2022 CEOS International Thermal Infrared Radiometer Comparison. Part I: Laboratory Comparison of Radiometers and Blackbodies

YOSHIRO YAMADA,[a] SUBRENA HARRIS,[a] MICHAEL HAYES,[a] ROB SIMPSON,[a] WERENFRID WIMMER,[b] RAYMOND HOLMES,[b] TIM NIGHTINGALE,[c] ARROW LEE,[c] NIS JEPSEN,[d] NICOLE MORGAN,[e] FRANK-M. GÖTTSCHE,[f] RAQUEL NICLÒS,[g] MARTÍN PERELLÓ,[g] CRAIG DONLON,[h] AND NIGEL FOX[a]

[a] *National Physical Laboratory, Teddington, United Kingdom*
[b] *University of Southampton, Southampton, United Kingdom*
[c] *Rutherford Appleton Laboratory, Science and Technology Facilities Council, Oxon, United Kingdom*
[d] *Danish Meteorological Institute, Copenhagen, Denmark*
[e] *CSIRO/Australian Bureau of Meteorology, Battery Point, Tasmania, Australia*
[f] *Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany*
[g] *University of Valencia, Valencia, Spain*
[h] *European Space Agency, Noordwijk, Netherlands*

ABSTRACT: An international comparison of field deployed radiometers for sea surface skin temperature (SST_{skin}) retrieval was conducted in June 2022. The campaign comprised a laboratory comparison and a field comparison. In the laboratory part, the radiometers were compared with reference standard blackbodies, while the same was done with the blackbodies used for the calibration of the radiometers against a transfer standard radiometer. Reference values were provided by the National Physical Laboratory (NPL), traceable to the primary standard on the International Temperature Scale of 1990. This was followed by the field comparison at a seaside pier on the south coast of England, where the radiometers were compared against each other while viewing the closely adjacent surface of the sea. This paper reports the results of the laboratory comparison of radiometers and blackbodies. For the blackbody comparison, the brightness temperature of the blackbody reported by the participants agreed with the reference value measured by the NPL transfer standard radiometer within the uncertainties for all temperatures and for all blackbodies. For the radiometer comparison, the temperature range of most interest from the SST_{skin} retrieval point of view is 10°–30°C, and in this temperature range, and up to the maximum comparison temperature of 50°C, all participants' reported results were in agreement with the reference. On the other hand, below 0°C the reported values showed divergence from the reference and the differences exceeded the uncertainties. The divergence shows there is room for improvement in uncertainty estimation at lower temperatures, although it will have limited implication in the SST_{skin} retrieval.

KEYWORDS: Ocean; Sea surface temperature; Infrared radiation; Instrumentation/sensors; Remote sensing; Measurements

1. Introduction

The temperature of Earth's surface is a fundamental and integral parameter within the larger system of the global climate. Patterns of sea surface temperature (SST) reveal the subsurface ocean variability, while long-term evolution of the global, regional and seasonal averages of SST are potential indicators of climate change (Minnett and Barton 2010). As such, SST is defined as one of the essential climate variables (Bojinski et al. 2014) that critically contributes to the characterization of Earth's climate by the World Meteorological Organization Global Climate Observing System (GCOS) (WMO 2022). Satellites have been monitoring global surface temperature for several decades, and have established sufficient consistency and accuracy between on-orbit sensors. This includes measurement of the SST, in which case the derived variable is the surface skin temperature (SST_{skin}) (Donlon et al. 2007). However, it is essential that such measurements are fully anchored to the International System of Units (SI) and that there is a direct regular correlation with "true" surface/in situ based measurements.

The most accurate of these surface-based measurements (used for validation) are derived from field-deployed infrared radiometers (or technically "radiation thermometers," although in this article the term "radiometer" will be used following the common usage of the terminology in this field). These are in principle calibrated traceable to SI, generally through a reference standard radiance blackbody (BB) source. Such radiometers are of varying design, operated by different teams in different parts of the globe. It is essential for the integrity of their use, both to provide validation data for satellites on orbit and to provide the links to future sensors, that any differences in the results obtained between them are understood. This knowledge will allow any potential biases to be removed and not to be transferred to satellite sensors. This knowledge can only be determined through formal comparison of the instrumentation, both in terms of its measurement capabilities in relation to primary "laboratory-based" calibration facilities, and its use in the field. The provision of a fully traceable link to SI as part of this process ensures that the data are

Corresponding author: Yoshiro Yamada, yoshiro.yamada@npl.co.uk

evidentially robust and can claim their status as a "climate data record." In Ohring et al. (2005), the target accuracy for satellite-derived SST is given as 0.1 K, and the shipborne infrared radiometers therefore aim to have similar accuracies, which is better than the measurement requirements for current satellite missions of <0.3 K (Donlon et al. 2012).

The calibration and validation community within the Committee on Earth Observation Satellites (CEOS) is well versed in the need and value of such comparisons, and has held highly successful exercises in Miami, Florida, in 2001 (Rice et al. 2004; Barton et al. 2004), and at the National Physical Laboratory (NPL), Teddington, United Kingdom, and in Miami in 2009 (Theocharous et al. 2010; Theocharous and Fox 2010) and at the NPL in 2016 (Theocharous et al. 2017; Barker-Snook et al. 2017a,b; Theocharous et al. 2019), all carried out under the auspices of CEOS. However, 6 years had passed since the last comparison, and it was considered timely to repeat/update the process, and so a similar comparison was conducted in 2022. The 2022 comparison included the following:

1) comparison of the BB reference standards used for calibrating the radiometers (laboratory based),
2) comparison of the radiometer response with a common SI-traceable BB target (laboratory based), and
3) evaluation of differences in radiometer response when viewing sea surface targets—in particular, the effects of external environmental conditions such as sky brightness (field based).

The comparison took place during two weeks in June of 2022. The first week involved the laboratory-based comparisons (comparisons 1 and 2) at NPL. The second week was devoted to the field-based comparison (comparison 3), at the tip of Boscombe Pier in Bournemouth, United Kingdom, and this part of the comparison is reported in an accompanying paper (Yamada et al. 2024).

This paper covers the result of the laboratory comparison of both the radiometers of the participants and of the BBs used to calibrate the radiometers. Detail of the comparison results can be found in two reports (Yamada et al. 2023a,b).

2. Overview of the comparison

As in the recent prior comparisons, NPL, the U.K. National Metrology Institute (NMI), served as the pilot for the 2022 comparison, by coordinating the comparison, preparing the protocol, providing the reference value traceable to the SI, analyzing the results, and preparing the reports. The protocol agreed by the participants can be seen in Yamada and Fox (2022). Seven participants including the pilot took part. This is a reduction from the previous 2016 comparison where 11 institutes, including the pilot, were present. No institute could participate from the United States and China, primarily due to travel restrictions imposed by the COVID-19 pandemic.

The laboratory comparison was undertaken in the week 13 17 June 2022. For the radiometer comparison, participants took turns to measure the reference standard BBs belonging to NPL. These BBs had calibrated platinum resistance thermometers monitoring their temperature that provided the reference value. For the BB comparison, a transfer standard radiometer was used to measure the brightness temperature of the participant BBs. The transfer radiometer was itself calibrated against the NPL reference standard radiometer traceable to the NPL primary temperature standards, and thus served to provide the reference value. Details on the reference standards are provided in the next section.

TABLE 1. Variable-temperature reference BB specifications.

	NH3-BB	SL-BB
Aperture diameter	ϕ 75-mm max	ϕ 160-mm max
Aperture distance from front panel	75 mm	35 mm
Emissivity	0.9993 @ 10 μm	0.9998 @ 11 μm[a]
Temperature range	From −40° to 50°C	From −10° to 40°C
Reference thermometer	PRT	PRT

[a] With ϕ 80-mm aperture applied to the cavity opening.

3. Reference standards

a. Reference standard BBs

Two variable temperature BBs were utilized in the radiometer comparison. One is an ammonia heat pipe BB (NH3-BB) (Chu and Machin 1999), and the other is a large aperture stirred liquid bath BB (SL-BB). The comparison reference values are given by the standard platinum resistance thermometers (SPRTs), which are calibrated traceable to the NPL primary temperature standards, measuring the temperature of the BBs. Of the two, the NH3-BB was the same as used in the previous comparison, and a diagram can be found in the paper by Theocharous et al. (2019). In the current comparison, the second BB source (SL-BB) was introduced, so that two measurements could be run side by side dividing the temperature range to be shared by the two BBs for improved efficiency.

The specifications for the two NPL variable temperature BBs are shown in Table 1. In recent years the uncertainty for the NH3-BB has been reevaluated and its day-to-day working uncertainty is slightly increased from what is shown in (Chu and Machin 1999), now being in the range from 0.13 to 0.10 K below 0°C and 0.095 K above 20°C (k 2). The SL-BB has a smaller uncertainty, which is around 0.05 K at 0° to 30°C (k 2), due to its higher emissivity.

Both BBs have purge systems utilizing a flow of dry nitrogen gas that is used below 10°C. They also have detachable, black-painted apertures that have been applied during the comparison measurements at set-point temperatures below the dewpoint to prevent ambient air from entering the cavity that can cause condensation of dew and frost. When the aperture is removed at around room temperature, the resulting decrease in cavity emissivity is almost completely balanced by the increase in the cavity reflectance of the ambient radiation, so the same correction and uncertainty due

TABLE 2. Reference and transfer standard radiometer specifications.

	AMBER (reference standard radiometer)	Heitronics TRT-IV.82 (transfer standard radiometer)
Wavelength	10.1 μm (9–11 μm)	8–14 μm
Target size	ϕ 5 mm	ϕ 8.7 mm
Measurement distance	70 mm	503 mm
Effective window/lens diameter	ϕ 13 mm	ϕ 57 mm
Scale realization	Through relative spectral response measurement, and BB measurement at the Ga melting point and at a second reference temperature at −30°C	By comparison with AMBER

to cavity emissivity have been applied for with and without the aperture.

b. *Reference and transfer standard radiometers*

The reference standard for the BB comparison was the NPL's reference standard radiometer Absolute Measurements of BB Emitted Radiance (AMBER) (Theocharous et al. 1998). In previous comparisons the AMBER was radiometrically calibrated by evaluating the radiance ratio against the gallium (Ga) melting point (29.7646°C) realized by an NPL reference fixed-point BB (Machin and Chu 1998). When the AMBER views an object the target temperature is derived from the measured radiance ratio. Here, radiance is evaluated as the spectral integration of the Planck's function multiplied by the preevaluated relative spectral responsivity function of the instrument. This is analogous to the definition of the ITS-90 above the silver point (Preston-Thomas 1990), although applied at a much lower temperature. The calibration had previously been verified through comparison with Physikalisch-Technische Bundesanstalt (PTB), Germany (Gutschwager et al. 2013). However, the scheme requires the knowledge of the zero-radiance signal that is derived through measurement of a zero-radiance source such as a cryogenic blackbody, which is hard to implement in practice. For the current comparison exercise, a new calibration scheme was applied employing a second reference temperature at around −30°C through measurement of the NH3-BB and extrapolating down to determine the zero-radiance signal, thus rendering the problematic realization of the zero-radiance source unnecessary. A detailed description of this two-point interpolation scale realization is described in a separate article (Yamada et al. 2023d).

A transfer standard radiometer was introduced for the first time in this comparison, which was the NPL TRT-IV.82 manufactured by Heitronics Infrarot Messtechnik GmbH (hereinafter referred to as "Heitronics"). This transfer standard was introduced following the positive contribution to the previous comparison by a radiometer of a similar model belonging to PTB (Theocharous et al. 2017). The Heitronics transfer standard radiometer was calibrated by comparison against the AMBER reference standard utilizing as the comparator sources the same NPL variable temperature NH3-BB and SL-BB described above. Then the Heitronics transfer standard was used to measure the temperature of the participant BBs.

The AMBER has a relatively small but not insignificant size-of-source effect (SSE). Therefore, a correction was made to account for the difference in the size of the two sources used for scale realization (30 mm diameter for the Ga-point BB and 75 mm diameter for the NH3-BB) and the AMBER SSE. For the Heitronics, a correction was made for the effect of the difference in the source size of the NH3-BB used to calibrate the Heitronics by comparison with the AMBER, and the participants' BB sizes. For this, a SSE correction scheme was applied that enables correction up to large source sizes at all measurement temperatures, based on a method described in Bloembergen (1999). The stability of the Heitronics was monitored by measurement of the Ga-point BB a few times a day before and during the comparison period. An abrupt shift of approximately 70 mK was detected after the calibration and just before the comparison, and a correction was applied to the measurements made of the participants' BBs to account for this. The uncertainty in this correction was also included in the uncertainty of the reference temperature.

The specifications of the AMBER and Heitronics relevant to the comparison measurements are given in Table 2.

4. Participants' instruments

a. *BBs*

There are, in general, two types of BBs used for calibration of radiometers for SST_{skin} retrieval. One is a BB cavity immersed in a stirred liquid bath, and the other is a BB cavity formed in a metal block. The BBs that participated in this comparison all belong to one of these two types. None of the BBs had a purge system to prevent formation of dew and frost. Therefore, their operation was limited to above the dewpoint of the laboratory, which was just below 10°C during the comparison.

1) Specialized BB with cavity in stirred liquid bath

BBs of the stirred liquid bath type that participated in the comparison are the CASOTS (Donlon et al. 1999) and CASOTS-II (Donlon et al. 2014) BBs. Both are similar in configuration and operation, the difference being in the improved thermal insulation leading to better temperature uniformity for the latter. The BB consists of cylindroconical cavity of copper with internal black coating of NEXTEL Suede Coating (NEXTEL Velvet Coating for CASOTS), leading to a high estimated

TABLE 3. Participants' BBs and radiometers. Acronyms are used in graphs. UoV-1 and 2 radiometers each have six bands, which have acronym extensions from B1 to B6, and each band is treated as an independent instrument in the comparison.

	BB		Radiometer	
Institute	Model	Acronym	Model	Acronym
University of Valencia	Landcal P80P	UoV	CIMEL Electronique CE312-2	
			Unit 1	UoV-1 B1–B6
			Unit 2	UoV-2 B1–B6
Karlsruhe Institute of Technology	Landcal P80P	KIT	Heitronics KT15.85 IIP	
			SN 9353; "surface" radiometer	KIT-1
			SN 13794; "sky" radiometer	KIT-2
CSIRO/Australian Bureau of Meteorology	CASOTS-II	CSIRO	ISAR5-E serial No. 16	CSIRO
University of Southampton	CASOTS-II	UoS	ISAR5-C serial No. 3	UoS
STFC Rutherford Appleton Laboratory	CASOTS	RAL	SISTeR	RAL
Danish Meteorological Institute	—	—	ISAR-5D	DMI

emissivity of 0.999 81 with a 50-mm-diameter aperture plate (CASOTS-II). The bath has no temperature control and the adjustment is made by adding or removing hot water, cold water, or ice. The temperature of the bath is monitored by a thermistor or a platinum resistance thermometer.

The Science and Technology Facilities Council (STFC) Rutherford Appleton Laboratory (RAL) brought with them their CASOTS, while University of Southampton (UoS) and CSIRO/Australian Bureau of Meteorology (CSIRO) took part in the comparison with their CASOTS-II.

2) COMMERCIAL BB WITH CAVITY IN METAL BLOCK

Two participants participated with the same commercial BB system (manufacturer: AMETEK-LAND; model: Landcal P80P). This system comprises a cylindroconical BB cavity with black, high temperature refractory coating in an aluminum block to achieve an emissivity higher than 0.995. The temperature of the block, heated and cooled by Peltier elements, can be monitored by a platinum resistance thermometer. University of Valencia (UoV) and Karlsruhe Institute of Technology (KIT) participated in the BB comparison with this type.

b. *Radiometers*

The radiometers that participated in this comparison can be categorized in to two types: dedicated systems for SST_{skin} retrieval equipped with internal BB references, and systems based on a commercially available instrument for general use without internal BB references. Other types of radiometer that were present in the previous comparison, such as Fourier transform infrared spectroradiometer type, were not among those that participated this time.

1) DEDICATED SYSTEMS FOR SST_{SKIN} RETRIEVAL WITH INTERNAL BBS

The SISTeR radiometer (Barton et al. 2004; Theocharous et al. 2019) of RAL and ISAR radiometer (Donlon et al. 2008) manufactured by UoS belong to the category of radiometer with an internal BB. Both have two reference BB cavities, one at ambient temperature, and the other, with a constant heater power supplied, at a slightly higher temperature (approximately 12 K higher for ISAR and 17 K higher for SISTeR). Both have a 45° scanning mirror that deflects the field of view of the radiometer to successively measure the radiation from the sea, the sky, and the two BBs. ISAR's detection is made by use of a radiometer (manufacturer: Heitronics; model: KT15.85) with detecting wavelength range from 9.6 to 11.5 μm. SISTeR utilized a pyroelectric detector in combination with a bandpass filter centered at 10.85 μm with full width at half maximum of 0.88 μm.

RAL participated with SISTeR, and UoS, CSIRO, and the Danish Meteorological Institute (DMI) participated with various models of ISAR. For all ISAR models, the optics and the detectors are of the same design.

2) COMMERCIAL INSTRUMENTS WITHOUT INTERNAL BB REFERENCE

Two institutes participated with radiometers of the category without internal BBs. KIT brought a set of two radiometers (KIT-1 and KIT-2, manufacturer: Heitronics; model: KT15.85 IIP) with wavelength range from 9.6 to 11.5 μm. One was intended for sea surface radiance measurement while the other was intended for sky radiance measurement. Similarly, UoV took part with a set of two radiometers (manufacturer: CIMEL Electronique; model: CE312-2), each with six selectable spectral bands (B1: 8.0 13.3 μm; B2: 10.9 11.7 μm; B3: 10.2 11.0 μm; B4: 9.0 9.3 μm; B5: 8.5 8.9 μm; B6: 8.3 8.6 μm) utilizing thermopile detectors. In this comparison each of these bands was treated as an independent participating instrument. A summary of the participants' instrumentation is given in Table 3.

A view of the laboratory where the BB comparison was carried out is shown in Fig. 1a. On the left, the NPL transfer radiometer (Heitronics) is shown viewing the reference standard compact Ga-point BB placed on the optical bench. On the left of the Ga-point BB a red CASOTS BB is seen, together with two blue CASOTS-II BBs to its left. At the far end of the row of BBs a Landcal P80P BB is seen. A second Landcal P80P was present but is not shown in the photograph. A view of the laboratory during the comparison of radiometers is shown in Fig. 1b. On the left, the CSIRO ISAR radiometer is measuring the SL-BB. On the right, the UoV CIMEL radiometer is being set up to measure the NH3-BB.

FIG. 1. View of laboratory during comparison measurements: (a) BB comparison, showing, from left to right, Landcal P80P, CASOTS-II, CASOTS-II, CASOTS, and Ga-point BB. Facing the BBs is the Heitronics transfer radiometer. (b) Radiometer comparison measurements showing, at left, ISAR measuring the SL-BB and, at right, CIMEL being prepared for NH3-BB measurement.

5. Measurement temperatures and measurand

In both the BB and the radiometer comparisons, the principal measurand was the brightness temperature of the BB sources at 10 μm. Therefore, where the temperature values are derived from contact thermometers monitoring the BB (as in the case of participant reported values in BB comparison, and pilot reference value in the radiometer comparison) corrections in the brightness temperature are required for source emissivity and ambient reflection. Where temperatures are measured by a 10-μm-range radiometer (as in the participant-reported value in the radiometer comparison, and the pilot reference value in the BB comparison) corrections are required for the SSE of the radiometer (when possible) but nothing else. Temperature here refers to that on the International Temperature Scale of 1990 (ITS-90) (BIPM 1989).

a. BB comparison

For the BB comparison the participants' BBs were compared at the nominal temperatures covering the range from 10° to 50°C as shown in Table 4 using the Heitronics transfer

TABLE 4. Measurement temperature points. Italics indicate temperature points at which a limited number of participants made measurements. The values in parentheses are outside the scope of comparison, as explained in more detail in the main text.

Comparison type	Nominal temperature (°C)
BB comparison	10, 15, 20, 25, 30, 35, *40*, *45*, *50*, (*55*, *60*)
Radiometer comparison	
NH3-BB source	−30, −15, 0, 30, 35, 40, *50*
SL-BB source	0, 10, 20, 30

standard radiometer. All BBs participated at all temperature points up to 35°C, above which only two participants (KIT, UoV) participated. Measurements at 55° and 60°C were also made by KIT, but these are not considered a part of the comparison since AMBER and Heitronics were not calibrated prior to the comparison up to these temperatures and measurement uncertainties with these instruments at these temperatures were not available.

b. Radiometer comparison

For the radiometer comparison the NPL BBs were set at the nominal temperatures covering the range from −30° to 50°C as shown in Table 4. The temperature range of main interest for SST_{skin} retrieval is 10° 30°C, so the SL-BB, having better temperature stability and higher emissivity as well as larger aperture for ease in alignment, was assigned to cover this range. The NH3-BB, being able to rapidly change set-point temperature and covering a wider range, was assigned to cover the higher and lower ends. At 0° and 30°C, both BBs were measured so that a check could be made of the agreement of the radiometer measurements made with the two BBs. All participants participated in all temperature points except at 50°C, where only UoV, KIT, and RAL participated. However, CSIRO later withdrew from submitting their results for three of the points (−15°, 35°, and 40°C) after noticing an issue with the alignment of their radiometer against the NH3-BB.

6. Measurement and reporting

a. BB comparison

The participants set the BBs to the set-point temperature and, when the temperature was sufficiently stable, the pilot took measurements of the BB brightness temperature with the transfer standard Heitronics radiometer. At least three repeated measurements were made with the Heitronics, including a realignment. The participants themselves took measurements of the BB temperature through contact thermometers. This was continuously logged for CASOTS and CASOTS-II BBs, and time-stamped data was reported. For the P80P BBs it was checked that the BB temperature stayed constant during the pilot's measurement.

Participants reported values of BB temperature after correcting for the BB emissivity and for the ambient reflection. The values measured with the transfer standard Heitronics by the pilot were corrected both for the difference from the reference AMBER scale, and for the Heitronics' SSE to account for the difference in the size of the source used to calibrate the transfer radiometer and the size of each of the participants' BBs. The SSE was measured with a flat plate radiator as a source in front of which apertures with varying diameters were placed, with sufficient space and tilt between them to prevent multiple reflections. The method for the SSE correction follows that presented by Bloembergen (1999), which allows correction to be applied to temperatures at which SSE is not directly evaluated. Heitronics temperature values after the corrections became the reference brightness temperatures of the BBs. The mean of the differences between the three reference brightness temperature values and the participant temperature values made at the same time was evaluated, and the reproducibility was assessed and included in the comparison uncertainty. When the participant reported a single value instead of temporal data the difference of this value from the simple mean of the reference measurements was used. Participants reported uncertainties of the measurement accompanying each measured value. The uncertainties included such sources as BB emissivity uncertainty, thermometer calibration uncertainty, cavity temperature nonuniformity, BB temperature stability, and reflected ambient radiation, as well as type A uncertainties. Details of the uncertainty estimations are found in Yamada et al. (2023a).

An example of measurement data from a participant BB with the transfer radiometer is shown in Fig. 2. Here, the UoS CASOTS-II BB is measured intermittently by the Heitronics while the water-bath temperature is continuously monitored by a thermistor. The logging of the data was done on two separate computer systems whose clocks were synchronized prior to data acquisition. CASOTS-II BB does not have temperature control capability, so the temperature is seen to gradually heat up due to the heat generated by the bath stirrer pump. The transfer standard Heitronics radiometer measurement is seen to follow this trend. Since the data acquisition synchronization (better than 30 s) is well within the time for significant temperature drift (at a rate of less than 0.01°C min^{-1}), the change in BB temperature has insignificant effect on the measurement.

b. Radiometer comparison

The participants took turns to align and measure the NPL BBs, with their radiometers, when the BB temperature was stable at the set point. The pilot measured the BB temperature continuously with SPRTs whose calibration was traceable to the NPL primary temperature standards, with corrections to the temperature applied for the cavity emissivity and ambient reflection to derive the reference brightness temperature value. The participants reported the time-stamped measured brightness temperature values and the associated uncertainties. The uncertainties included such sources as the primary calibration uncertainty of the radiometer, linearity, drift since calibration, and ambient temperature effect, as well as type A uncertainties. Details of the uncertainty estimations reported by each participant are found in Yamada et al. (2023b). The

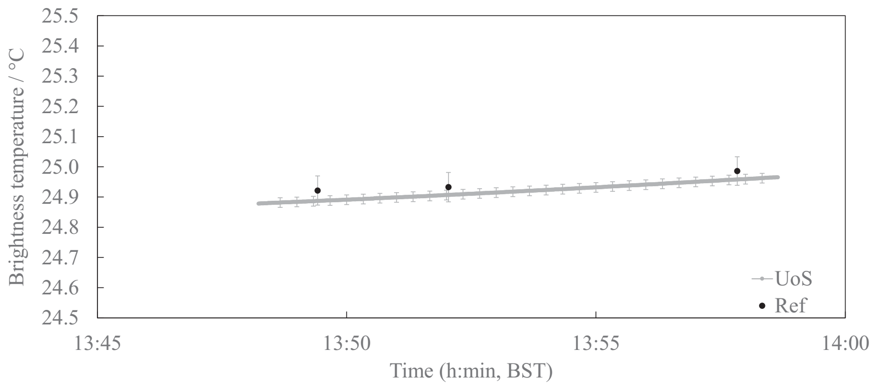

FIG. 2. An example of measurement data of participant's BB brightness temperature: UoS CASOTS-II BB at a nominal temperature of 25°C measured by a thermistor monitoring the water-bath temperature ("UoS") and by the transfer standard Heitronics radiometer ("Ref"). Error bars denote standard uncertainties.

measurements reported by the participants were compared with the reference value at the same point in time. The mean of the differences of the temperature values corresponding to the same timing was evaluated and the uncertainty of the mean was included in the comparison uncertainty.

An example of the measurement data of the reference SL-BB with participant radiometers is plotted in Fig. 3.

7. Comparison results

a. BB comparison result

Agreement with the reference value is evaluated by plotting the data with error bars added to both the participant reported values and the reference values, as shown in Figs. 4a and 4b. The error bars are the expanded uncertainties ($k = 2$). Plots are shifted slightly to make them distinguishable.

Since the CASOTS and CASOTS-II BBs have significantly smaller reported uncertainties than those for the other participant BBs, Fig. 4a shows results for these BBs, while Fig. 4b shows those for the others. Note the difference in the vertical scales.

b. Radiometer comparison result

Agreement with the reference value is evaluated by plotting the data with error bars added to both the participant reported values and the reference values, as shown in Figs. 5 and 6. Plots are shifted slightly to make them distinguishable. The error bars are the expanded uncertainties ($k = 2$). At each temperature point, each participant reported either a set of time-stamped measurements or a single averaged value. For the former, evaluation of the standard error of the mean of the temperature difference from the reference was evaluated for each set of measurements, and this was combined with the participant claimed combined measurement uncertainty.

Since the ISAR and SISTeR radiometers have an internal reference BB to improve the accuracy of measurement, the quoted uncertainties are significantly smaller than for the other two types of radiometer. Therefore, Figs. 5a and 6a, 6b

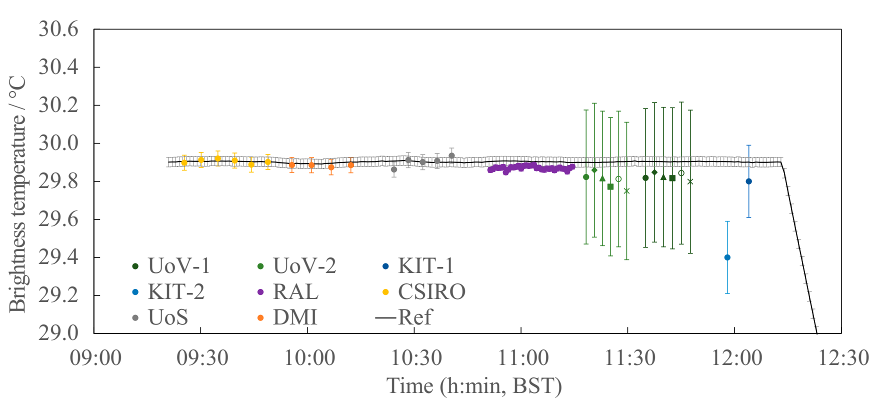

FIG. 3. An example of measurement data of the reference BB brightness temperature: NPL's SL-BB at a nominal temperature of 30°C is measured by reference SPRT monitoring the water-bath temperature ("Ref") and by participants' radiometers. Error bars denote standard uncertainties. For both UoV-1 and UoV-2 radiometers, each plot corresponds to a spectral band, from B1 to B6 in this order from left to right.

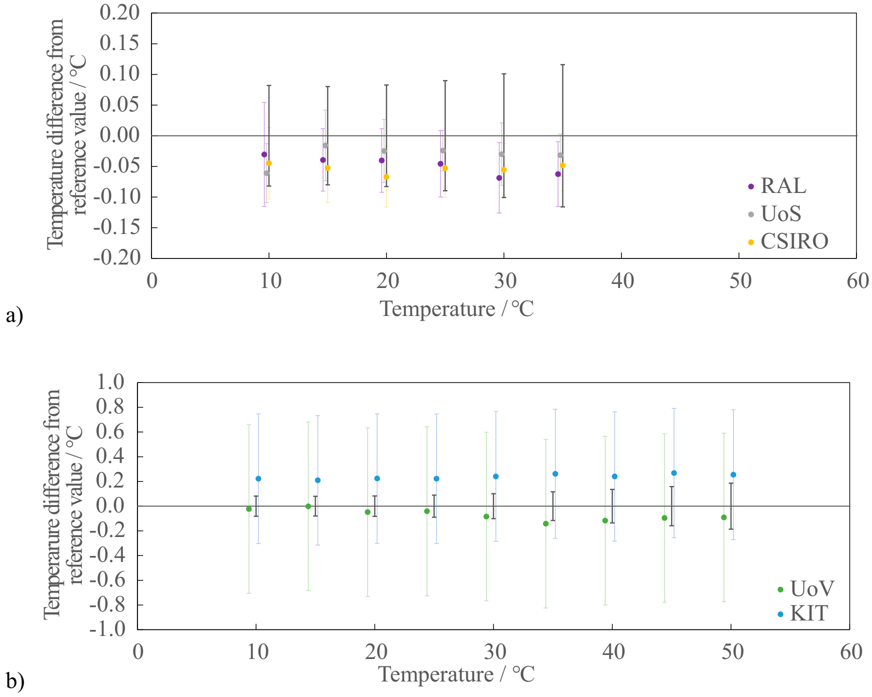

FIG. 4. BB comparison result: (a) specialized BBs (CASOTS and CASOTS-II), and (b) commercial BBs (Landcal P80P). Error bars denote expanded uncertainty ($k = 2$). The expanded uncertainties of the reference values are the black bars.

show results for the ISAR and SISTeR instruments, while Figs. 5b and 6c show those for the others. Note the difference in the vertical scales.

8. Discussions

a. BB comparison

Figure 2 shows an example of a measurement of the CASOTS-II BB, which is the type with the BB cavity immersed in a stirred bath. The figure shows that, although the BB temperature is slowly fluctuating, the Heitronics' reading follows the temporal fluctuation of the monitored BB temperature. The figure also verifies that the stability of the participant CASOTS-II BB is sufficient as long as the timings of the temperature readings are matched with the timings of the radiometric measurements, as is done in the current comparison. The other BB type, with the cavity in a temperature-controlled metal block (the Landcal P80P), is believed to have stable enough temperature control to assume its temperature to be constant, although this was not verified from the data since no temporal data were provided by the participants. In the case of the UoV BB, the stability was studied, before the comparison, using external PRT readings at fixed BB temperatures (10°, 20°, 40° and 50°C) during 90 min, obtaining a maximum standard deviation value of 0.03 K.

The standard measurement uncertainty with the Heitronics, including the scale realization on the AMBER, was approximately 45 mK at 20°C, which is comparable to or slightly smaller than the 53 mK reported for AMBER in the previous comparison (Theocharous et al. 2017). This is due mainly to the employment of the novel two-point interpolation scale realization on the AMBER, and the improved short-term stability and reproducibility achieved by using the Heitronics as the transfer standard for the comparison measurements. The short-term repeatability of the Heitronics was good and including this uncertainty term did not increase the calibration uncertainty.

Figure 4a shows that the deviations of the participant reported temperatures for the CASOTS and CASOTS-II BBs (belonging to RAL, UoS, and CSIRO) from the reference (brightness temperature) values are relatively small and are all less than 50 mK. The Landcal P80P BBs (of UoV and KIT) show larger deviations exceeding 0.1 K in some cases (Fig. 4b). No apparent dependence of the deviations on BB temperature is observed. The Landcal P80P BBs have an emissivity of 0.995 (cf. section 3a) and may be affected by reflection of objects that are at different temperatures from the ambient.

In the figures, the error bars, corresponding to the expanded participant and reference value uncertainties (k 2), overlap with each other, confirming the agreement with the reference value within the uncertainties for all BBs at all temperatures. The uncertainty of the reference is larger than the claimed uncertainties for the CASOTS and CASOTS-II BBs, but the former is sufficiently small to claim the comparison supports the reliability of the compared artifacts.

The temperature range of comparison for the CASOTS and CASOTS-II BBs was from 10° to 35°C, and good agreement with the reference value was confirmed in this range. This range is largely sufficient for the intended application,

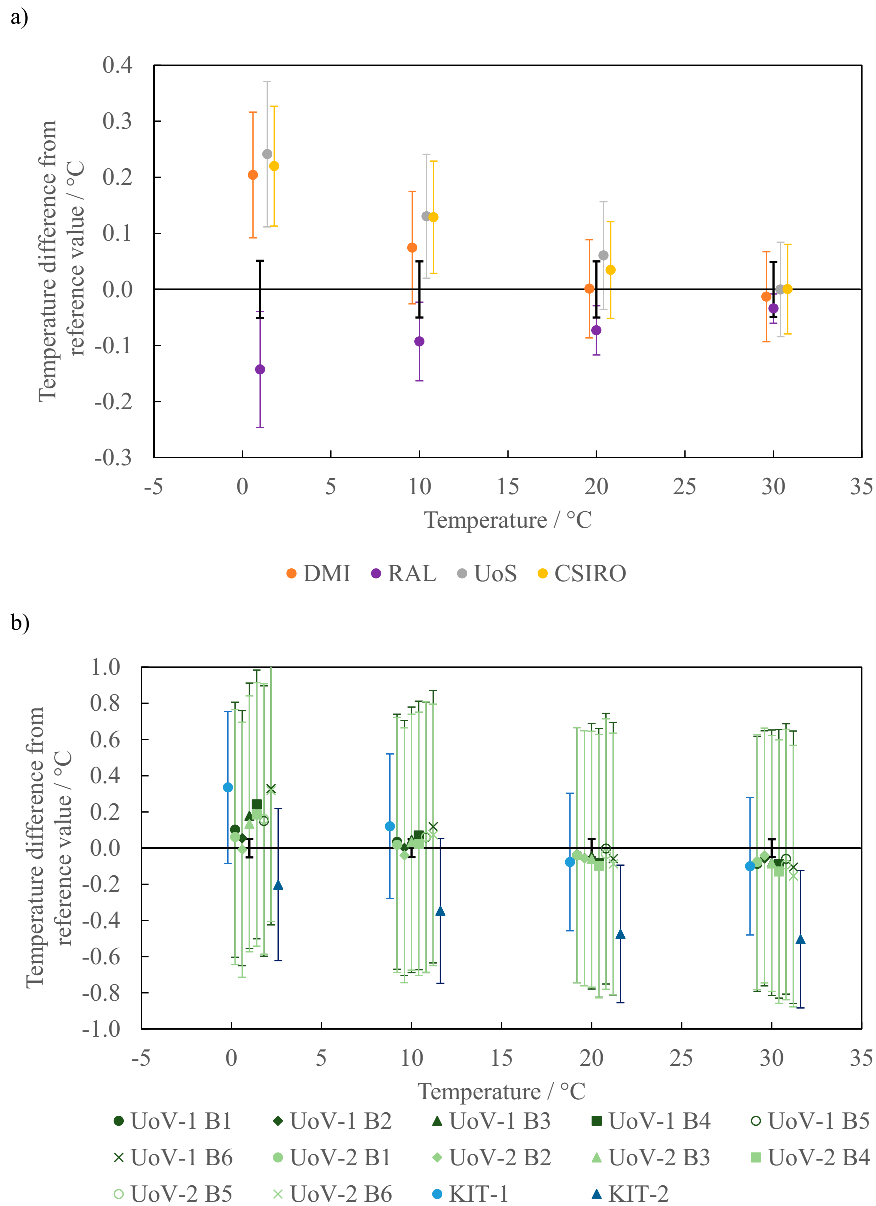

FIG. 5. Radiometer comparison result with SL-BB: (a) CSIRO, UoS, RAL, and DMI radiometers, and (b) UoV and KIT radiometers. Error bars are the expanded uncertainties ($k = 2$) for the participant measured values and for the reference value (with the latter shown as black bars). Plots are shifted slightly to make them distinguishable.

namely, SST_{skin} retrieval. If similar accuracy is to be required for ice or land surface temperature retrieval, the BB operation temperature range needs to be expanded. It should be noted that formation of dew and frost will not be an issue if the ambient temperature can also be lowered together with the set point so that it corresponds better to the actual condition in the field, for by doing so the dewpoint will also be lowered.

b. *Radiometer comparison*

Figure 3 shows that the stability of the reference SL-BB was sufficient to evaluate the agreement of the participants' temperature scales with the SI. For both this BB and the NH3-BB, the evaluated standard errors of the mean for each set of measurements were all small enough that including these only increased the combined uncertainty by less than 5%. Exceptions were some cases at −15° and −30°C, but, for these extreme cases, it could be confirmed from the scatter of the data that the poor repeatability was caused by the radiometer and not the reference BB. For the temperature range from 0° to 30°C, which is of most interest from the SST_{skin} retrieval objective, the SL-BB was used, and the introduction of this additional reference source for this

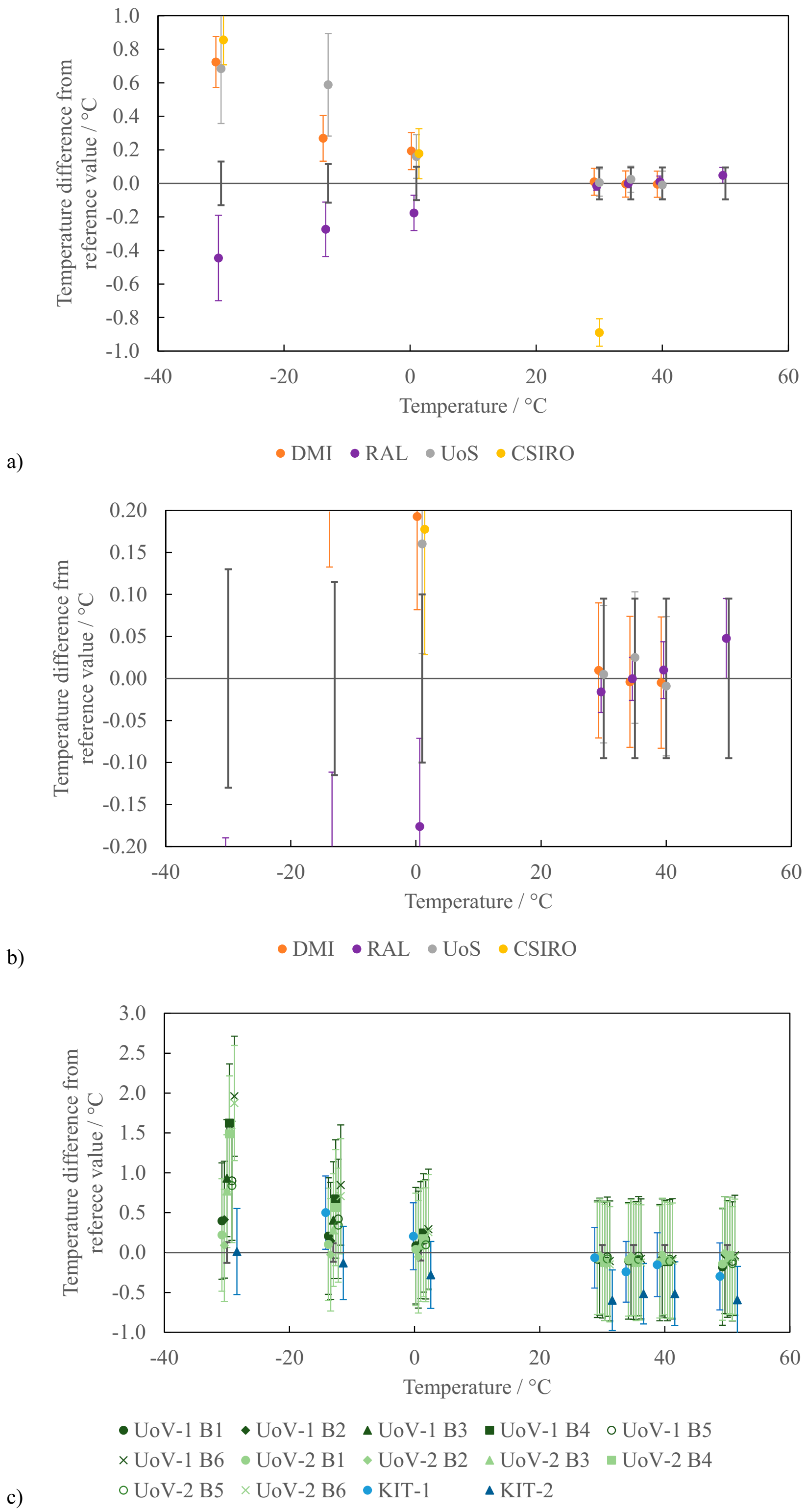

FIG. 6. Radiometer comparison result with NH3-BB: (a) CSIRO, UoS, RAL, and DMI radiometers; (b) CSIRO, pUoS, RAL, and DMI radiometers (magnified vertical scale); and (c) UoV and KIT radiometers. Error bars are the expanded uncertainties (k = 2) for the participant measured values and for the reference value (with the latter shown as black bars). Plots are shifted slightly to make them distinguishable.

comparison has made a positive impact through its exceptional stability.

In Figs. 5 and 6, the agreement of the participants' values with the reference value is evaluated. The expanded uncertainties (k 2) are expressed by error bars for both the participant measurements and for the reference. Overlap of the error bars for the measurement and the reference value, indicating the agreement of the two, is confirmed for all participants in the range 10° 30°C. The main source of the uncertainty for the UoV and KIT radiometers corresponds to the primary calibration uncertainty (from the Landcal P80P BB used), which was estimated as 0.34 K for UoV and 0.15 K for KIT (k 1; cf. Fig. 4b), and was mainly due to the BB cavity temperature nonuniformity effect. However, the good agreement observed in the comparison result indicates this is likely an overestimation. This is further confirmed in the BB comparison (Fig. 4). Investigation is envisaged to determine a more realistic reduced calibration uncertainty.

Separate graphs of the differences of the participant values from the reference value are given for the two sources, the SL-BB (Fig. 5) and the NH3-BB (Fig. 6). At 0° and 30°C both sources are measured by the radiometers, and it can be verified that the two sources are practically equivalent; that is, the differences (participant value − reference value) agree. The single outlier at these two temperatures is the measurement by CSIRO of the NH3-BB at 30°C, which shows an almost 1°C lower value than with the SL-BB. This is most likely caused by an issue with the alignment of the radiometer against the NH3-BB aperture, the wide field of view of the radiometer not being fully contained within the aperture that is located deep inside from the BB front face.

In Figs. 5 and 6, results for the ISAR and SISTeR radiometers are plotted separately from the other instruments with larger uncertainties. It is clear from the graphs that all three ISARs agree very well with each other while the SISTeR shows a different trend. A systematic error in the ISAR instrument may be present. An investigation into the cause is recommended for improved reliability.

In Figs. 5 and 6, it can also be seen that the scatter of the data increases as the temperature becomes lower, and also larger differences from the reference are observed. This is natural since the detected radiance signal of the radiometers becomes lower and the signal-to-noise ratio decreases, leading to more "noise" (scatter) in the results. Furthermore, all radiometers have some kind of an internal temperature reference kept at around ambient temperature and often a second reference slightly above this, and therefore have the highest accuracy around these temperatures. The farther away the target temperature becomes from ambient, the larger the extrapolation from the internal references is, and therefore the larger the uncertainty is. Last, the BBs used to calibrate the radiometers, a Landcal P80P or a CASOTS/CASOTS-II, are not equipped with purge systems to prevent formation of dew and frost in the BB cavity. This means that the use of the BBs is limited to above the dewpoint, which is normally above 0°C, or, if they are used below the dewpoint, they could be affected by dew and frost. The participant scales in the temperature range to below 0°C are therefore most likely realized by extrapolation, leading to increased uncertainty at these temperatures.

Even though a lower target temperature introduces various difficulties for accurate temperature measurement the declared uncertainties do not increase as expected, and for some participants they are almost the same as in the ambient temperature range. In the temperature range below 0°C, the error bars of the measurements do not necessarily overlap with that of the reference, indicating that the uncertainty estimation does not fully represent the true measurement capabilities of the participants. The result suggests that all participants need to reconsider the uncertainty budget so that such effects as extrapolation from the calibration temperature, low signal level due to reduced radiance, and larger deviation of the target temperature from the internal blackbody temperature are adequately taken into account in order that the uncertainties reflect the true measurement capabilities.

From the point of view of SST_{skin} retrieval, increasing the uncertainty is not an issue, since measurement at these low temperatures is required only for measurement of the sky brightness temperature and not for the sea surface brightness temperature. Sky brightness temperature is used in the correction for reflection at the sea surface when deriving the SST_{skin} from the sea surface brightness temperature. Since the emissivity of the sea surface is high, this correction is small especially when the sky has no overcast cloud and its brightness temperature is low. For instance, sky brightness temperature measurement error of 10 K at −30°C will only introduce an error of around 50 mK in the derived SST_{skin}. Thus the requirement for accuracy in the sky brightness temperature is much more relaxed.

9. Conclusions

Six SST_{skin} retrieval radiometers as well as five BBs used for calibrating them were gathered at NPL and their realized brightness temperatures were compared against the NPL reference standard scale as a part of the CEOS International Thermal Infrared Radiometer Intercomparison (CRIC). During the comparison, which took place during five days in June 2022, the BBs were measured with the transfer standard radiometer calibrated against the reference standard radiometer, AMBER, on which the scale was realized radiometrically traceable to the ITS-90 primary standards of NPL. The six radiometers viewed the cavities of an NH3-BB and an SL-BB, and brightness temperatures detected by the radiometers were compared against the values derived from the platinum resistance thermometers measuring the BBs, which were calibrated traceable to the ITS-90 primary standards of NPL.

The temperature range of the BB comparison covered from 10° to 35°C for all participants, and to 50°C for two of the participants. The brightness temperature reported by the participants agreed with the reference value measured by the NPL transfer standard radiometer within the uncertainties for all temperatures and for all BBs.

The SL-BB was applied for comparison in the range 0° 30°C. The temperature range of most interest from the SST_{skin} retrieval point of view is 10° 30°C, and in this temperature range all participants reported results that were in good agreement with the reference.

The NH3-BB was applied to the extreme temperatures at −30°, −15°, 0°, 30°, 35°, 40°, and 50°C. The temperatures above 30°C showed good agreement, similar to 30°C. On the other hand, at and below 0°C, the participant reported that values showed divergence from the reference that grew as the temperature became lower, and the divergence exceeded the uncertainties. This will not have a major significance in the derivation of the SST_{skin}, since this low temperature range is only required for sky brightness temperature measurement, is used for correction of the reflection at sea surface, and this requires lower accuracy. However, it indicates there is deficiency in the uncertainty estimation capability for all participants, especially when the derived SST_{skin} deviates from the ambient, and this should be improved in future if the participants are to maintain confidence in their SST_{skin} retrieval capabilities.

Three new features were introduced in the current comparison as compared with the previous comparison in 2016. The first is the introduction of the transfer standard radiometer to perform the measurement of the BBs. This overcomes the issue of the short-term stability of the AMBER, eliminates the thermal interaction of the cryogenically cooled AMBER with the BB, and reduces the problem with its poor operability encountered during practical measurements. The second is the employment of a novel scale realization on the AMBER utilizing two reference temperatures, which resulted in reduced uncertainty and made the realization of a zero-radiance source unnecessary. The third is the introduction of the new SL-BB. This proved to have a positive impact, not only for improved efficiency of the measurements, but also from the point of view of improvement in measurement accuracy by the participants owing to its large aperture, high emissivity, and temporal stability. Measurement of the two BBs made at same temperatures showed similar agreement, which confirmed that they produce identical comparison results.

Also note that the comparison in the laboratory is not always equivalent to measurement in the field. The comparison results show that divergence from the reference is noticeable where the target temperature diverts from the ambient temperature. The instruments tested here utilize internal reference BBs at temperatures at around the ambient, which means high accuracy is expected if the target is around the ambient. In the laboratory, this is not always the case, for the room temperature is maintained at ~23°C regardless of the BB source temperature. On the other hand, in the field the ambient temperature is nearly always close to the SST_{skin}. Performance of the instruments when deriving SST_{skin} should therefore be expected to be better in practice when compared with what this comparison shows, and the results shown in this comparison should be interpreted as a worst-case scenario. The result of the accompanying field comparison shows very good agreement among participants, and this seems to support the above observation (Yamada et al. 2023c, 2024).

In recent years, new improved radiometers for SST_{skin} retrieval have been developed, and more radiometers are being deployed at the sea. A future repeat of the current comparison exercise will be needed, possibly with a reduced interval between comparisons than the current six to eight years, when the new radiometers are being used in the field.

Acknowledgments. This work was funded by the ESA Contract FRM4SST Phase II. The UoV participants took part in the comparison with the support of the Research Projects PID2020-118797RBI00 (MCIN/AEI/10.13039/501100011033) and PROMETEO/2021/016 (Generalitat Valenciana).

Data availability statement. Datasets for this research are included in Yamada et al. (2023a; https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICR-NPL-001_ISSUE-1.pdf) and Yamada et al. (2023b; https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICR-NPL-002_ISSUE-1.pdf).

REFERENCES

Barker-Snook, I., E. Theocharous, and N. P. Fox, 2017a: 2016 comparison of IR brightness temperature measurements in support of satellite validation. Part 2: Laboratory comparison of radiation thermometers. NPL Tech. Rep. ENV 14, 127 pp., http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/FRM4STS_D100_TR-2_Part2_Radiometer_23Jun17-signed.pdf.

——, ——, and ——, 2017b: 2016 comparison of IR brightness temperature measurements in support of satellite validation. Part 3: Sea surface temperature comparison of radiation thermometers. NPL Tech. Rep. ENV 15, 58 pp., http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/FRM4STS_D100_TR-2_Part3_WST_23Jun17-signed.pdf.

Barton, I. J., P. J. Minnett, K. A. Maillet, C. J. Donlon, S. J. Hook, A. T. Jessup, and T. J. Nightingale, 2004: The Miami2001 infrared radiometer calibration and intercomparison: Part II: Shipboard results. *J. Atmos. Oceanic Technol.*, **21**, 268–283, https://doi.org/10.1175/1520-0426(2004)021<0268:TMIRCA>2.0.CO;2.

BIPM, 1989: The International Temperature Scale of 1990 (ITS-90). Bureau International des Poids et Mesures Tech. Doc., 20 pp., https://www.bipm.org/documents/20126/41791796/ITS-90.pdf.

Bloembergen, P., 1999: On the correction for the size-of-source effect corrupted by background radiation. *Proc. Seventh Int. Symp. on Temperature and Thermal Measurements in Industry and Science*, Delft, Netherlands, IMEKO/NMi Van Swinden Laboratorium, 607–612.

Bojinski, S., M. Verstraete, T. C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Amer. Meteor. Soc.*, **95**, 1431–1443, https://doi.org/10.1175/BAMS-D-13-00047.1.

Chu, B., and G. Machin, 1999: A low-temperature blackbody reference source to −40 °C. *Meas. Sci. Technol.*, **10**, 1–6, https://doi.org/10.1088/0957-0233/10/1/004.

Donlon, C. J., T. Nightingale, L. Fiedler, G. Fisher, D. Baldwin, and I. S. Robinson, 1999: The calibration and intercalibration of sea-going infrared radiometer systems using a low cost blackbody cavity. *J. Atmos. Oceanic Technol.*, **16**, 1183–1197, https://doi.org/10.1175/1520-0426(1999)016<1183:TCAIOS>2.0.CO;2.

——, and Coauthors, 2007: The Global Ocean Data Assimilation Experiment High-Resolution Sea Surface Temperature Pilot

Project. *Bull. Amer. Meteor. Soc.*, **88**, 1197–1214, https://doi.org/10.1175/BAMS-88-8-1197.

——, I. S. Robinson, W. Wimmer, G. Fisher, M. Reynolds, R. Edwards, and T. J. Nightingale, 2008: An Infrared Sea surface temperature Autonomous Radiometer (ISAR) for deployment aboard volunteer observing ships (VOS). *J. Atmos. Oceanic Technol.*, **25**, 93–113, https://doi.org/10.1175/2007JTECHO505.1.

——, and Coauthors, 2012: The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sens. Environ.*, **120**, 37–57, https://doi.org/10.1016/j.rse.2011.07.024.

——, W. Wimmer, I. Robinson, G. Fisher, M. Ferlet, T. J. Nightingale, and B. Bras, 2014: A second-generation blackbody system for the calibration and verification of seagoing infrared radiometers. *J. Atmos. Oceanic Technol.*, **31**, 1104–1127, https://doi.org/10.1175/JTECH-D-13-00151.1.

Gutschwager, B., E. Theocharous, C. Monte, A. Adibekyan, M. Reiniger, N. P. Fox, and J. Hollandt, 2013: Comparison of the radiation temperature scales of the PTB and the NPL in the temperature range from −57 °C to 50 °C. *Meas. Sci. Technol.*, **24**, 065002, https://doi.org/10.1088/0957-0233/24/6/065002.

Machin, G., and B. Chu, 1998: A transportable gallium melting point blackbody for radiation thermometry calibration. *Meas. Sci. Technol.*, **9**, 1653–1656, https://doi.org/10.1088/0957-0233/9/10/002.

Minnett, P. J., and I. J. Barton, 2010: Remote sensing of the Earth's surface temperature. *Radiometric Temperature Measurements: II. Applications*, Z. M. Zhang, B. K. Tsai, and G. Machin, Eds., Elsevier, 334–394.

Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, 2005: Satellite instrument calibration for measuring global climate change: Report of a workshop. *Bull. Amer. Meteor. Soc.*, **86**, 1303–1314, https://doi.org/10.1175/BAMS-86-9-1303,.

Preston-Thomas, H., 1990: The International Temperature Scale of 1990 (ITS-90). *Metrologia*, **27**, 3–10, https://doi.org/10.1088/0026-1394/27/1/002.

Rice, J. P., and Coauthors, 2004: The Miami2001 infrared radiometer calibration and intercomparison. Part I: Laboratory characterization of blackbody targets. *J. Atmos. Oceanic Technol.*, **21**, 258–267, https://doi.org/10.1175/1520-0426(2004)021<0258:TMIRCA>2.0.CO;2.

Theocharous, E., and N. P. Fox, 2010: CEOS comparison of IR brightness temperature measurements in support of satellite validation. Part II: Laboratory comparison of the brightness temperature of blackbodies. NPL Rep. OP 4, 49 pp., https://eprintspublications.npl.co.uk/4759/.

——, ——, V. I. Sapritsky, S. N. Mekhontsev, and S. P. Morozova, 1998: Absolute measurements of black-body emitted radiance. *Metrologia*, **35**, 549–554, https://doi.org/10.1088/0026-1394/35/4/58.

——, E. Usadi, and N. P. Fox, 2010: CEOS comparison of IR brightness temperature measurements in support of satellite validation. Part I: Laboratory and ocean surface temperature comparison of radiation thermometers. NPL Rep. OP3, 136 pp., http://eprintspublications.npl.co.uk/4744/.

——, I. Barker-Snook, and N. P. Fox, 2017: 2016 comparison of IR brightness temperature measurements in support of satellite validation. Part 1: Laboratory comparison of the brightness temperature of BBs. NPL Rep. ENV 12, 113 pp., http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/FRM4STS_D100_TR-2_Part1_Blackbody_23Jun17-signed.pdf.

——, and Coauthors, 2019: The 2016 CEOS Infrared Radiometer Comparison: Part II: Laboratory comparison of radiation thermometers. *J. Atmos. Oceanic Technol.*, **36**, 1079–1092, https://doi.org/10.1175/JTECH-D-18-0032.1.

WMO, 2022: The 2022 GCOS Implementation Plan. WMO Tech. Rep. GCOS-244, 98 pp., https://library.wmo.int/doc_num.php?explnum_id=11317.

Yamada, Y., and N. Fox, 2022: Fiducial reference measurements for sea surface temperature (FRM4SST): Protocol for FRM4SST CRIC laboratory comparison of radiometers and blackbodies. NPL Tech. Rep. FRM4SST: D-80-Ver-1-Final, 18 pp., https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICIP-NPL-001_Issue-1-signed.pdf.

——, and Coauthors, 2023a: Fiducial reference measurements for sea surface temperature (FRM4SST): D90—Results from CEOS International Thermal Infrared Radiometer Inter-Comparison (CRIC), Part 1 of 3: Laboratory comparison of blackbodies. NPL Tech. Rep. FRM4SST-CRICR-NPL-001_ISSUE-1, 40 pp., https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICR-NPL-001_ISSUE-1.pdf.

——, and Coauthors, 2023b: Fiducial reference measurements for sea surface temperature (FRM4SST): D90—Results from CEOS International Thermal Infrared Radiometer Inter-Comparison (CRIC), Part 2 of 3: Laboratory comparison of radiometers. NPL Tech. Rep. FRM4SST-CRICR-NPL-002_ISSUE-1, 52 pp., https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICR-NPL-002_ISSUE-1.pdf.

——, and Coauthors, 2023c: Fiducial reference measurements for sea surface temperature (FRM4SST): D90—Results from CEOS International Thermal Infrared Radiometer Inter-Comparison (CRIC), Part 3 of 3: Field comparison of radiometers. NPL Tech. Rep. FRM4SST-CRICR-NPL-003_ISSUE-1, 45 pp., https://ships4sst.org/sites/shipborne-radiometer/files/documents/FRM4SST-CRICR-NPL-003_ISSUE-1.pdf.

——, S. Harris, E. Theocharous, and M. Hayes, 2023d: A novel radiometric calibration of an infrared thermometer. *Metrologia*, **60**, 065005, https://doi.org/10.1088/1681-7575/acffcb.

——, and Coauthors, 2024: 2022 CEOS International Thermal Infrared Radiometer Comparison. Part II: Field comparison of radiometers. *J. Atmos. Oceanic Technol.*, **41**, 309–318, https://doi.org/10.1175/JTECH-D-23-0060.1.