = C_{\mathrm{nl}}(\lambda) \cdot C_T(\lambda) \cdot C_t.(\lambda) \tag{17}$$

To achieve the corrected results for each sensor, the non-corrected calibration coefficients were multiplied by the respective combined correction coefficients.

3.2.2. Temporal Changes in Comparison Transfer Radiometers

The full duration of the laboratory comparison measurements was 564 days. The stability of radiometers was evaluated based on seven measurements at the Pilot laboratory. According to the timetable of the comparison, the Pilot made the measurements in January 2022 before the first loop, three times in April 2022 after receiving the comparison radiometers back, and before sending them to the second loop, and three times in August 2023 after receiving the radiometers back from the last participant. The relative yearly drift of the comparison radiometers normalized to responsivity at the date of P2 measurements is shown in the Figure 4. Already during the three months of the first loop, in the range of 400 nm to 800 nm, a responsivity drift of transfer standards up to 1% was visible for all

radiometers except for SAT2072. Therefore, compensation for temporal drift was needed for the comparison analysis.

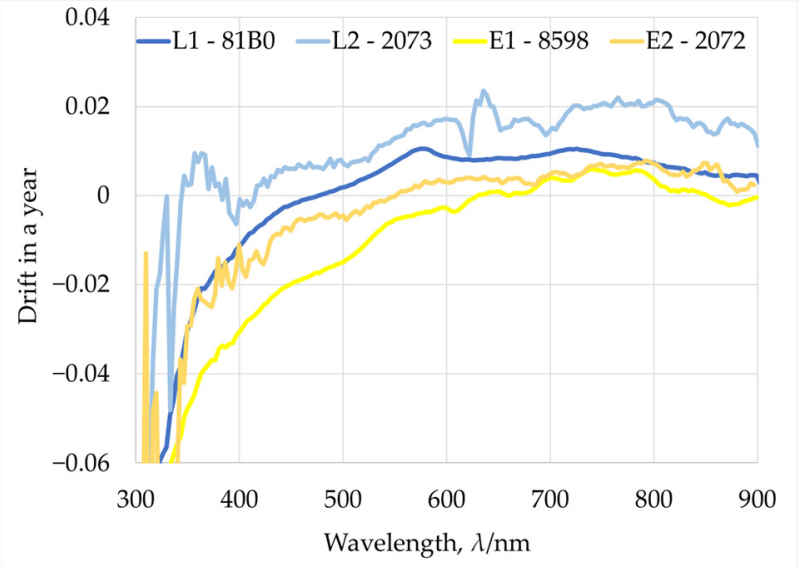


Figure 4. Relative drift of the comparison transfer radiometers during a year as a function of wavelength; irradiance sensors plotted in yellow and radiance sensors in blue.

For the drift compensation in the comparison analysis, the measurement date of P2 was taken as the central date, and all temporal corrections were referenced to that same date using a correction proportional to the time difference with the P2 measurements: P5—130 days, P1—53 days, P4—48 days, and P6—379 days. The results of the Pilot were determined by the least squares method from the seven spectra, where the solution contains a calibration spectrum and a drift rate at a particular time. The drift coefficient, together with the Pilot's spectrum used for the comparison, was referenced for that same date as the P2 measurements time.

After the instruments were returned to the Pilot laboratory from the last participant, the radiometers were calibrated before and after cleaning the input optics. No significant cleaning effect was detected. During the second loop of the comparisons, the input optics of the RAMSES sensor SAM_8598 were damaged: a thin fragment of the cosine collector was broken off. The instrument was recalibrated; however, the effects of the damage cannot be clearly separated from the temporal drift effect.

The correction factors of different participants for the drift are shown in Figure 5. For RAMSES sensors, the correction factors are calculated using Formula (13), and for HyperOCR sensors, they are calculated using Formula (16). As all temporal corrections were calculated to the same date with the measurement time of P2, the drift correction factor of the P2 results is always 1.

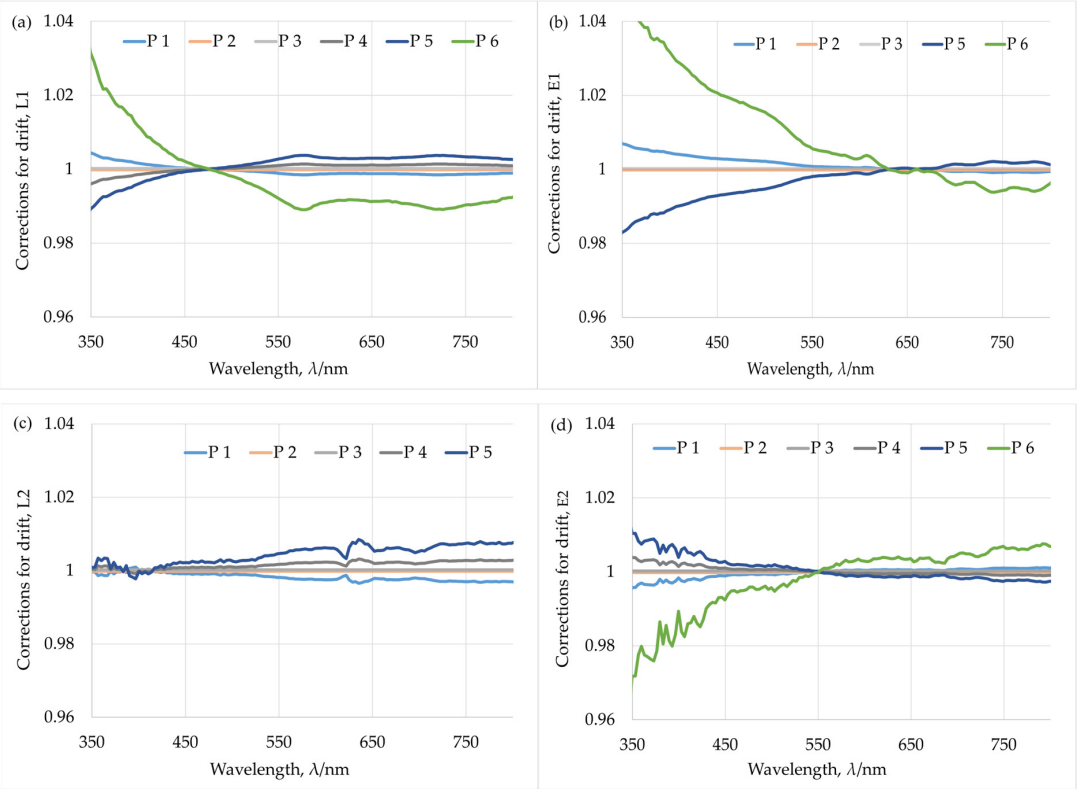


Figure 5. Corrections for temporal drift. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for the HyperOCR irradiance sensor SAT2072.

3.2.3. Non-Linearity

The longest integration times of transfer standards reported during comparison calibrations are given in Table 5. The various participant laboratories set different integration times (up to a factor of 32) for the calibrations of the radiance sensors (SAM_81B0 and SAT2073). This large variability implies that the choice, range, and setting accuracy of integration times need more attention. For the irradiance sensors, the same integration time was used by all labs, except for one instance where it was half of that used elsewhere.

Table 5. The longest integration times which were used at the participant's laboratories during comparison calibrations for different transfer standards.

	Integration Time, ms			
Participant	**L1 (SAM_81B0)**	**E1 (SAM_8598)**	**E2 (SAT2072)**	**L2 (SAT2073)**
P1	1024	256	512	8192
P2	32	256	256	256
P3	128	256	512	1024
P4	512	N/A	512	8192
P5	512	256	512	512
P6	128	256	512	N/A

The non-linearity correction factors of different participants are given in Figure 6, and applied to the individual raw spectra following [36–38]. The correction factors for RAMSES sensors are calculated using Formula (11), and for HyperOCR sensors using Formula (14).

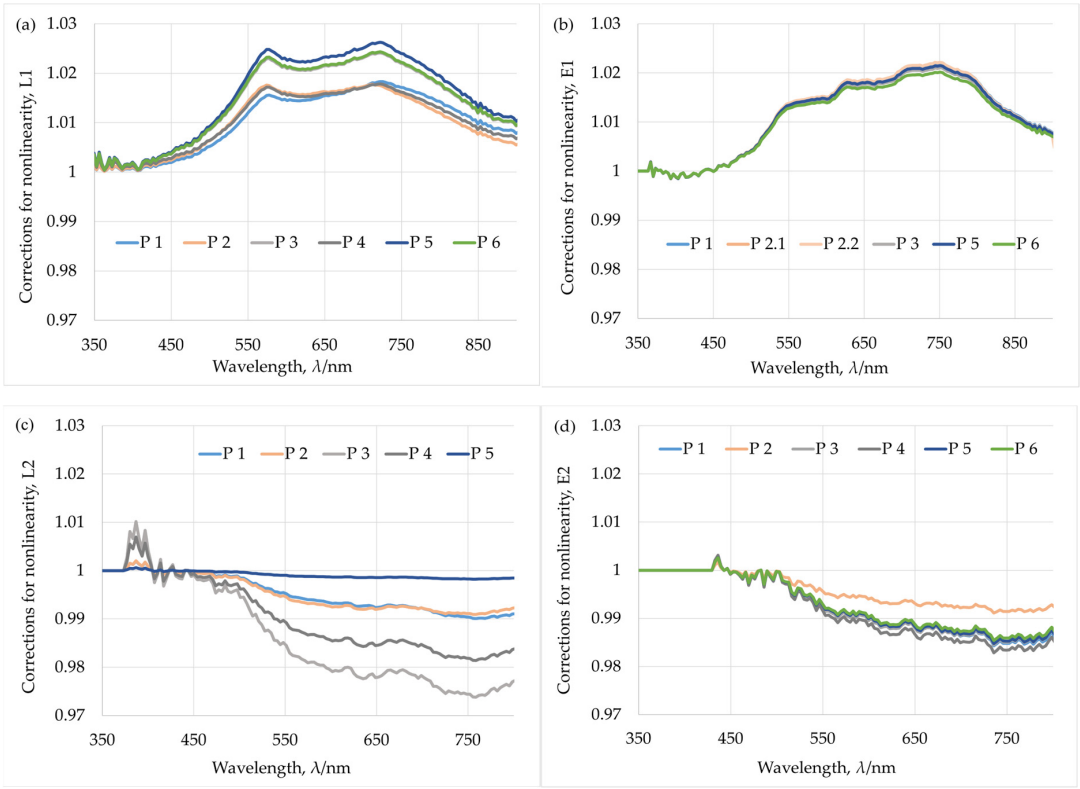


Figure 6. Corrections for non-linearity. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for HyperOCR radiance sensor SAT2073. (**d**) results for HyperOCR irradiance sensor SAT2072.

3.2.4. Stray Light

The stray light correction was applied by using the individual stray light matrices and the iterative method described in [23,39]. Stray light correction factors for radiance calibrations using FEL + panel and/or integrating sphere methods showed notable differences in the UV range, see Figure 7. Stray light corrections are larger for spectra measured using the integrating sphere compared to the FEL + panel method. For both methods, the reproducibility of stray light corrections, as replicated for different participants, was good. All stray light corrections for irradiance calibrations of different participants were close: in the spectral range 350 nm to 400 nm, the spread in correction factors between participants was within ±0.3% and from 400 nm to 800 nm, within ±0.1%, and these corrections were not applied.

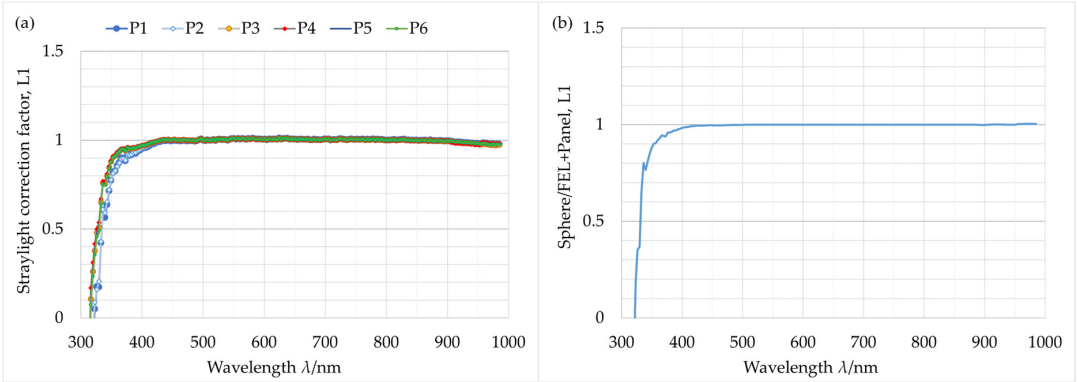


Figure 7. (**a**) Stray light correction factor for radiance calibrations (SAM_81B0). (**b**) Ratio of correction factors for the integrating sphere and FEL + panel methods.

3.2.5. Temperature Correction

Temperatures reported at the participant's laboratories during calibrations of transfer standards are given in Table 6. During calibration, temperature differences up to 5 °C were evident. These temperature differences can cause biases of over 2% in the NIR range around 800 nm if not corrected.

Table 6. Temperatures at the participant's laboratories during calibrations.

Participant	Measurement Location	Temperature, °C			
		L1 (SAM_81B0)	E1 (SAM_8598)	E2 (SAT2072)	L2 (SAT2073)
P1	Ambient	25.7	26.5	26.2	26.0
	Device internal	26.8	27.7	30.3	30.1
P2	Ambient	N/A	23.8–25.1	N/A	N/A
	Device internal	23.3–25	24–24.3	26.3–30.0	26.2–26.8
P3	Ambient	21.5	21.5	21.5	21.5
	Device internal	23.5	22.8	24.3–24.5	23.5–24.5
P4	Ambient	21.9–22.4	N/A	22.6–23.1	22.9–23.1
	Device internal	N/A	N/A	26.3–27.2	26.8–27.3
P5	Ambient	26	26	26	26
	Device internal	N/A	N/A	N/A	N/A
P6	Ambient	21.5–21.8	21.3–22.1	21.5–22.1	N/A
	Device internal	22–22.3	21.3–21.7	24.1–25.1	N/A

According to Table 6, the ambient temperature ranged from 21.3 °C to 26.2 °C. Temperature differences for individual OCRs reach up to 6 °C. The participant's results were normalized to 23 °C of the device's internal temperature. The temperature correction was applied to the individual spectra following [23,40]. The correction factors for temperature differences between participants' laboratories are given in Figure 8. The correction factors were calculated using Formula (12) for RAMSES and Formula (15) for the HyperOCR sensors.

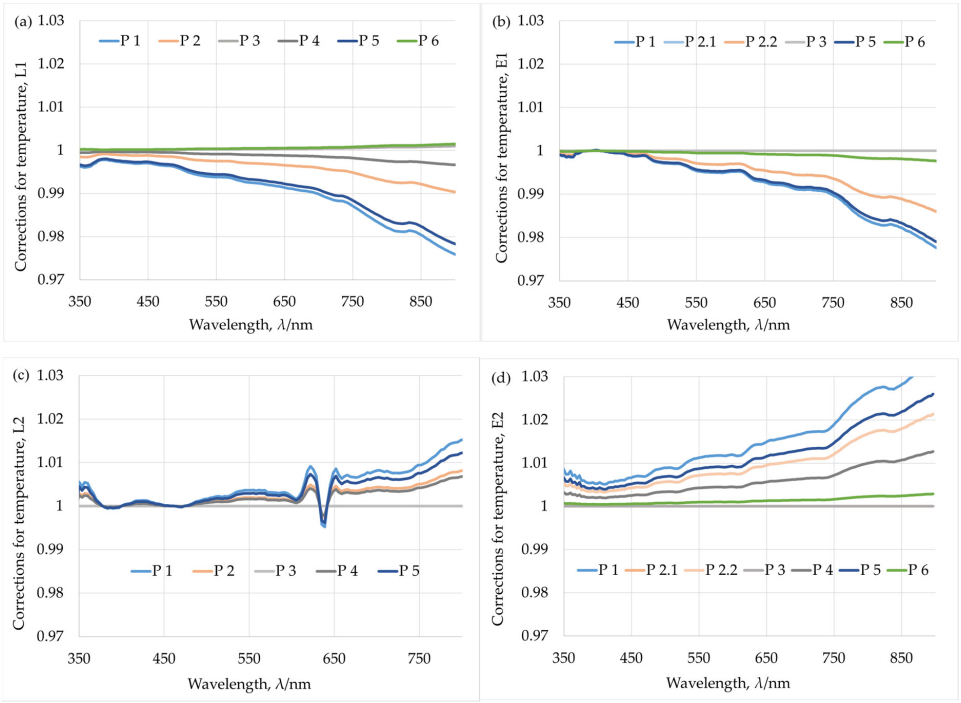


Figure 8. Correction factors for temperature differences in laboratories. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for the HyperOCR irradiance sensor SAT2072.

3.2.6. Combined Correction for Drift, Non-Linearity, and Temperature

The combined correction factors to compensate for the drift, non-linearity, and temperature differences in the participants are calculated using Formula (17) and given in Figure 9. Uncertainty bars show the relative standard uncertainty of combined correction coefficients, Formula (30).

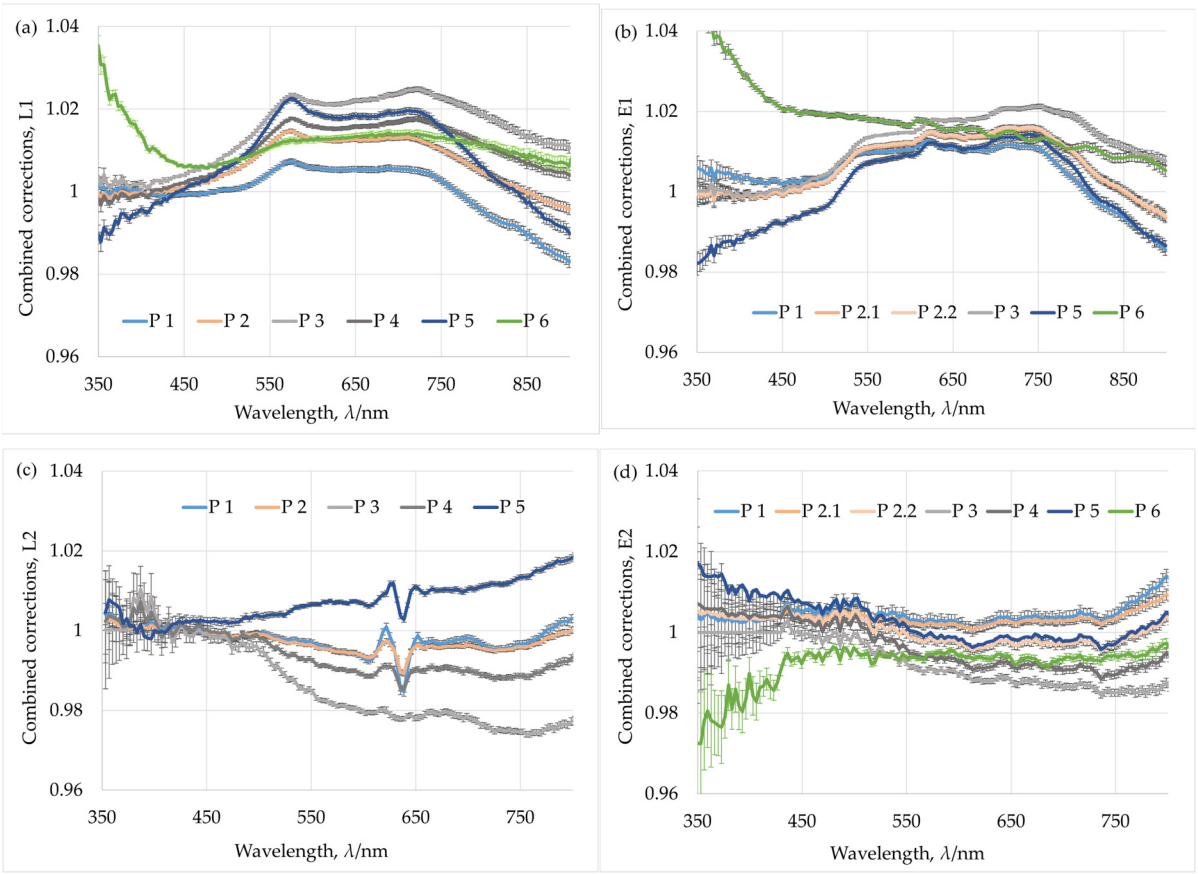


Figure 9. Combined drift, non-linearity, and temperature correction. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for the HyperOCR irradiance sensor SAT2072.

3.3. Results with Correction Applied

3.3.1. Relative Difference from CCV of Corrected Results

First, the data reported by the participants were recalculated, starting from the raw data files whenever available. Secondly, the instrument-specific corrections for temporal drift, non-linearity, temperature, and straylight were applied to the individual data. The agreement between the participants after applying the corrections is presented in this section. The relative difference in corrected responsivity coefficients from the respective consensus values for the radiance L and irradiance E radiometers is given in Figure 10.

After correcting for temporal drift, linearity, temperature, and straylight, in the spectral range from 400 nm to 800 nm for irradiance sensors and from 450 nm to 800 nm for radiance sensors, differences in the participant's results from consensus values stay within ±3%. Only one result of P5 for SAT2072 deviates slightly by over 3%. Referring to the SIRREX-7 experiments [11], and assuming for irradiance and radiance calibration uncertainties of participants, the secondary ranking, all results presented in Figure 10 would demonstrate satisfactory metrological consistency with determined CCVs.

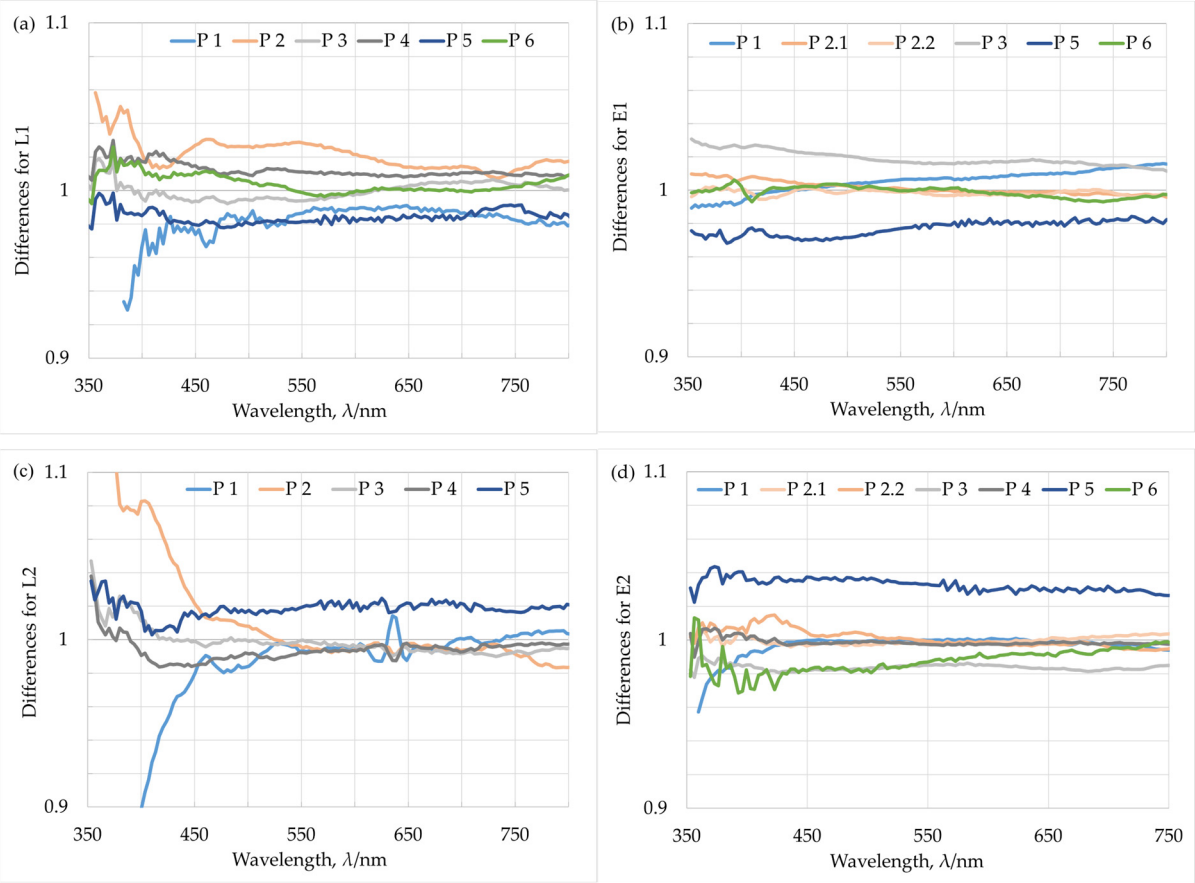


Figure 10. Relative difference from the CCV after applying all corrections. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598. (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for the HyperOCR irradiance sensor SAT2072.

3.3.2. Agreement of Participant's Results Using E_n Numbers

To evaluate the level of agreement of a participant's measurement result with the CCV, the performance statistic E_n number [26] is used. This statistic is also known as a conformance check to determine if a participant's result is satisfactory. The E_n number is calculated as:

$$E_n = \frac{x_{Li} - X_{ref}}{\sqrt{U_{Li}^2 + U_{ref}^2}} \quad (18)$$

where x_{Li} is the participant's i result, X_{ref} is the assigned value based on the results of all participants (CCV in this case), U_{Li} is the expanded uncertainty of the participant's result x_{Li}, and U_{ref} is the expanded uncertainty of X_{ref}. It is common to use a critical value of 1.0 with E_n numbers, where any result for which $|E_n| > 1$ is considered inconsistent. However, while E_n values in the range ($1 < |E_n| \leq 1.5$) can be considered as questionable results, values with ($|E_n| > 1.5$) represent unsatisfactory results.

Agreement of the participants' results with the comparison consensus value characterized in E_n numbers is given in Figure 11. Between all participants, reasonably good agreement is evident. Among the 20 results, most are within acceptable limits near ± 1, but each of the two comparison radiometers has one result with E_n values between 1.5 and 2, which is considered unsatisfactory.

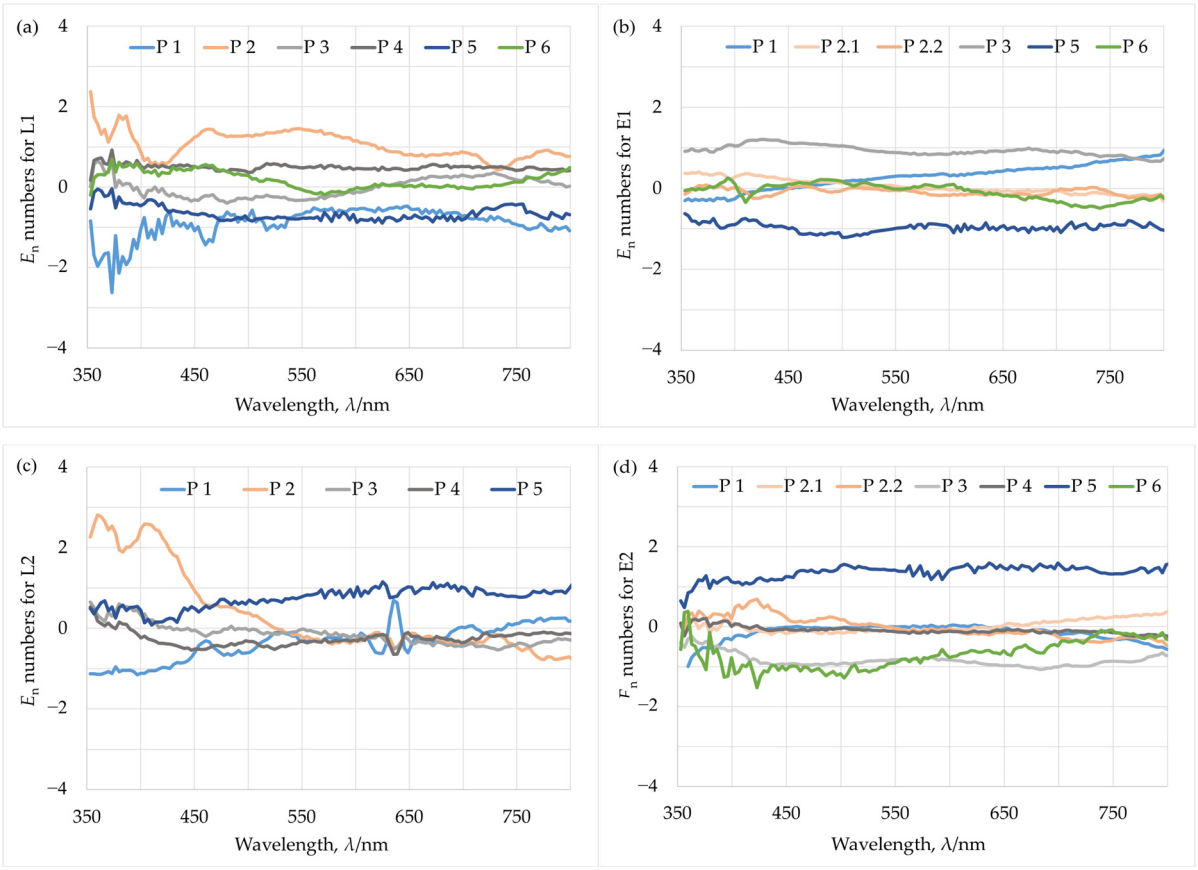


Figure 11. Agreement between participants shown with E_n numbers. (**a**) results for RAMSES radiance sensor SAM_81B0. (**b**) results for RAMSES irradiance sensor SAM_8598 (e.g., P2.2 indicates the results with the second lamp of the second participant). (**c**) results for the HyperOCR radiance sensor SAT2073. (**d**) results for the HyperOCR irradiance sensor SAT2072.

3.4. Measurement Uncertainty of Responsivity Calibration

3.4.1. Type B Contributions

The uncertainty analysis is based on the methodology of Guide to the Expression of Uncertainty in Measurement (GUM) [41]. Type B contributions to uncertainty are those that are not characterized by repeated measurements. The uncertainty budgets of calibrated hyperspectral radiometers include the following components with significant contributions to the laboratory comparison exercise [34,41,42]:

- Realization of the radiometric scale, including aging estimates of standards
- Spectral interpolation of irradiance/radiance values
- Distance from the working standard lamp to the calibrated DUT sensor or panel
- Wavelength assignment error of the radiometer pixels
- Operating current of the working standard lamp
- Alignment of the position of the lamp, radiometer, and panel
- Correction for directional-hemispherical spectral reflectance of the panel if required
- Scattered light control
- Variability of the calibration temperature

Differences between the laboratories due to different measurement conditions were treated by the Pilot during the analysis.

(1) Traceability of radiometric standards of participants

The uncertainty of the radiometric scale is taken as the uncertainty from the calibration certificate(s) of the following sources, depending on which source the hyperspectral radiometer was calibrated against by the participant:

1. The working standard lamp
2. The lamp—panel standard or
3. The integrating sphere or a transfer radiometer, depending on how the sphere's radiance is derived

(2) Uncertainty associated with the aging of the working standard

Drift of a new FEL lamp is less than 0.01%/h [43]. Thus, the usual drift estimate after 50 working hours is about 0.5%. However, unpredictable stepwise changes up to ±1% still may occur. As an example, the uncertainty associated with lamp aging for a lamp with a working time of 40 h (neglecting the above-mentioned stepwise changes) is:

$$u(aging) = \frac{0.5\%}{\sqrt{3}} \frac{40\ \text{h}}{50\ \text{h}} = 0.23\% \tag{19}$$

The last calibration of FEL lamps used for the comparison was often dated more than 5 years ago, and although the period of hours of use was around 10 to 20 h, one must be careful with too optimistic drift estimation. It is one of the most difficult components to evaluate, and its contributions may be significant, especially for the lamps with the smallest uncertainties. For monitoring the stability of FEL lamps during use [44], parameters such as lamp current, lamp voltage, and/or signals from an independent filter radiometer should be continuously recorded to discover any possible instability caused by lamp aging. Such evaluation requires:

1. Recording the lamp voltage, comparison with historical data
2. Having at least two standard lamps and performing regular comparisons between lamps
3. Using a monitoring radiometer concurrently with a lamp
4. Analyzing calibration history, if available
5. Regular stability check of lamps with a reference filter radiometer

Filter radiometers demonstrate superior stability compared to lamps; the alignment process using them for system verification is relatively simple, and they can easily be used for uncertainty evaluation. Two FELs with stable differences allow us to use longer recalibration intervals. If the difference between the two lamps or the drift detected by the monitoring radiometer is large in comparison with the expanded uncertainty of the lamps, then they should be recalibrated.

(3) Uncertainty associated with the interpolation of irradiance/radiance values

The wavelength scale of a hyperspectral radiometer is individual and unique; therefore, the irradiance calibration data from the FEL lamp must be transferred for each radiometer by determining the irradiance values at each radiometer pixel's wavelength. The spectral irradiance of the FEL lamps departs from the ideal Planckian distribution. It is usually approximated by multiplying the Planck function by an Nth-degree polynomial, where typically, the degree N varies between 3 and 7 [44–49]. Institutes calibrating the FELs are using different approximation/interpolation formulas. Some institutes also provide the interpolation formula together with the calibration certificates, and users can interpolate the irradiance values for each needed wavelength. Such interpolated irradiance values are covered with the uncertainty stated in the certificate. Some institutes provide irradiance data tables with a small wavelength step (0.25 nm–1 nm). In the case of such tables, linear interpolation can be applied without the need for an extra uncertainty contribution in

addition to the certificate. If the wavelength step provided in the certificate is 10 nm or larger, using a Planckian distribution-based interpolation formula is strongly advised.

For evaluation of the quality of the interpolation, the absolute (squared) residuals between certificates and interpolated points may be used. Dependence of the residual on wavelength is rather likely and is associated with larger errors at the beginning and end of the spectrum.

(4) Uncertainty associated with the wavelength assignment error of the radiometer

For hyperspectral radiometers, the accuracy of the relation between pixel number and wavelength must be established to obtain the signal as a function of wavelength λ, and uncertainty $u_{\Delta\lambda}(E)$ is associated with this accuracy.

$$\frac{u_{\Delta\lambda}(E)}{E(\lambda)} = \frac{u(\Delta\lambda)[\mathrm{nm}]}{\sqrt{3}} \frac{100}{E(\lambda)} \left|\frac{\partial E}{\partial \lambda}\right| [\%]. \tag{20}$$

Here, the wavelength error of the radiometer is $\Delta\lambda = 0.3$ nm according to the manufacturer's specification, and $|\partial E/\partial\lambda|$ depends on the spectral irradiance $E(\lambda)$ of the radiation source.

(5) Uncertainty associated with the distance from the working standard lamp to the calibrated sensor/panel

The standard uncertainty of the reference distance (stated in the calibration certificate, usually 500.00 mm) from a lamp to the radiometer consists of at least two components: systematic, related to the calibration of the length instrument used, and random, due to a particular measurement procedure. The uncertainty component associated with the mechanical stability of the holders should also be accounted for, especially for the distance setting and measurement with the smallest possible uncertainties. For determining the impact of distance variations on the uncertainty, the inverse square law for point light sources (1) can be used.

Taylor series expansion of (1) yields that a small deviation δd from the standard distance d will cause a relative change in radiant intensity following

$$E(d + \delta d) \cong E(d)\left(1 - \frac{2 \cdot \delta d}{d}\right) \tag{21}$$

For example, if the combined standard uncertainty of the distance is 0.5 mm, the resulting relative standard uncertainty of irradiance will be about $u = 0.2\%$, with coverage factor $k = 1$. A special approach may be needed if the realized distance is different from the standard distance of 500 mm. The spectral irradiances at distance d can be determined using Equation (1). If the reference distance of 500 mm is used, the uncertainty of the lamp offset Δd is not needed.

$$\frac{u_d(E)}{E} \cong \frac{2u(d)}{d} \tag{22}$$

The offset Δd and its uncertainty $u(\Delta d)$ shall be accounted for if the realized distance d differs from the standard distance d_{ref}.

$$\frac{u_{\Delta d}(E)}{E} \cong \frac{2u(\Delta d)}{d} \left|1 - \frac{d}{d_{\mathrm{ref}}}\right| \tag{23}$$

Using the certified distance (500 mm) gives the simplest reliable distance-related uncertainty estimate. For other distances, the additional uncertainty of the lamp's offset shall be accounted for. The lamp's offset is often determined from a set of 5 to 10 measured distances along with the net signals at those distances using a least-squares fit.

(6) Uncertainty associated with the operating current of the working standard lamp

The operating electrical current of a lamp is a critical parameter for spectral irradiance. According to [39,50], the relative uncertainty of the lamp's spectral irradiance as a function of the lamp's current uncertainty in the UV/VIS/IR range can be estimated as

$$\frac{u_{\mathrm{cur}}(E(\lambda))}{E(\lambda)} \cong 0.0006\left(\frac{654.6[\mathrm{nm}]}{\lambda[\mathrm{nm}]}\right)u(I)[\mathrm{mA}]. \tag{24}$$

Formula (24) can be easily tested experimentally by deviating the lamp's operating current slightly away from its nominal value and measuring the respective signal change. This was checked for the lamps used at the Pilot, and good consistency with (24) was evident.

The component $u(I)$ includes accuracy and stability of the current source, shunt, and voltage measurement, and can be estimated as:

$$u(I) = \sqrt{u_{\mathrm{stab}}^2 + u_{\mathrm{measure}}^2} \tag{25}$$

Here u_{stab} is the standard uncertainty of a lamp current stability, u_{measure} is the standard uncertainty of a lamp current measurement, including the standard uncertainty of a shunt.

(7) Uncertainty associated with the correction for directional-hemispherical spectral reflectance to bidirectional

Directional-hemispherical spectral reflectance $R(6^\circ/\mathrm{H})$ or $R(8^\circ/\mathrm{H})$ of the diffuse reflectance plaque is typically specified in the manufacturer's certificate. This must be transformed to the bidirectional reflectance factor $R(0^\circ/45^\circ)$ needed for radiance calculations. Based on the Pilot's internal data, combined with published data [18,51], the recommended adjustment is:

$$R(0/45) = R(8/\mathrm{H})\cdot 1.023 \tag{26}$$

The relative standard uncertainty for this correction factor should be at least $u(1.023) = 0.25\%$, with $k = 1$.

(8) Uncertainty associated with the alignment of the lamp position

Alignment errors of the lamp across the optical axis less than ± 1 mm in the y or z directions (Figure 12), rotation of the lamp around y and z axes less than $\pm 0.1^\circ$, and around x axes less than 2° will cause uncertainty in the irradiance less than 0.1% . . . 0.2% [50]. This accuracy can be achieved only by very careful alignment using lasers and levels.

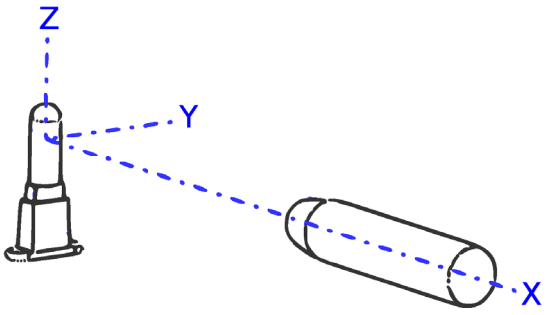


Figure 12. Coordinate system of irradiance calibration setup, in which the axes origin is placed inside the vertical FEL lamp, and the sensor is positioned and aligned along the *x-axis*.

(9) Uncertainty associated with the alignment of the radiometer and panel

Positioning errors in the input optics of the radiometer lead to an additional irradiance uncertainty of about 0.1%. Alignment errors concerning the diffusing panel led to an additional uncertainty of radiance of about 0.1%.

(10) Uncertainty associated with the variability of the calibration temperature

For irradiance comparisons, the ambient temperature range for calibrations using FEL lamp standards recommended is between 20 °C and 25 °C.

Room temperature in the laboratory should vary no more than ± (1 . . . 1.5) °C for the whole period of the comparison measurements. When calibrating hyperspectral spectroradiometers with silicon detectors, the estimate of standard uncertainty due to temperature variability in the spectral region from 400 nm to 700 nm is around 0.1%/°C and will increase up to 0.6%/°C for longer wavelengths (950 nm).

(11) Scattered light control

External light reflected or scattered in the surroundings may hit the radiometer's measuring head and cause an additional signal that does not originate directly from the reference light source. It must be suppressed by using baffles and corrected together with the dark signal. The placement and number of baffles should ensure that the ambient radiation is less than ~0.1% . . . 0.3% of the net signal. The diameter of suitable baffles depends on the distance between the lamp and the radiometer.

3.4.2. Type a Contributions

Type A contributions are the uncertainty contributions that can be evaluated using statistical analysis of repeated measurements [41].

(1) Reproducibility

The standard deviation of the mean is calculated from the results of independent measurements, where each independent measurement is carried out after realignment of the working standard, and the radiometer is subject to calibration.

(2) Repeatability

The standard deviation of the mean is calculated from a set of spectra when calibrating the hyperspectral radiometer.

(3) Uncertainties of corrections applied by the Pilot

The relative standard uncertainty of each correction factor from Equations (11) to (16) has two components. For example, the uncertainty of the correction factor due to temperature difference according to (12) and (15) is calculated as

$$u(C_T(\lambda)) = \sqrt{[u(\alpha_T)\cdot\Delta T]^2 + [\alpha_T\cdot u(\Delta T)]^2} \quad (27)$$

Here $u(\alpha_T)$ is the standard uncertainty (coverage factor k = 1) of the temperature coefficient $\alpha_T(\lambda)$, and $u(\Delta T)$ is the standard uncertainty of the temperature difference $\Delta T = T_{lab} - T_{ref}$.

Similar expressions are also valid for the uncertainties in the non-linearity and drift corrections, but as a second member ($\alpha_{\text{lin}}\cdot u(S_{DN})$) or ($\alpha_t\cdot u(\Delta t)$) is much smaller than the first member; only the first is used. Thus:

$$u(C_{\text{lin}}(\lambda)) = u(\alpha_{\text{lin}})\cdot S_{DN}(\lambda) \quad (28)$$

$$u(C_t(\lambda)) = u(\alpha_t)\cdot\Delta t. \quad (29)$$

The combined uncertainty of correction factors containing multiplications is the square root of the sum of squares of each of the component terms:

$$u(C_{comb}(\lambda)) = k\sqrt{[u(C_{\text{lin}}(\lambda))]^2 + [u(C_T(\lambda))]^2 + [u(C_t(\lambda))]^2}. \quad (30)$$

where k is the coverage factor.

3.5. Uncertainties of the Participant's Results

3.5.1. Traceability of Radiometric Standards of Participants

The participants had different SI-traceability routes for their lamps. Two participants had lamps directly calibrated at the National Institute of Standards and Technology (NIST), one had traceability to NIST scale via Gooch and Housego (now Optronics Laboratories), one via Optronics Laboratories, one had the lamps calibrated at the National Physical Laboratory (NPL), and one had traceability to the Metrology Research Institute (MRI) of Aalto University, Finland. Some participants calibrated their own working standard lamps onsite against those SI-traceable lamps. The relative standard uncertainty of the radiometric standards of participants is shown in Figure 13. The calibration uncertainties of the FEL lamps are in Figure 13a, uncertainties of the lamp-panel setups or integrating spheres are in Figure 13b. Four participants used two lamps for the comparison measurements, while the remaining two used a single lamp. The calibration uncertainty of the bidirectional reflectance factor $R(0°/45°)$ of the plaques used in the lamp-panel setups for radiance calibration is shown in Figure 13c. The reference reflectance panel of P5 was calibrated for the 8°: hemispherical reflectance geometry. To correct it for the proper reflectance factor geometry at 0°:45°, a correction factor of 1.023 was advised based on the Pilot internal data combined with published data by NPL and NIST [18,51].

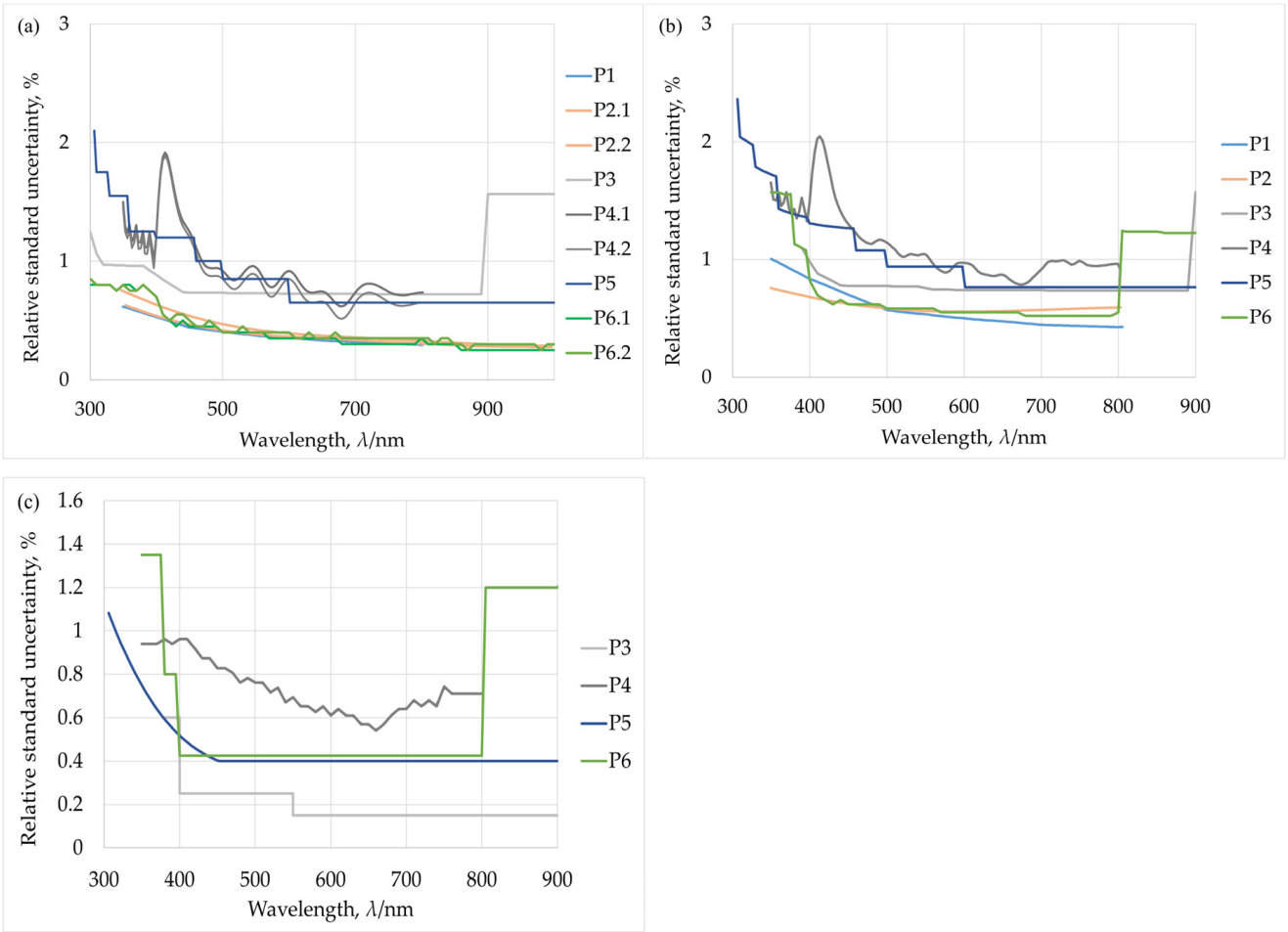


Figure 13. Calibration uncertainty of the radiometric standards. (**a**) FEL lamps, where each curve indicates a lamp of a given participant (e.g., P6.2 indicates the second lamp of the sixth participant). (**b**) Radiance standards—lamp-panel setups or integrating spheres. (**c**) Diffusing reflectance panel standards.

3.5.2. Calibration Uncertainties Reported by the Participants

The relative standard uncertainty of the responsivity coefficients of the radiometers reported by participants is given in Figure 14. The uncertainties stated by the participants for the responsivity of comparison standards are rather similar. Although the calibration uncertainties of the radiometric standards (Figure 13) differ more than a factor of two among the participants, the spread between participants is reduced significantly due to the contributions discussed in Section 4.2 when these standards are applied in responsivity calibrations (Figure 14).

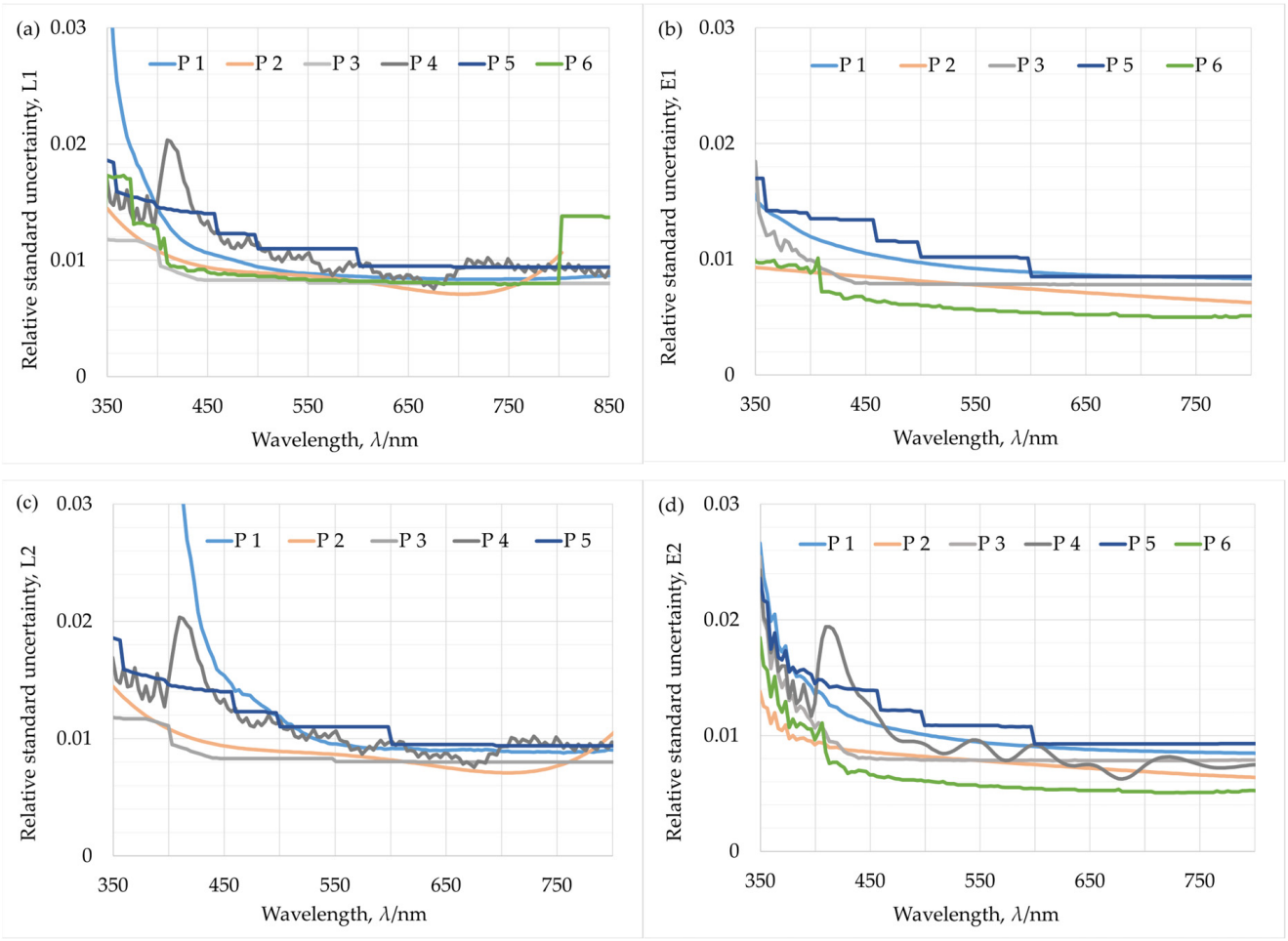


Figure 14. Standard uncertainty of responsivity coefficients reported by participants. (**a**) Uncertainty for RAMSES radiance sensor SAM_81B0. (**b**) uncertainty for RAMSES irradiance sensor SAM_8598. (**c**) uncertainty for HyperOCR radiance sensor SAT2073. (**d**) uncertainty for HyperOCR irradiance sensor SAT2072.

3.5.3. Combined Calibration Uncertainties After Adding Correction Contributions

The corrections for temporal drift, non-linearity, and temperature differences that were evaluated and corrected for by the Pilot brought additional uncertainty contributions. The combined relative standard uncertainty of the responsivity coefficients after adding the contributions of corrections applied by the Pilot is given in Figure 15. As the combined uncertainty of corrections was around 0.1 . . . 0.2% in the spectral range from 400 nm to 800 nm, and up to 1% in shorter wavelengths, the effect of the corrections to the combined relative uncertainty is insignificant. The biases before correcting varied from ±1 . . . 2% up to ±5%, see Figure 9.

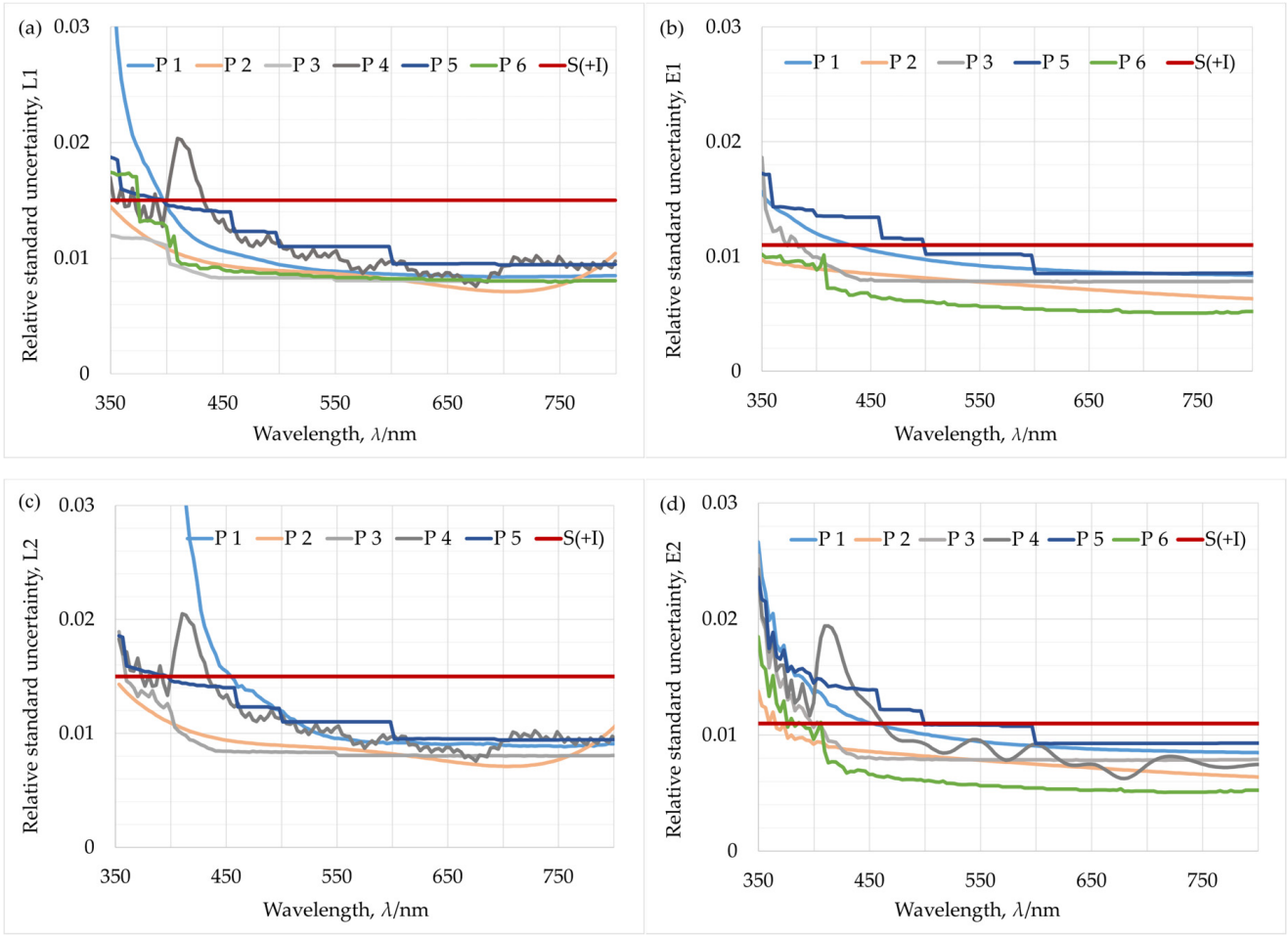


Figure 15. Uncertainty of RAMSES responsivity coefficients after correction by the Pilot. (**a**) Uncertainty for RAMSES radiance sensor SAM_81B0. (**b**) uncertainty for RAMSES irradiance sensor SAM_8598. (**c**) uncertainty for HyperOCR radiance sensor SAT2073. (**d**) uncertainty for HyperOCR irradiance sensor SAT207. Uncertainty according to SIRREX-7 rank-I is shown with a red line.

The combined calibration uncertainties can be classified according to the SIRREX-7 experiments [11] where the significance of uncertainties is ranked as primary, secondary, or tertiary. In the SIRREX-7 experiments, the uncertainties were ranked based on the difficulty of reducing the respective contributions. Primary contributions include the uncertainty of standards sources and minimum contributions from adjustment and operation; they are close to the state of the art and very difficult to reduce. Secondary and tertiary contributions, respectively, can be reduced much more easily. In the spectral range from 500 nm to 800 nm, all responsivity uncertainties of the participants could be classified as primary (Figure 15). For shorter wavelengths, the uncertainties of participants exceed the primary for most participants. Referring to the results of [18] where agreement between applied FEL lamp irradiance standards was within ±1.5%, and discrepancies between the radiance standards were ±4%, in the current study, the variability of the OCR responsivity coefficients for radiance was smaller and for irradiance comparable. The uncertainty claims were mostly confirmed by the comparison results (Figures 10 and 11).

4. Discussion

4.1. Interpolation of Standard Lamps

The wavelength scale of a hyperspectral radiometer is individual and unique; therefore, the irradiance calibration data from the FEL lamp must be transferred for each radiometer

by determining the irradiance values at each pixel's wavelength. As an example, the linear interpolation error of the tabulated calibration data of an incandescent calibration standard compared to the Planckian function-based interpolation formula is shown in Figure 16. At 350 nm, the error due to linear interpolation of irradiance data tabulated with a wavelength step of 10 nm is about 0.5%. With the 5 nm step, the error is about 0.12%, which is ten times smaller than what is usually stated for irradiance uncertainty. For data with steps smaller than 5 nm, linear interpolation is suitable, and related errors in the range from 300 nm to 900 nm can be considered insignificant.

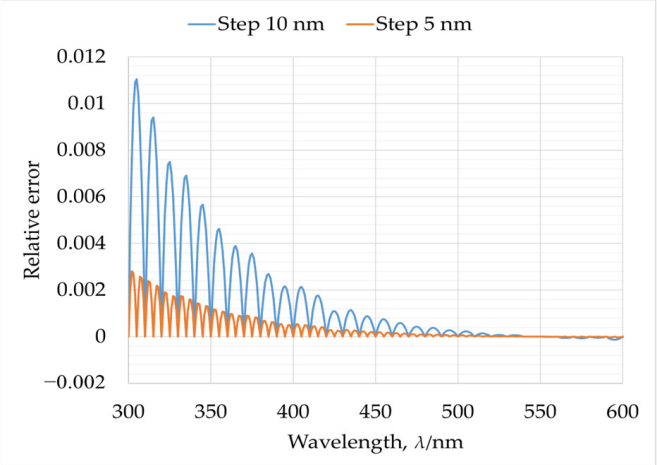


Figure 16. Linear interpolation error as a function of wavelength step of the tabulated calibration data.

Irradiance interpolations performed by the participants were mainly in sufficiently close conformance with the calibration certificate of the used radiation source. However, in two cases, the interpolated spectral irradiances of two standard sources used by participants showed a spectral shift error by one pixel (Figure 17). Empty blue circles (Figure 17a,b) are the initial points from the submitted comparison report; filled blue circles are shifted by one pixel towards shorter wavelengths. After the shift, the points of both radiometers match the calibration curve of the source.

Using the spectrally shifted-by-one-pixel irradiance or radiance values for calculating the radiometer's calibration coefficients could introduce a relative error of more than 10% in the UV-blue part of the spectrum (Figure 17c). Such an error is significantly larger than the uncertainty usually stated for calibration coefficients. Figure 17d presents the agreement obtained with two different FEL lamps for two radiometers (SAM_8598 and SAT2072). The initial reported values are shown with solid lines. Normal agreement between calibrations with different FELs is within ±1% (SAT2072), and a significant difference of about 10% for SAM_8598 implies an error in measurement or data handling. After shifting the irradiance spectrum determined for SAM_8598 by one pixel, a similar agreement for both radiometers became evident. Calibration coefficients of the RAMSES are inversely proportional to the calibration coefficients of HypeOCR, and as expected, after applying the shift, a mirror symmetry between the ratios is visible above 400 nm (Figure 17d). Therefore, interpolating the source spectrum and intermediate calculations performed during data handling should be performed with great care. Checking the agreement of interpolated values by comparing them with the reference calibration data is highly advisable.

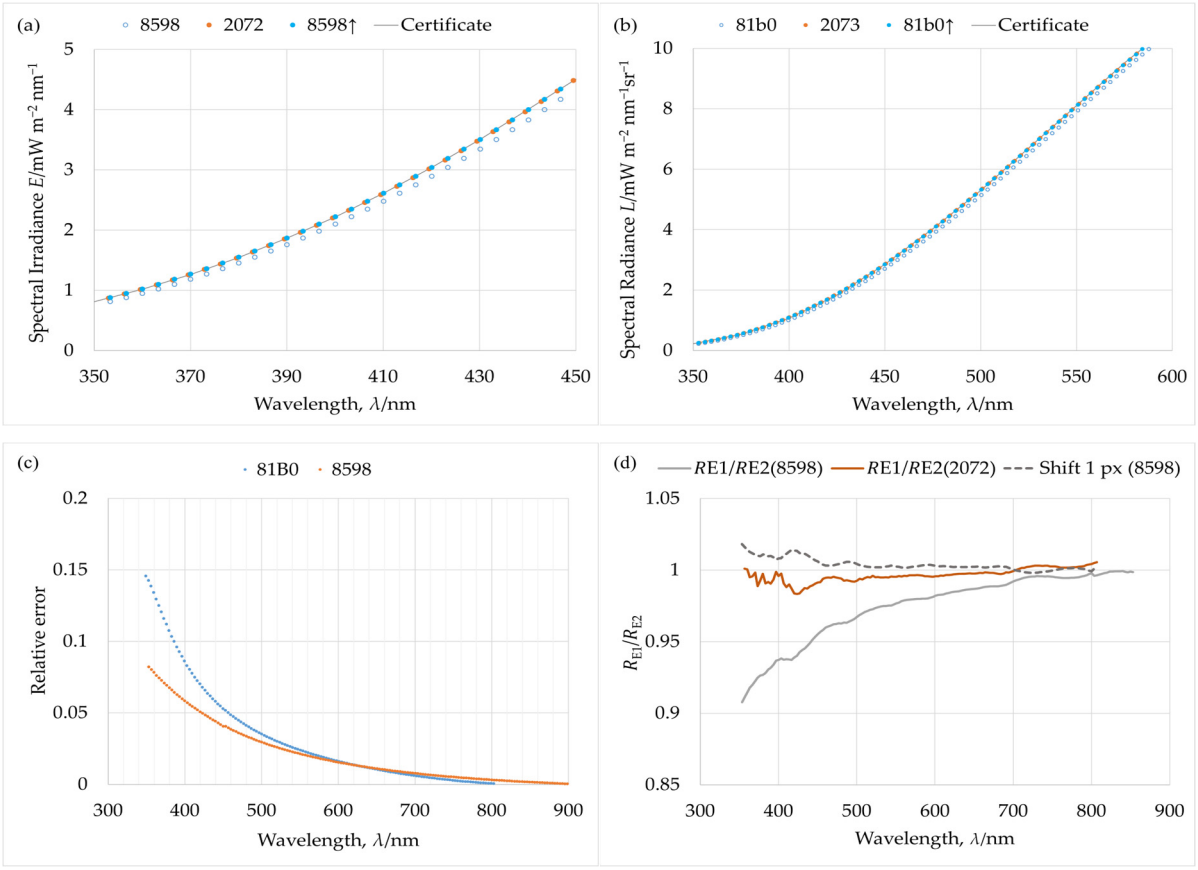


Figure 17. The relative error of the interpolated values due to spectral shift in wavelength scales by 1 pixel. (**a**) Interpolated values for irradiance and (**b**) radiance sensor as circles, the calibration curve as a continuous line. Initial disagreement with the calibration curve shown with empty circles was eliminated by a pixel shift. (**c**) Ratios of the calibration coefficients -1 obtained with the shifted and correct values. (**d**) The calibration coefficients were measured with two FELs, before (solid gray) and after the shift (dashed gray) by 1 pixel of one of the interpolated FEL values.

4.2. Errors Arising Due to Pixel Shift

During the data handling, three types of errors may happen: (1) spectrally shifted-by-one-pixel irradiance or radiance values; (2) spectrally shifted-by-one-pixel calibration coefficients; (3) spectral shift by one pixel for both the irradiance/radiance values and the calibration coefficients.

All these errors will cause a bias proportional to the relative change in respective values with one pixel step, see Figure 18. A 1-pixel shift in the radiation spectrum will cause smooth curves, with maximum bias in the UV range (up to ±15%) and a diminishing effect in the NIR range due to the nature of the Planckian curve. The raw output signals and calibration coefficients show many spectral peaks and will, when shifted, produce pulsating bias curves with significant effects in the UV and NIR parts of the spectrum. The combined error can reach ±20% in the UV, and about ±5% in the NIR range. When shifts occur in the calibration coefficient because of mistakes in data handling, then significant tilts or pulsation effects will be observed.

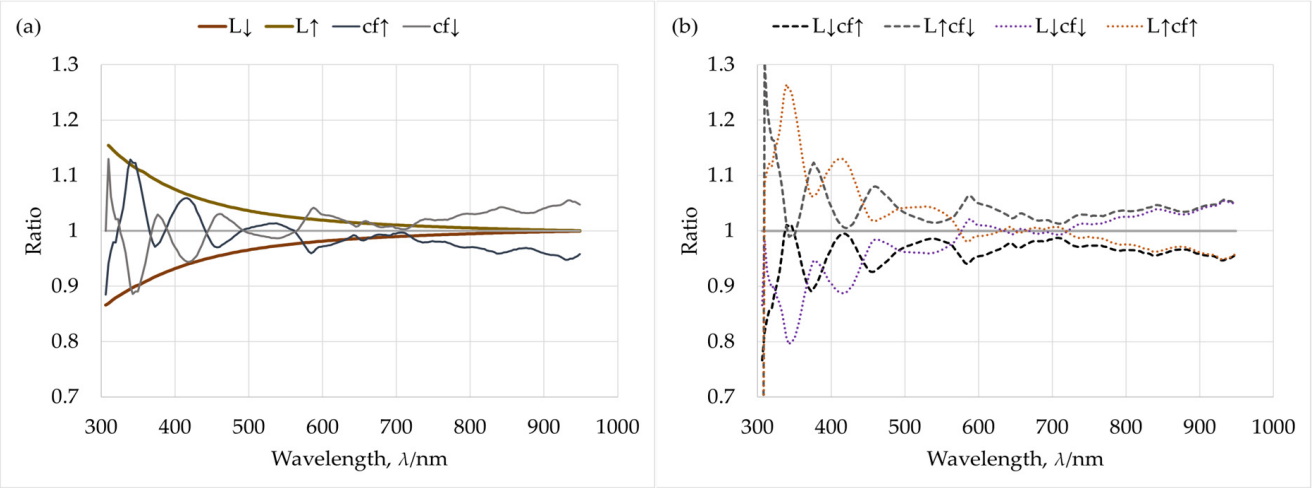


Figure 18. Ratio of calibration coefficients due to a 1one pixel spectral shift. (**a**) shifting radiance (L↑ and L↓) values will cause smooth error curves, while shifted values of calibration coefficients (cf↑ and cf↓) will yield pulsating curves; (**b**) combined errors if both values have the spectral shift in one pixel.

4.3. Comparison with the SIRREX-7 Experiments

Differences between the results reported by the participants are compared with uncertainties in irradiance and radiance calibrations as determined from the SIRREX-7 experiments [11]. In the spectral range from 400 nm to 800 nm, for non-corrected results (Figure 19a,c,e,g) almost all differences between the participant's results did stay in the rank-II uncertainty limits. After correcting for data handling and different measurement conditions (Figure 19b,d,f,h), the radiance *L* differences were within rank-I uncertainty limits, and the irradiance *E* differences were mostly within rank-I.

In Figure 19a,c, e, and g, the agreement between the participants' results as initially reported is shown. However, all results were converted to the same sensor-specific units (Equations (4), (5), (7) and (8)) and, in the case of HyperOCR, to the same integration times as by the latest factory calibration (Equations (9) and (10)). Substantially better consistency was obtained with unified data handling starting from the raw spectra and after correcting for biases due to different measurement conditions (Figure 19b,d,f,h). Regarding future laboratory comparisons, sufficiently detailed and easily accessible documentation, better protocols for the measurement and data handling, suitable quality checks, and some training, hopefully, would allow satisfactory agreement in the initially reported results of participants.

For further improvement of the calibration uncertainty of OCRs, development of the radiometric scales in laboratories is needed, but these activities are very resource- and time-consuming. The SI-traceability of radiometric standards can be improved by using suitable NMIs as calibration providers. However, without the enhancement of requirements for the laboratory conditions, aging and maintenance of radiometric standards, and adequate auxiliary equipment, the effect may be modest. An improvement of the radiometric scale can reduce the calibration uncertainty of OCRs only after the combined contribution to the uncertainty from all other reasons is comparable or smaller than the contribution from the radiometric scale.

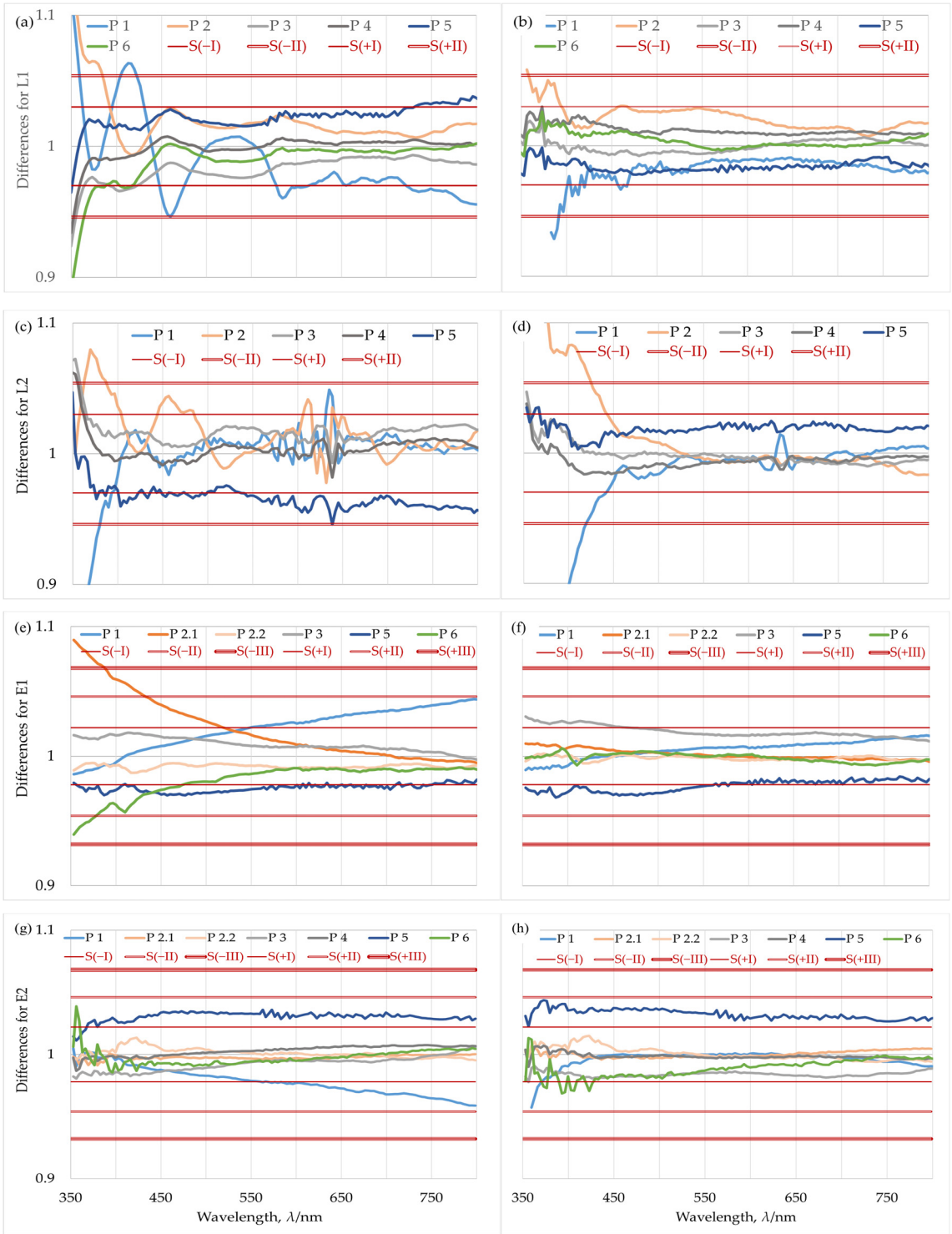


Figure 19. The relative difference from CCV of L and E sensors with the frame of SIRREX-7 uncertainty ranking. Left: non-corrected responsivity coefficients; right: corrected responsivity coefficients. (**a**,**b**) RAMSES radiance sensor SAM_81B0; (**c**,**d**) HyperOCR radiance sensor SAT2073; (**e**,**f**) RAMSES irradiance sensor SAM_8598; (**g**,**h**) HyperOCR irradiance sensor SAT2072.

4.4. Definition of Calibration Coefficients

When measuring a constant source using hyperspectral radiometers like the OCR models in this study, twelve settings with sensitivity from 1 to 2048 were possible, each differing by a factor of 2. Sensitivity can be set by selecting an integration time from twelve values from 4 ms, 8 ms, 16 ms, . . . to 8.192 s. In metrology, to provide firm SI traceability,

before use, each measurement range should be calibrated, but due to limitations of the radiation sources used for the calibration of OCRs, the calibration procedure cannot be realized for all integration times. The integration times of the radiometer (t_{cal} and t_{meas}) shall be explicitly accounted for in Equations (2) and (3) as in [31,34,52–54]. The normalization of the signal to the used integration time will make the calibration coefficients from different sources comparable and enable us to use the calibrated instrument with different integration times later for different purposes. Therefore, the calibration coefficients shall always be clearly defined in the documentation of the radiometer and software provided by the manufacturer. During the current study, information about the determination and use of calibration coefficients was likely insufficient, as discussed before.

5. Conclusions

We presented the results of the international laboratory comparison measurements of the spectral irradiance and radiance by six participants in the framework of the Copernicus FRM4SOC Phase 2 in 2022–2023. Four radiometers were sent in turn to each participant to obtain responsivity calibration coefficients for the radiometers using their in-house SI-traceable spectral irradiance and radiance facilities and their usual calibration procedures. The purpose of the exercise was to compare the measurement and data handling processes followed by each participant, calibrating the same set of transfer radiometers, without requiring them to modify their routine procedure. Participants determined the responsivity calibration coefficients in the wavelength range from 350 nm to 900 nm. The transfer standards were the two most commonly used OCR models: two TriOS RAMSES (radiance and irradiance) and two Sea-Bird HyperOCR hyperspectral radiometers (radiance and irradiance). The quantitative results of the comparison are presented in terms of differences between each participant's measurement and the comparison consensus value (CCV), and as En numbers, Equation (18).

The participants had different SI-traceability routes for their standard lamps, lamp-panel setups, or integrating spheres with uncertainties from 0.3% to 2% within the wavelength range of this study, but the expanded calibration uncertainties ($k = 2$) for the comparison radiometers stated by the participants were much closer, around 2%. The stated uncertainty fulfilled the IOCCG protocol requirement of 1% standard uncertainty ($k = 1$) in calibration.

After correcting for temporal drift, linearity, temperature, and straylight, in the spectral range from 400 nm to 800 nm for irradiance sensors and from 450 nm to 800 nm for radiance sensors, differences in the participants' results from consensus values were within $\pm 3\%$. Agreement in E_n numbers was also mostly satisfactory: from a total of 20 results, only two results had absolute values of E_n numbers larger than one, around 1.5 . . . 2. The degree of equivalence presently evident should be maintained and improved. This requires a special inter-laboratory comparison exercise to be regularly arranged for OCR calibration laboratories. Assuming that in the future, data handling errors can be avoided and that each participant addresses the relevant corrections for hyperspectral instrument characteristics independently (rather than relying on the Pilot), the results of this comparison demonstrate the capability of OCR calibration laboratories to meet the IOCCG protocol requirement of approximately 1% ($k = 1$) calibration uncertainty for irradiance, and slightly higher for radiance. The observed agreement among participants within $\pm 3\%$, as confirmed by E_n values, can be interpreted as validation of their stated uncertainty budgets. It is important to emphasize that the 1% uncertainty requirement at a coverage factor of $k = 1$ translates to an acceptable difference of up to 3% between participants. Although this may initially seem counterintuitive, it aligns with the statistical interpretation of coverage factors in uncertainty evaluation. Specifically, a standard uncertainty ($k = 1$) encompasses

approximately 67% of the distribution of possible measurement results. In contrast, En values are calculated using expanded uncertainties ($k = 2$), which cover about 95% of the distribution and are typically double the $k = 1$ values. For two independent measurements each with a standard uncertainty of 1% ($k = 1$), the combined expanded uncertainty ($k = 2$) is calculated using the root-sum-of-squares method, resulting in approximately 2.83%. Therefore, a difference approaching 3% between measurements is statistically expected and meets the IOCCG protocol requirement.

A specific calculation formula was needed for the responsivity coefficients of each transfer radiometer. The guidance provided within the comparison protocol was likely insufficient and did not cover all the steps and details of data handling. Thus, significant recalculation and/or reformatting of submitted data by the Pilot was necessary to make a preliminary comparison of the results from different laboratories possible.

The Pilot laboratory evaluated the known systematic effects due to different measurement conditions in different laboratories and corrected the biases using uniform algorithms. Corrections applied by the Pilot to the submitted results were significant if compared with the stated uncertainties, changing relative differences between the participants from $\pm 1 \ldots 2\%$ up to $\pm 5\%$. To achieve smaller corrections, improved transfer radiometers would be desirable. For example, if transfer radiometers through internal temperature stabilization exhibit temperature sensitivity of <0.1%/°C, the differences in laboratory temperatures would not need corrections. And more stable transfer radiometers with responsivity drift of <1%/year would not need corrections for drift during comparison with similar timescales and uncertainty requirements.

Differences in hardware and software of the OCRs used in the comparison caused major problems for participants. As the calibration and data handling procedures have some differences depending on the radiometer model, software, and data formats, several data handling errors, such as (I) pixel shifts, (II) interpolation errors, (III) errors during intermediate calculations, and (4) errors related to correcting biases, were present. The errors imply that the initial available documentation and the intermediate checks were insufficient. Eventually, a satisfactory metrological consistency of comparison results could be achieved only after having corrected these errors during the data handling. This experience sets a precedent to avoid such inconveniences for new round robins of a similar nature.

Recommendations for the establishment of common processes in OCR calibration laboratories and for future laboratory comparisons are as follows:

- Harmonization of the hardware and software of OC radiometers.
- A uniform detailed protocol for measurements and data handling.
- Training of the laboratory staff according to the specific instrument types.
- More stable transfer radiometers and a shorter comparison duration.
- Standardization of calibration conditions.
- Sufficiently detailed calibration and characterization guidelines for the OCR calibration laboratories.
- Regular laboratory comparisons to maintain and improve the quality of the field data.

The implications of this experiment highlight the need for additional corrections when working with hyperspectral instruments. These include addressing detector non-linearity, thermal sensitivity, spectral stray light, and drift. Although this exercise focused solely on laboratory comparisons, these effects are clearly present during in situ measurements as well. Therefore, to achieve accurate ocean color radiometry (OCR) measurements, it is essential to correct for these effects during in situ data acquisition and processing.

Author Contributions: V.V., I.A. and R.V. formulated the original concept of this study. I.A., A.B., M.E.F., R.L., S.M., M.O., C.R. and E.R. performed the measurements. V.V. and I.A. processed and analyzed the data. R.V., J.K., A.B., K.J.V., J.I.G. and E.K. advised on the analysis. V.V. and I.A. prepared the figures. V.V. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The work presented in this article was performed in the framework of the project Fiducial Reference Measurements for Satellite Ocean Color (FRM4SOC) Phase-2 (EUMETSAT contract no. EUM/CO/21/460002539/JIG) funded by the European Commission (Copernicus Programme) and implemented by EUMETSAT. The research has also been funded through an Estonian Research Council grant PRG2646 "Methods, Traceability and Validation of the In-Water Ocean Color Measurements". The work was also supported by the laboratory facilities of Tartu Observatory, University of Tartu, funded by the Estonian Research Council project TT8 and the Estonian State Shared Service Center project KosEST.

Data Availability Statement: The comparison results are anonymized to avoid any risk of bias. The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: The authors would like to thank Carol Johnson, Giuseppe Zibordi, and Julian Gröbner for valuable suggestions and comments during the Copernicus FRM4SOC—Phase 2 project.

Conflicts of Interest: The authors declare no conflict of interest. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. This study does not constitute endorsement of any of the products tested by any of the authors or their organizations.

Abbreviations

The following abbreviations are used in this manuscript:

CCV	Comparison consensus value
CEOS	The Committee on Earth Observation Satellites
CIPM	Comité International des Poids et Mesures (International Committee for Weights and Measures)
DUT	Device under test
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
FRM	Fiducial reference measurements
FRM4SOC	Fiducial Reference Measurements for Satellite Ocean Color
GUM	Guide to the Expression of Uncertainty in Measurement
IOCCG	International Ocean Color Coordination Group
MLML	Moss Landing Marine Laboratories
MOBY	Marine Optical Buoy
N/A	Not available
NIR	Near-infrared
NIVA	Norsk Institutt for Vannforskning (The Norwegian Institute for Water Research)
NMI	National Metrology Institute
NOAA	National Oceanic and Atmospheric Administration
NPL	National Physical Laboratory
OCR	Ocean Color Radiometer
QTH	Quartz Tungsten Halogen
SB	Sea-Bird Scientific
SeaWiFS	Sea-Viewing Wide Field-of-View Sensor
SI	International System of Units
SIMBIOS	The Sensor Intercomparison for Marine Biological and Interdisciplinary Ocean Studies

SIMRIC-2	The Second SIMBIOS Radiometric Intercomparison
SIRREX	SeaWiFS Intercalibration Round Robin Experiments
TO	Tartu Observatory, University of Tartu
UT	University of Tartu
Y	Data presented

References

1. Banks, A.C.; Vendt, R.; Alikas, K.; Bialek, A.; Kuusk, J.; Lerebourg, C.; Ruddick, K.; Tilstone, G.; Vabson, V.; Donlon, C.; et al. Fiducial Reference Measurements for Satellite Ocean Colour (FRM4SOC). *Remote Sens.* **2020**, *12*, 1322. [CrossRef]
2. The CEOS Cal/Val Portal. Cal/Val Portal. Available online: https://calvalportal.ceos.org (accessed on 29 April 2025).
3. Goryl, P.; Fox, N.; Donlon, C.; Castracane, P. Fiducial Reference Measurements (FRMs): What Are They? *Remote Sens.* **2023**, *15*, 5017. [CrossRef]
4. *JCGM 200:2008*; International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM). Joint Committee for Guides in Metrology: Sèvres, France, 2011.
5. Vabson, V.; Ansko, I.; Duong, K.; Vendt, R.; Kuusk, J.; Ruddick, K.; Bialek, A.; Tilstone, G.H.; Gossn, J.I.; Kwiatkowska, E. Complete characterization of ocean color radiometers. *Front. Remote Sens.* **2024**, *5*, 1320454. [CrossRef]
6. CIPM. *Mutual Recognition of National Measurement Standards and of Calibration and Measurement Certificates Issued by National Metrology Institutes*; CIPM: Paris, France, 1999.
7. Mueller, J.L. *The First SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-1, July 1992*; Hooker, S.B., Firestone, E.R., Eds.; National Aeronautics and Space Administration, Goddard Space Flight Center: Greenbelt, MD, USA, 1993; Volume 14.
8. Mueller, J.L.; Johnson, B.C.; Cromer, C.L.; Cooper, J.; Mclean, J.T.; Hooker, S.B.; Westphal, T.L. *The Second SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-2*; Hooker, S.B., Firestone, E.R., Eds.; NIST: Gaithersburg, MD, USA, 1994. Available online: https://www.nist.gov/publications/second-seawifs-intercalibration-round-robin-experiment-sirrex-2ed-sb-hooker-and-er (accessed on 15 August 2025).
9. Johnson, B.C.; Bruce, S.S.; Early, E.A.; Houston, J.M.; O'Brian, T.R.; Thompson, E.A.; Hooker, S.B.; Mueller, J.L. *The Fourth SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-4*; Hooker, S.B., Firestone, E.R., Eds.; NASA Tech Memo 104566 NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1996; Volume 37. Available online: https://www.nist.gov/publications/fourth-seawifs-intercalibration-round-robin-experiment-sirrex-4-ed-sb-hooker-and-er (accessed on 29 June 2020).
10. Johnson, B.C.; Yoon, H.W.; Bruce, S.S.; Shaw, P.-S.; Thompson, E.A.; Hooker, S.B.; Barnes, R.; Eplee, R.; Maritorena, S.; Mueller, J. The Fifth SeaWiFS Intercalibration Round-Robin Experiment (SIRREX-S), July 1996. 1999. Volume 7. Available online: https://www.nist.gov/publications/fifth-seawifs-intercalibration-round-robin-experiment-sirrex-s-july-1996 (accessed on 12 April 2020).
11. Hooker, S.B.; Firestone, E.R.; McLean, S.; Sherman, J.; Small, M.; Lazin, G.; Zibordi, G.; Brown, J.W.; McClain, C.R. *The Seventh SeaWiFS Intercalibration Round-Robin Experiment (SIRREX-7), TM-2003-206892*; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 2002; Volume 17. Available online: http://ntrs.nasa.gov/search.jsp?R=20020045342 (accessed on 8 February 2017).
12. Meister, G.; Abel, P.; McClain, C.; Barnes, R.; Fargion, G.; Cooper, J.; Davis, C.; Korwan, D.; Godin, M.; Maffione, R. The First Simbios Radiometric Intercomparison (Simric-1), April–September 2001. Vancouver. 2002. Available online: https://ntrs.nasa.gov/citations/20020045345 (accessed on 4 November 2025).
13. Meister, G.; Abel, P.; Carder, K.L.; Chapin, A.L.; Clark, D.K.; Cooper, J.W.; Davis, C.; English, D.; Fargion, G.S.; Feinholz, M.; et al. The second SIMBIOS Radiometric Intercomparison (SIMRIC-2), March–November 2002. *NASA Technol. Memo.* **2003**, *2*, 1–65.
14. Zibordi, G.; D'Alimonte, D.; Van der Linde, D.; Berthon, J.F.; Hooker, S.B.; Mueller, J.L.; Lazin, G.; McLean, S. The Eighth SeaWiFS Intercalibration Round-Robin Exercise (SIRREX-8), September–December 2001. *NASA Technol. Memo.* **2002**, *206892*, 39.
15. Zibordi, G.; Ruddick, K.; Ansko, I.; Moore, G.; Kratzer, S.; Icely, J.; Reinart, A. In situ determination of the remote sensing reflectance: An inter-comparison. *Ocean Sci.* **2012**, *8*, 567–586. [CrossRef]
16. Vabson, V.; Kuusk, J.; Ansko, I.; Vendt, R.; Alikas, K.; Ruddick, K.; Ansper, A.; Bresciani, M.; Burmester, H.; Costa, M.; et al. Laboratory Intercomparison of Radiometers Used for Satellite Validation in the 400–900 nm Range. *Remote Sens.* **2019**, *11*, 1101. [CrossRef]
17. Vabson, V.; Kuusk, J.; Ansko, I.; Vendt, R.; Alikas, K.; Ruddick, K.; Ansper, A.; Bresciani, M.; Burmester, H.; Costa, M.; et al. Field Intercomparison of Radiometers Used for Satellite Validation in the 400–900 nm Range. *Remote Sens.* **2019**, *11*, 1129. [CrossRef]
18. Białek, A.; Goodman, T.; Woolliams, E.; Brachmann, J.F.S.; Schwarzmaier, T.; Kuusk, J.; Ansko, I.; Vabson, V.; Lau, I.C.; MacLellan, C.; et al. Results from Verification of Reference Irradiance and Radiance Sources Laboratory Calibration Experiment Campaign. *Remote Sens.* **2020**, *12*, 2220. [CrossRef]
19. Tilstone, G.; Dall'olmo, G.; Hieronymi, M.; Ruddick, K.; Beck, M.; Ligi, M.; Costa, M.; D'alimonte, D.; Vellucci, V.; Vansteenwegen, D.; et al. Field Intercomparison of Radiometer Measurements for Ocean Colour Validation. *Remote Sens.* **2020**, *12*, 1587. [CrossRef]

20. Tilstone, G.H.; Jordan, T.M.; Aurin, D.; Białek, A.; Deru, A.; Ramsay, A.; Hieronymi, M.; Dall'oLmo, G.; Ligi, M.; Kovach, C.; et al. Radiometric field inter-comparison of fiducial reference measurements using an open source community processor. *Opt. Express* **2025**, *33*, 15756–15781. [CrossRef] [PubMed]
21. Mueller, J.L.; Johnson, B.C.; Hooker, S.B.; Mclean, J.T.; Biggar, S.F. *The Third SeaWiFS International Round-Robin Experiment [SIRREX-3]*; Hooker, S.R., Firestone, E.R., Acker, J.G., Eds.; NIST: Gaithersburg, MD, USA, 1996. Available online: https://www.nist.gov/publications/third-seawifs-international-round-robin-experiment-sirrex-3-ed-sr-hooker-er-firestone (accessed on 15 August 2025).
22. Hooker, S.; McClain, C.; Mannino, A. *T1-NASA Strategic Planning Document: A Comprehensive Plan for the Long-Term Calibration and Validation of Oceanic Biogeochemical Satellite Data*; Goddard Space Flight Center: Greenbelt, MD, USA, 2007.
23. IOCCG. Protocols for Satellite Ocean Colour Data Validation: In Situ Optical Radiometry. In *IOCCG Ocean Optics and Biogeochemistry Protocols for Satellite Ocean Colour Sensor Validation, Volume 3.0*; IOCCG Protocol Series; Zibordi, G., Voss, K., Johnson, B.C., Mueller, J.L., Eds.; International Ocean Colour Coordinating Group (IOCCG): Dartmouth, NS, Canada, 2019. [CrossRef]
24. IOCCG. *Request to Manufacturers of In Situ and Above-Water Spectral Imaging Radiometers in the UV, VIS and NIR Range. Addendum to IOCCG Protocol Series (2019)*; IOCCG: Dartmouth, NS, Canada, 2024; Volume 3. Available online: https://ioccg.org/what-we-do/ioccg-publications/ioccg-protocols/ (accessed on 23 September 2025).
25. Vabson, V.; Ansko, I.; Vendt, R.; Kuusk, J. *Technical Report D7: Complete Characterisation and Calibration Results for FRMOCnet OCR Models and Re-Characterisation Routine: An Update*; EUMETSAT: Darmstadt, Germany, 2022.
26. *ISO/IEC 17043:2023*; Conformity Assessment—General Requirements for the Competence of Proficiency Testing Providers. ISO: Geneva, Switzerland, 2023.
27. Mueller, J.L.; Morel, A.; Frouin, R.; Davis, C.; Arnone, R.; Carder, K.; Lee, Z.P.; Steward, R.G.; Hooker, S.; Mobley, C.D.; et al. *Ocean Optics Protocols For Satellite Ocean Color Sensor Validation, Revision 4. Volume III: Radiometric Measurements and Data Analysis Protocols*; Goddard Space Flight Space Centre: Greenbelt, MD, USA, 2003. [CrossRef]
28. Ylianttila, L.; Visuri, R.; Huurto, L.; Jokela, K. Evaluation of a Single-monochromator Diode Array Spectroradiometer for Sunbed UV-radiation Measurements. *Photochem. Photobiol.* **2005**, *81*, 333–341. [CrossRef] [PubMed]
29. Seckmeyer, G.; Bais, A.F.; Bernhard, G.; Blumthaler, M.; Drueke, S.; Kiedron, P.; Lantz, K.; McKenzie, R.L.; Riechelmann, S. *Instruments to Measure Solar Ultraviolet Radiation Part 4: Array Spectroradiometers (WMO/TD No. 1538)*; World Meteorological Organization: Geneva, Switzerland, 2010; p. 44.
30. Johnson, C.; Yoon, H.; Rice, J.P.; Parr, A.C. Chapter 1.2—Principles of Optical Radiometry and Measurement Uncertainty. In *Experimental Methods in the Physical Sciences*; Zibordi, G., Donlon, C.J., Parr, A.C., Eds.; Optical Radiometry for Ocean Climate Measurements; Academic Press: Cambridge, MA, USA, 2014; Volume 47, pp. 13–67. [CrossRef]
31. Salim, S.G.R.; Woolliams, E.R.; Fox, N.P. Calibration of a Photodiode Array Spectrometer Against the Copper Point. *Int. J. Thermophys.* **2014**, *35*, 504–515. [CrossRef]
32. Manninen, P.; Kärhä, P.; Ikonen, E. Determining the irradiance signal from an asymmetric source with directional detectors: Application to calibrations of radiometers with diffusers. *Appl. Opt.* **2008**, *47*, 4714–4722. [CrossRef]
33. Manninen, P.; Hovila, J.; Seppälä, L.; Kärhä, P.; Ylianttila, L.; Ikonen, E. Determination of distance offsets of diffusers for accurate radiometric measurements. *Metrologia* **2006**, *43*, S120. [CrossRef]
34. Schinke, C.; Pollex, H.; Hinken, D.; Wolf, M.; Bothe, K.; Kröger, I.; Nevas, S.; Winter, S. Calibrating spectrometers for measurements of the spectral irradiance caused by solar radiation. *Metrologia* **2020**, *57*, 065027. [CrossRef]
35. Vabson, V.; Ansko, I.; Vendt, R.; Kuusk, J. *Harmonised Cal/Char Lab Guidelines, Including Lab Protocols for FRMOCnet OCR Models*; Technical Report D12; EUMETSAT: Darmstadt, Germany, 2022.
36. Lin, J.; Dall'Olmo, G.; Tilstone, G.H.; Brewin, R.J.W.; Vabson, V.; Ansko, I.; Evers-King, H.; Casal, T.; Donlon, C. Derivation of uncertainty budgets for continuous above-water radiometric measurements along an Atlantic Meridional Transect. *Opt. Express* **2022**, *30*, 45648–45675. [CrossRef]
37. Talone, M.; Zibordi, G. Non-linear response of a class of hyper-spectral radiometers. *Metrologia* **2018**, *55*, 747–758. [CrossRef]
38. Talone, M.; Zibordi, G.; Bialek, A. Reduction of non-linearity effects for a class of hyper-spectral radiometers. *Metrologia* **2020**, *57*, 025008. [CrossRef]
39. Kostkowski, H. *Reliable Spectroradiometry*; Spectroradiometry Consulting: La Plata, MD, USA, 1997.
40. Zibordi, G.; Talone, M.; Jankowski, L. Response to Temperature of a Class of In Situ Hyperspectral Radiometers. *J. Atmospheric Ocean. Technol.* **2017**, *34*, 1795–1805. [CrossRef]
41. JCGM. JCGM 100, Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM), First Edition, September 2008. Available online: https://www.bipm.org/en/committees/jc/jcgm/publications (accessed on 30 September 2025).
42. Khlevnoy, B.B.; Solodilov, M.V.; Kolesnikova, S.S.; Otryaskin, D.A.; Pons, A.; Campos, J.; Shin, D.J.; Park, S.; Obein, G.; Valin, M.H.; et al. CIPM key comparison CCPR-K1.a.2017 for spectral irradiance 250 nm to 2500 nm. Final report. *Metrologia* **2023**, *60*, 02002. [CrossRef]

43. Metzdorf, J.; Sperling, A.; Winter, S.; Raatz, K.-H.; Möller, W. A new FEL-type quartz-halogen lamp as an improved standard of spectral irradiance. *Metrologia* **1998**, *35*, 423. [CrossRef]
44. Shaw, P.-S.; Woodward, J.T.; Johnson, B.C.; Brown, S.W.; Yoon, H.W. Incandescent lamp-based facility for measuring the long-term radiometric stability of spectrographs. *Appl. Opt.* **2021**, *60*, 10935–10944. [CrossRef] [PubMed]
45. Andor, G. New data-reduction method in detector-based spectral-irradiance measurements. *Metrologia* **1995**, *32*, 495. [CrossRef]
46. Andor, G. Approximation function of spectral irradiance of standard lamps. *Metrologia* **1998**, *35*, 427. [CrossRef]
47. Huang, L.K.; Cebula, R.P.; Hilsenrath, E. New procedure for interpolating NIST FEL lamp irradiances. *Metrologia* **1998**, *35*, 381. [CrossRef]
48. Kübarsepp, T.; Kärhä, P.; Manoocheri, F.; Nevas, S.; Ylianttila, L.; Ikonen, E. Spectral irradiance measurements of tungsten lamps with filter radiometers in the spectral range 290 nm to 900 nm. *Metrologia* **2000**, *37*, 305. [CrossRef]
49. Ojanen, M.; Kärhä, P.; Ikonen, E. Spectral irradiance model for tungsten halogen lamps in 340-850 nm wavelength range. *Appl. Opt.* **2010**, *49*, 880–886. [CrossRef]
50. Bernhard, G.; Seckmeyer, G. Uncertainty of measurements of spectral solar UV irradiance. *J. Geophys. Res. Atmos.* **1999**, *104*, 14321–14345. [CrossRef]
51. Nadal, M.E.; Barnes, P.Y. Near infrared 45 degrees/0 degrees reflectance factor of pressed polytetrafluoroethylene (PTFE) powder. *J. Res. Natl. Inst. Stand. Technol.* **1999**, *104*, 185. [CrossRef]
52. Pissulla, D.; Seckmeyer, G.; Cordero, R.R.; Blumthaler, M.; Schallhart, B.; Webb, A.; Kift, R.; Smedley, A.; Bais, A.F.; Kouremeti, N.; et al. Comparison of atmospheric spectral radiance measurements from five independently calibrated systems. *Photochem. Photobiol. Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol.* **2009**, *8*, 516–527. [CrossRef]
53. Gröbner, J.; Kouremeti, N. The Precision Solar Spectroradiometer (PSR) for direct solar irradiance measurements. *Sol. Energy* **2019**, *185*, 199–210. [CrossRef]
54. Rammeloo, C.; Baumgartner, A. Spectroradiometer Calibration for Radiance Transfer Measurements. *Sensors* **2023**, *23*, 2339. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.