PAPER • OPEN ACCESS

Survey of bidirectional transmittance distribution function measurement facilities by multilateral scale comparisons

To cite this article: Jinglin Fu *et al* 2024 *Metrologia* **61** 035006

View the article online for updates and enhancements.

Survey of bidirectional transmittance distribution function measurement facilities by multilateral scale comparisons

Jinglin Fu[1,*], Alejandro Ferrero[2], Tatjana Quast[1], Michael Esslinger[1],
Pablo Santafé-Gabarda[2], Néstor Tejedor[2], Joaquín Campos[2], Lou Gevaux[3],
Gaël Obein[3], Robin Aschan[4], Farshid Manoocheri[4], Erkki Ikonen[4,13],
Geiland Porrovecchio[5], Marek Šmíd[5], Ellie Molloy[6], Annette Koo[6], Søren A Jensen[7],
Rafael Oser[8], Jan Audenaert[9], Youri Meuret[9], Stefan Källberg[10], Iryna Gozhyk[11],
Tobias Kraus[12] and Alfred Schirmacher[1]

[1] Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany
[2] Consejo Superior de Investigaciones Científicas, Instituto de Óptica 'Daza de Valdés', Madrid, Spain
[3] LNE-CNAM (EA 2367), La Plaine St Denis, France
[4] Aalto University, Metrology Research Institute, Espoo, Finland
[5] Cesky Metrologicky Institut (CMI), Brno, Czech Republic
[6] Measurement Standards Laboratory of New Zealand, Lower Hutt, New Zealand
[7] Danish Fundamental Metrology A/S (DFM), Hørsholm, Denmark
[8] Covestro Deutschland AG, PCS-R&D-DR-Optics Lab, Leverkusen, Germany
[9] KU Leuven, Light & Lighting Laboratory, Gent, Belgium
[10] RISE Research Institutes of Sweden, Borås, Sweden
[11] Saint-Gobain Recherche, Labo mixte, Aubervilliers, France
[12] Temicon GmbH, Freiburg, Germany
[13] VTT MIKES, Espoo, Finland

E-mail: jinglin.fu@ptb.de

CrossMark

Abstract
In recent years, a growing demand for the capability of performing accurate measurements of the bidirectional transmittance distribution function (BTDF) has been observed in industry, research and development, and aerospace applications. However, there exists no calibration and measurement capabilities-entry for BTDF in the database of the *Bureau International des Poids et Mesures* and to date no BTDF comparison has been conducted between different national metrology institutes (NMIs) or designated institutes (DIs). As a first step to a possible future key comparison and to test the existing capabilities of determining this measurand, two interlaboratory comparisons were performed. In comparison one, five samples of three different types of optical transmissive diffusers were measured by five NMIs and one DI. By specific

---

* Author to whom any correspondence should be addressed.

sample choice, the focus for this study lay more on orientation-dependent scatter properties. In comparison two, where one NMI, one DI, one university, and three industrial partners investigated their measurement capabilities, the dependence on the orientation was not assessed, but two additional samples of the same material and different thickness were measured. Results of the two comparisons are presented, giving a good overview of existing experimental solutions, and showing specific sample-related problems to be solved for improved future BTDF measurements.

Supplementary material for this article is available online

Keywords: BTDF, intercomparison, diffuse transmission, goniospectrophotometry, applied radiometry, transmissive diffuser

1. Introduction

Fields of application for optical diffuse transmission are manifold and range from greenhouse ceilings, architectural accessory, to beam forming in lighting, and satellite radiation calibration, resulting in a demand for the traceable characterisation and calibration of optical transmission with high accuracy. This can be accomplished by the measurement of the bidirectional transmittance distribution function (BTDF), which would allow transmittance for any irradiation and detection conditions to be obtained. However, there exists neither a calibration and measurement capabilities-entry in the *Bureau International des Poids et Mesures* (BIPM) database for the BTDF measurand [1], nor has a key comparison arranged by the Consultative Committee of Photometry and Radiometry (CCPR) of the International Committee for Weights and Measures (CIPM). Only a bilateral comparison on the BTDF scale of two quasi-Lambertian diffusers performed by the National Aeronautics and Space Administration (NASA) and the National Institute of Standards and Technology (NIST) has shown good consistency between the two participating instruments [2]. To get an insight on existing European measurement capabilities and gain experience in the characterisation of a larger variety of different diffusers, intercomparisons on the BTDF scale between partners of the EMPIR project 18SIB03 (BxDiff) [3] were performed.

This study reports on two multilateral scale comparisons of BTDF measurements. Comparison one was performed between national metrology institutes (NMIs) and designated institutes (DIs) from Germany—*Physikalisch-Technische Bundesanstalt* (PTB), Denmark—*Danish Fundamental Metrology* (DFM), France—*Conservatoire National des Arts et Métiers* (CNAM), Finland—*Aalto-yliopisto* (Aalto), Czech Republic—*Cesky Metrologicky Institut* (CMI), and the non-European project partner *Measurement Standards Laboratory* (MSL) from New Zealand. It mainly aimed to lay the foundation for a future scale comparison between the national standard facilities for diffuse transmission. Unlike comparison one, comparison two also involved commercially available instruments in order to evaluate the present capabilities of these systems. The NMIs and DIs taking part in this comparison were

from Spain—*Consejo Superior de Investigaciones Científicas* (CSIC) and Sweden—*Research Institutes of Sweden* (RISE). The other participants were the *Katholieke Universiteit Leuven* (KU Leuven, Belgium) with their home-built instrument, and three industrial partners with commercial measuring systems: *Saint-Gobain Recherche* (Saint-Gobain), *Covestro Deutschland AG* (Covestro), and *Temicon GmbH* (Temicon). The latter two participants supplied different types of samples for both comparisons.

Measurements were performed for samples with various scattering characteristics in different in-plane geometries. The measured BTDF results were compared at two different wavelengths in comparison one and at one wavelength in comparison two. Since the two comparisons have different details and parameters, an overview of the structure of this paper is given in the following. Readers are invited to focus on those sections that are most relevant for their interests or application.

- Section 2 introduces the measurand, the sample details, the measurement procedures, the participated measuring instruments, and the methodology of both comparisons.
- Section 3 presents the results and discussions of comparison one, in which the metrological community dealt with accurate measurement of more difficult samples.
- Section 4 presents the results and discussions of comparison two, in which less complicated samples were studied by a broader group of participants including metrology institutes, university, and industries.
- Section 5 concludes both comparisons.

2. Methods

2.1. Description of the measurand

The measurand in both comparisons is the BTDF, which was firstly described by Bartell *et al* [4] as an expanded counterpart in the measurement of transmittance to the bidirectional reflectance distribution function (BRDF), which was introduced by Nicodemus [5, 6] as the measurand for reflection. The BTDF is defined as the derivative of the transmitted radiance L_t with respect to the uniform incident irradiance E_i:
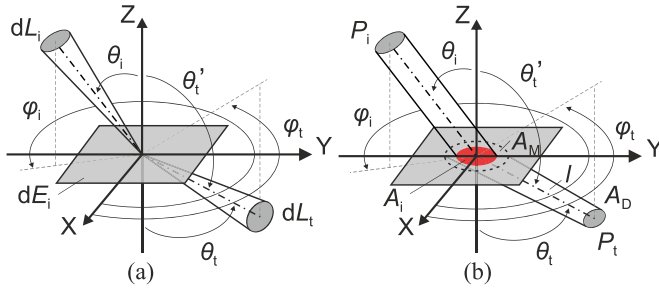
Figure 1. Geometrical variables of (a) BTDF definition and (b) BTDF measurements. Modified from figure 1 in [9] and figure 1 in Adapated from [10]. CC BY 4.0. Both licensed under a Creative Common Attribution (CC BY) licence. Meaning of the symbols in the figure is introduced in the text.

$$f_t(\theta_i, \varphi_i; \theta_t, \varphi_t) = \frac{dL_t(\theta_t, \varphi_t)}{dE_i(\theta_i, \varphi_i)}, \quad (1)$$

with the measurand BTDF referred to as f_t. It is dependent on the polar angle θ with respect to the sample normal (z-axis), and the azimuth angle φ with respect to the x-axis, both lying within the sample plane. The subscripts 'i' and 't' designate the direction of the incident and transmitted radiation, respectively. The measurand describes the scattering of a transmitting sample into the unit solid angle and has the unit of the inverse steradian (sr^{-1}). The BTDF also depends on the wavelength λ of the incident radiation and is typically expressed as the unpolarised value at a specific wavelength.

In the measurements, however, the infinitesimal values in the definition cannot be realised [7]. The measured BTDF is always an average over an area on the sample surface [4], and detected within a finite solid angle of the detector, as illustrated in figure 1.

The BTDF can be measured as an absolute value in two measurement schemes. In the under-irradiated scheme, where the measurement area A_M is larger than the irradiated area A_i, the BTDF is calculated as [2, 7, 8]:

$$f_t = \frac{P_t}{P_i} \cdot \frac{1}{\Omega_t \cos\theta_t}, \quad (2)$$

with P_i and P_t being the incident and transmitted optical power and Ω_t the detector solid angle determined by the detector aperture area A_D and the distance l between the detector aperture and the sample back surface:

$$\Omega_t = \frac{A_D}{l^2}, \quad (3)$$

where $\sqrt{A_D}$ should be much smaller than l. In the over-irradiated scheme (measurement area $A_M <$ irradiated area A_i), where the entire sample should be uniformly irradiated, the BTDF is measured as the ratio between the transmitted radiance L_t and the incident radiance L_i, along with the geometric factor [8]:

$$f_t = \frac{L_t}{L_i} \cdot \frac{1}{\Omega_i \cos\theta_i}. \quad (4)$$

For relative measurements, the measurand can be estimated by comparison with a reference standard for diffuse reflection, as no agreement on the reference standard for diffuse transmission has yet been reached. Thus, the BTDF of the sample under test is obtained with the aid of the calibrated BRDF of the reference standard f_{ref} and the ratio between the transmitted optical power of the sample under test P_t and the reflected optical power of the reference standard P_{ref} measured under the same condition [8]:

$$f_t = f_{ref} \cdot \frac{P_t}{P_{ref}}. \quad (5)$$

2.2. Description of measurement protocols

Altogether seven different sample types were chosen for the two comparisons. They give a good representation of diffusers being present on the market and used in different fields of application. These samples can be roughly divided into three categories as listed in table 1. Different sample types were used in the two comparisons so that varying key aspects were addressed. In this paper, all samples are mentioned by their designation as presented in table 1.

Each comparison comprised five different sample types, whose details are given in table 1. Unlike comparison one, comparison two lay the emphasis on different scattering magnitudes and thicknesses, rather than azimuthal orientation dependence. Both comparisons were performed in a star-type manner as schematically shown in figure 2, providing individual, nominally identical sample sets for each participant. Measurements were conducted by two pilot labs in each comparison. Each pilot measured two sample sets and distributed them to two non-pilot participants so that each one received one set. After their measurements, the pilots repeated the measurements on the two sets to ensure that no variation was introduced in the shipping process or during measurements. Additionally, one extra set was compared between the two pilots following a similar procedure, with repeated measurements carried out by the first pilot. Because of technical difficulties, deviations from the described procedure occurred in some cases, which are considered in the evaluation.

In all measurements, the sample under test was irradiated at the geometrical centre under normal incidence, i.e. (θ_i, φ_i) = (0°, 0°). Adapted to the scattering characteristics, the detection geometries vary for each sample type. For comparison one, the BTDF was measured in a narrower angular range than in comparison two. The samples B, C, and E were measured in the detection geometries (θ_t, φ_t) = (0°–35°, 0° & 180°) with angular step of $\Delta\theta_t = 5°$. For the other two orientation-dependent samples, the measurements were conducted in two perpendicular azimuthal directions. For sample A, the detection geometries were chosen as (θ_t, φ_t) = (0°–20°, 0° & 180°) and (0°–20°, 90° & 270°). For sample D, the parameters were (θ_t, φ_t) = (0°–40°, 0° & 180°) and (0°–20°, 90° & 270°). Both samples were measured with a smaller angular step of $\Delta\theta_t = 1°$. For comparison two, since all sample types involved are nominally independent of φ_t, the

Table 1. Details of sample types used in both comparisons[a].

Category	Comparison	Name	Designation	Material	Description of scattering properties	Dimensions (mm)
Quasi-Lambertian	1, 2	Heraeus HOD®-500	C	Fused synthetic SiO$_2$ with uniform air-bubbles (Ø < 20 μm)	Bulk scattering	Ø 50 × 2
	1, 2	SphereOptics Zenith® Polymer	E	Sintered polymer PTFE foil, sandwiched between two 90 mm × 110 mm plates	Bulk scattering	Ø 45 × 0.25
Dedicated angle-dependent scattering, no azimuthal dependence	1, 2	Thorlabs DG20-220	B	N-BK7 substrate, active surface roughened using 220 grit	Gaussian, FWHM 16°, surface scattering	Ø 50.8 × 2
	2	Covestro IOL1113-2, thin	Co1	Scatter centres in polycarbonate matrix	Gaussian, FWHM 50°, bulk scattering	50 × 50 × 1
	2	Covestro IOL1113-2, thick	Co3	Scatter centres in polycarbonate matrix	Gaussian, FWHM 130°, bulk scattering	50 × 50 × 3
Dedicated angle-dependent scattering, with pronounced azimuthal dependence	1	Thorlabs ED1-S20	A	Microlens units made of Zeonor© injection moulded on N-BK7 substrate	Square-shaped, FWHM ∼ 20°/20°, functional surface	Ø 25.4 × 1.5
	1	Temicon PAN001008	D	125 μm thick imprinted PET foil on standard glass substrate	Gaussian elliptical, FWHM 26°/13°, holographic diffuser	50 × 50 × 2

[a] The listed samples are examples of possibly suitable products. Their choice does not constitute an endorsement by the BxDiff consortium and similar commercial products may provide equivalent results.
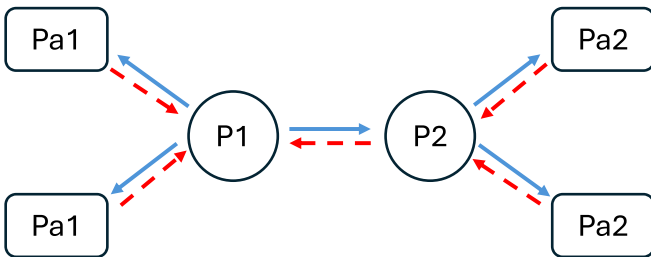
Figure 2. Schematic of the nominal star-type distribution of sample sets between pilots (P*x*) and non-pilot participants (Pa*x*) in each comparison. The solid blue arrow shows the distribution of one sample set after the first measurement, and the dashed red arrow represents the return of that sample set after the second measurement, followed by another repeated measurement by the original distributor. Deviations occurred in some cases due to technical difficulties.

Figure 3. Scattering patterns of sample types A (left) and D (right), with the dashed line indicating the $\varphi_t = 0°$ azimuth angle/detection scan plane.

detection geometries were chosen as $(\theta_t, \varphi_t) = (0°–80°, 0°)$ with angular step of $\Delta\theta_t = 5°$.

To minimise orientation-dependent errors, the mentioned azimuthal directions for samples A and D were carefully determined prior to the measurements. In figure 3, scattering patterns for these two sample types are shown. They were collected on a half-transparent diffuse screen behind the sample by a camera, which was aligned to the detector scan plane beforehand. From images consecutively acquired while rotating the sample around its normal, the optimal azimuthal orientation with respect to the laboratory coordinates was

determined and marked. For sample A, the zero azimuth was defined by aligning the upper and lower edges of the square-shaped pattern to the detector scan plane, as shown in the left image in figure 3, with the dashed line indicating the scan plane. For sample D, the zero azimuth was determined when the major axis of the elliptical scattering pattern was made parallel to the scan plane. To provide identical situations for the measurements to be carried out by different participants, the orientation-independent samples are also marked with a black dot on the back side of the sample for an arbitrarily chosen azimuth. In this way, the $\varphi_t = 0°$ azimuth is indicated by a (virtual) line connecting the sample geometrical centre and the marking dot. For square samples the $\varphi_t = 0°$ azimuth is referred to one specific sample edge.
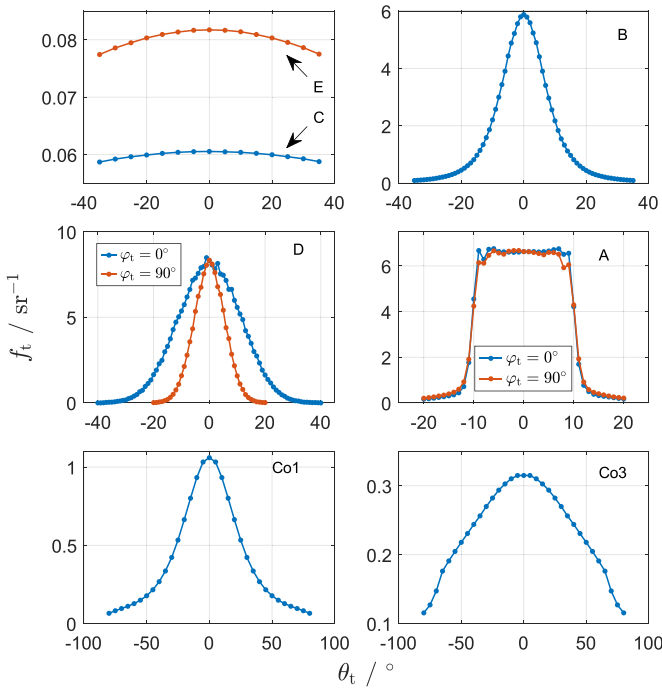
Figure 4. Angular transmission characteristics of all sample types used in the two comparisons. (Negative polar angles indicate azimuthal orientation $\varphi_t = 180°$.) The azimuthal dependence of samples A and D was analysed by the different angular transmission characteristics measured at two perpendicular azimuthal directions.

The influence of wavelength was dealt with as follows. In comparison one, BTDF values were determined for nominal wavelengths of 633 nm and 445 nm. In comparison two, the measurand was defined for the wavelength 633 nm. If the required wavelength was not accessible experimentally, a correction with respect to the sample spectral dependence was performed. This will be discussed in more detail in section 2.3 and in section S3 in the supplementary material (SM).

Figure 4 exemplifies the BTDF of all sample types. The two quasi-Lambertian samples C and E exhibit a slowly varying BTDF profile with a drop of approximately 10 % at $\theta_t = 35°$, but differ in amplitude and slightly in shape. The BTDF of sample B shows a much narrower angular distribution of approximately 16° FWHM with a peak at $\theta_t = 0°$ and a much higher maximum BTDF value. Sample D exhibits a Gaussian shaped BTDF profile with a width varying with the azimuth angle. The FWHM of the profile is approximately 26° at $\varphi_t = 0°$ and 13° at $\varphi_t = 90°$. Sample A shows distinctively almost constant BTDF values in the central angular range and a steep drop from maximum to nearly zero at around $\theta_t = 10°$. Samples Co1 and Co3, which only differ in their thickness (1 mm and 3 mm, respectively), present Gaussian shaped BTDF distributions with different widths of around 50° and 130° and maximum BTDF values around 1 sr^{-1} and 0.3 sr^{-1}.

The large variation in scattering characteristics indicates the need for applying different measurement conditions to properly measure the different types of samples. Especially for samples A, B, and D, a high detection angular resolution was recommended in the protocol to determine more accurately the rapidly changing BTDF values. For the Lambertian-like distributions of samples C and E, and for the wide Gaussian shaped distributions of samples Co1 and Co3, a lower detection angular resolution is suitable, which is also beneficial to measure the overall lower BTDF level as more scattered radiation is collected by a larger solid angle.

2.3. Description of measurement instruments

A brief description of the different measurement instruments used in both comparisons is given in the following section. More details of the participants' instrumentation are provided in section S1 in the SM.

The basic parameters of the instruments involved in comparison one are presented in table 2.

Measurements at PTB were performed on the modified NaNoRef setup, which is originally the national reference standard for specular reflection. Some detail about the original facility can be taken from [11, 12] and the modified setup used for the BTDF measurement is reported in detail in [9]. In brief, the samples mounted on several motorised translational and rotary stages were irradiated by collimated laser beams at wavelengths of 642 nm and 445 nm with a spatially uniform and speckle-reduced profile. The imaging detection system allowed different sizes of the measurement area A_M to be realised. All components of the detection system were mounted on another independent rotary arm, which rotates around the centre of the sample back surface at a distance of 494.5 mm to achieve different detection polar angles for the angle-resolved BTDF measurement.

The measuring system at DFM comprised a CW laser operating at 663 nm, which was weakly focused for the irradiation on the sample. The BTDF was measured at two orthogonal polarisations, and the average was reported. The detector with a detection aperture was mounted on a rotary arm, which rotates at a distance of 355 mm around the sample. For the nominal wavelengths prescribed in comparison one, PTB's and DFM's data were corrected using the angle-dependent spectral transmittance of all sample types as taken by extra measurements at PTB [9]. Details regarding the correction can be found in section S3 in the SM.

The facilities of CNAM, CMI, and MSL performed the measurement in a similar way. The monochromatic irradiation on all instruments was achieved by spectrally selecting broadband radiation with different spectral bandwidths. At CNAM a depolariser was used in irradiation, whereas CMI and MSL performed the measurement under polarised irradiation. The samples were held by a robot arm (CNAM and CMI) or mounted on motorised stages (MSL) and the detection system was mounted on a motorised rotation ring with different diameters on all three facilities. All detection systems consisted of an imaging mirror and diaphragms of different sizes so that the detection angular resolution could be varied. The size of the measurement area A_M on the sample back surface could be defined using the imaging mirror. The distance from the sample to the detection aperture varied for these facilities. CNAM has the largest distance among all participants,

Table 2. Basic parameters of the measurement instruments and assignment of the participants' roles in comparison one.

| Participant, role[a] | Irradiation source | | | Irradiated area[b] | Aperture angle (full) | | Measurement area ($\theta_t = 0°$)[b] | Distance sample to detector | Detection system | Measurement method | Measurement scheme |
	Type & Wavelength	Bandwidth	Polarisation		Irradiation	Detection[b]					
PTB P1	642 nm diode laser & 445 nm solid-state laser	0.5 nm (642 nm)/ 1 nm (445 nm)	s & p	9.6 mm²/ 29.2 mm²/ 44.2 mm²/ 80.1 mm²	0.1°	0.5°/1°/1.5°	50.3 mm²/75.4 mm²/ 265.9 mm²	494.5 mm	Si-photodiode on integrating sphere	Absolute	Under-irradiated
DFM P2	663 nm CW laser	1.2 nm	s & p	27 mm²	N/A	0.8°	27 mm²	355 mm	Si-photodiode	Absolute	Under-irradiated
CMI Pa1	Monochromatic 633 nm & 445 nm, from broadband source	5 nm	s & p	78.5 mm²	0.66°	2.1°	N/A	462.2 mm	Si-photodiode on integrating sphere	Absolute	Under-irradiated
MSL Pa1	Monochromatic 633 nm & 445 nm, from broadband source	1.5 nm	s & p	12.6 mm²/ 50.3 mm²	N/A	0.9°/1.9°	19.6 mm²/ 50.3 mm²/63.6 mm²	500 mm	Broadband Si-photodiode	Absolute	Under-irradiated
CNAM Pa2	Monochromatic 633 nm & 445 nm, from broadband source	2 nm[c]/5 nm	Unpolarised	28.3 mm²/ 78.5 mm²	0.2°–0.35°	0.8°/1°/2°	28.3 mm²/78.5 mm²	1150 mm	Photodiode	Absolute	Under-irradiated
Aalto Pa2	Monochromatic 633 nm & 445 nm, from broadband source	5 nm/4 nm[d]	s & p	50.3 mm²/ 113.1 mm²/ 50 mm²[d]	N/A	3°/8°[d]	N/A	472.9 mm/ 130 mm[d]	Si-photodiode	Absolute/ relative[d]	Under-irradiated

[a] Px = pilot *x*, Pa*x* = participant in the sub-group of pilot *x*.
[b] For some participants, this parameter varies depending on sample scattering characteristics.
[c] Only used in the measurement of sample type A at 445 nm.
[d] Only used in the measurement of sample type E.

while CMI and MSL applied distances similar to PTB. More details about the involved facilities are presented in [13, 14] for CNAM and MSL, respectively.

At Aalto, the measurements of samples A, B, C, and D were performed on the absolute goniospectrophotometer. The samples were held horizontally, and the irradiation and detection systems were mounted on two independent motorised rotary stages. The samples were irradiated with a monochromatic radiation which was spectrally selected from a broadband light source, and the optical signal was detected by a Si-photodiode. Details of this setup are described in [15]. Sample E was measured on the transfer standard goniospectrophotometer applying a much larger solid angle. This measurement used the relative method by measuring against a standard calibrated on Aalto's absolute setup. Therefore, the uncertainties of sample E turned out to be larger than the uncertainties of other samples.

The basic parameters of the instruments involved in comparison two are presented in table 3.

At CSIC, BTDF was measured using the goniospectrophotometer described in [16] and [17]. It was originally conceived for BRDF measurements and was adapted for BTDF measurements. The samples were irradiated using broadband irradiation with high directionality and uniformity achieved by using a Köhler optical system. A spectroradiometer was used for detection, providing spectral information from 380 nm to 780 nm. A relative measurement was performed in the over-irradiated scheme, using a diffuse reflectance standard as reference, whose BRDF had been previously calibrated in the standard geometry $0°{:}45°$.

The measuring system at Covestro consisted of a halogen lamp, providing a band-pass filter for spectral evaluation at 632 nm. The measurement was performed in the relative mode referring to a PTFE reflection standard in the under-irradiated scheme.

At Saint-Gobain, BTDF was measured with an OMS4 goniospectrophotometer commercialised by OPTIS. The OMS4 setup and its usage for BTDF measurements is described in [18]. For this work, a broadband source (Xe-lamp) filtered at 635 nm was used for sample illumination. The signal was detected using a photomultiplier tube.

A commercial instrument was used by Temicon to perform BTDF measurements with high resolution stepped detector and collimated illumination.

A spectrally broad illumination (Xe-lamp), modified by a bandpass filter (VIS region) and collimating optics was used at KU Leuven. The detection was a CCD-based spectrograph with variable integration time, in combination with ND filters in the optical path results in a high dynamic range of 6 decades. The measurement was relative using a reference tile (primed Barium Sulphate reflection standard—PTB), and was carried out under-irradiated. This instrument is described in detail in [19].

The BTDF measurement setup at RISE used a laser-driven light source with collimating optics, creating an irradiation spot of 10 mm in diameter. The primary collector was a 4 inch integrating sphere with a silicon detector, positioned at a distance of 500 mm from the sample. The spectral BTDF was determined using an interference filter. The measurement was performed in the under-irradiated scheme.

2.4. Methodology of the comparisons

The evaluation of both comparisons follows the instructions in [20] (procedure A) together with the Guidelines for CCPR Key Comparison Report Preparation using the cut-off-assisted Mandel-Paule method [21, 22]. The BTDF is assumed to be independently measured by all participants and is compared in consideration of inter-sample-set variations.

Starting from the BTDF values reported by each non-pilot participant $f_{k,\,\mathrm{par}}$ (k is the sample set number) with their associated standard uncertainties $u(f_{k,\,\mathrm{par}})$ and the BTDF values measured by pilot one, $f_{k,\,\mathrm{p}_1}$, and pilot two, $f_{k,\,\mathrm{p}_2}$, with their standard uncertainties $u(f_{k,\,\mathrm{p}_1})$ and $u(f_{k,\,\mathrm{p}_2})$, one specific sample set is chosen to be the reference set, to which the data from all participants are normalised in the way as shown in equations (6) and (7), assuming that the reference set was measured by pilot 1 ($f_{\mathrm{ref},\,\mathrm{p}_1}$). Note that the slash '/' in the subscript of some variables in the following discussion indicates that the same formula can be applied to either the variable with the subscript before '/' or the variable with the subscript after '/'.

$$f_{k,\,\mathrm{par/p}_1}^{\mathrm{norm}} = f_{k,\,\mathrm{par/p}_1} \cdot \frac{f_{\mathrm{ref},\,\mathrm{p}_1}}{f_{k,\,\mathrm{p}_1}}, \text{ sample set } k \text{ measured by pilot 1}$$

$$f_{k,\,\mathrm{par/p}_2}^{\mathrm{norm}} = f_{k,\,\mathrm{par/p}_2} \cdot \frac{f_{\mathrm{ref},\,\mathrm{p}_1}}{f_{k,\,\mathrm{p}_2}} \cdot \frac{f_{k^*,\,\mathrm{p}_2}}{f_{k^*,\,\mathrm{p}_1}}. \text{ sample set } k \text{ measured by pilot 2}$$

$$(6)$$

$f_{k^*,\,\mathrm{p}_{1/2}}$ is the value of the additional pilot-comparison set k^* measured by pilot 1/2, which helps connect the measurement scales between their sub-groups. Identical normalisation is also performed for the uncertainties:

$$u\left(f_{k,\,\mathrm{par/p}_1}^{\mathrm{norm}}\right) = u(f_{k,\,\mathrm{par/p}_1}) \cdot \frac{f_{\mathrm{ref},\,\mathrm{p}_1}}{f_{k,\,\mathrm{p}_1}}, \quad \text{sample set } k \text{ measured by pilot 1}$$

$$u\left(f_{k,\,\mathrm{par/p}_2}^{\mathrm{norm}}\right) = u(f_{k,\,\mathrm{par/p}_2}) \cdot \frac{f_{\mathrm{ref},\,\mathrm{p}_1}}{f_{k,\,\mathrm{p}_2}} \cdot \frac{f_{k^*,\,\mathrm{p}_2}}{f_{k^*,\,\mathrm{p}_1}}. \quad \begin{array}{l}\text{sample set } k \\ \text{measured by pilot 2}\end{array}$$

$$(7)$$

In comparison one, the reference set is selected as the pilot-comparison set, i.e. $f_{k^*,\,\mathrm{p}_{1/2}} = f_{\mathrm{ref},\,\mathrm{p}_{1/2}}$. In this way, the above equations for $f_{k,\,\mathrm{par/p}_2}^{\mathrm{norm}}$ and $u(f_{k,\,\mathrm{par/p}_2}^{\mathrm{norm}})$ are reduced to the first case in equations (6) and (7).

After being normalised, the data of multiple sample sets for each pilot are further reduced to one data set by means of the arithmetic mean. In the end, only one data set ($f_n^{\mathrm{norm}}, u(f_n^{\mathrm{norm}})$) per participant n contributes to the determination of the reference value R of the comparison. A provisional R for every sample type in every geometry is calculated as the weighted mean of all participants' data:

$$R = \frac{\sum_{n=1}^{N} w_n f_n^{\mathrm{norm}}}{\sum_{n=1}^{N} w_n}, \tag{8}$$

Table 3. Basic parameters of the measurement instruments and assignment of the participants' roles in comparison two.

| Participant, role [a] | Irradiation source | | | | Aperture angle (full) | | Measurement area ($\theta_t = 0°$) | Distance sample to detector | Detection system | Measurement method | Measurement scheme |
	Type & Wavelength	Bandwidth	Polarisation	Irradiated area	Irradiation	Detection					
CSIC P1	Xe-lamp, broadband irradiation	N/A	Unpolarised	707 mm²	1.5°	5°	38.5 mm²	350 mm	Spectro-radiometer	Relative	Over-irradiated
Covestro P2	632 nm Halogen light source (100 W) with band filter	5 nm	Unpolarised	12.5 mm²	N/A	2.2°	50.2 mm²	170 mm	Integrated flux detection	Relative	Under-irradiated
Saint-Gobain Pa1	635 nm, Xe-lamp	10 nm	Unpolarised	50 mm²	0.2°	0.2°	50 mm²	1000 mm	Photomultiplier	Relative	Under-irradiated
Temicon Pa1	633 nm diode laser	N/A	s & p	50.3 mm²	≪1°	<0.05°	N/A	750 mm	Photodiode with AD converter	Relative	Under-irradiated
KU Leuven Pa2	Xe-lamp, broadband irradiation	3 nm	Unpolarised	132.7 mm²	0.1°	1.6°	706.9 mm²	886 mm	Spectrograph + CCD	Relative	Under-irradiated
RISE Pa2	Broadband irradiation source with bandpass filter	10 nm	Unpolarised	78.5 mm²	N/A	2.9°	176.7 mm²	500 mm	Si-photodiode on integrating sphere	Relative	Under-irradiated

[a] Px = pilot x, Pax = participant in the sub-group of pilot x.

8

with the weights:

$$w_n = u^{-2}(f_n^{\mathrm{norm}}), \tag{9}$$

and the standard uncertainty of the reference value R:

$$u(R) = \sqrt{\frac{1}{\sum_{n=1}^N w_n}}. \tag{10}$$

By applying a chi-squared (χ^2) test, the overall consistency of the results can be checked:

$$\chi^2_{\mathrm{obs}} = \sum_{n=1}^N \frac{(f_n^{\mathrm{norm}} - R)^2}{u^2(f_n^{\mathrm{norm}})}. \tag{11}$$

If more than 95 % of the calculated χ^2_{obs} are smaller than the χ^2 distribution with $N-1$ degrees of freedom, i.e. 11.07 for $\chi^2_{0.05}(5)$ for most of the cases in both comparisons and 9.488 for $\chi^2_{0.05}(4)$ for one exceptional case in comparison one, the measured data can be regarded as consistent and the degree of equivalence (DoE) for participant n is expressed as (d_n, $U(d_n)$):

$$\begin{aligned} d_n &= f_n^{\mathrm{norm}} - R, \\ U(d_n) &= 2\sqrt{u^2(f_n^{\mathrm{norm}}) - u^2(R)}, \end{aligned} \tag{12}$$

which indicates the consistency of the data provided by each participating lab to the reference value R depending on the relation between $|d_n|$ and $U(d_n)$. Consistency is achieved when 95 % of the data fulfil $|d_n| < U(d_n)$. If more than 5 % of the data show $|d_n| > U(d_n)$ or $|d_n| \ll U(d_n)$, this indicates either inconsistent measurement data, or overestimated uncertainties.

If the χ^2 test fails, the participants' uncertainties $u(f_n^{\mathrm{norm}})$ are adjusted by introducing a cut-off value $u_{\mathrm{cut-off}}$:

$$\begin{aligned} u_{\mathrm{cut}}(f_n^{\mathrm{norm}}) &= u(f_n^{\mathrm{norm}}), \quad \mathrm{for}\, u(f_n^{\mathrm{norm}}) \geqslant u_{\mathrm{cut-off}} \\ u_{\mathrm{cut}}(f_n^{\mathrm{norm}}) &= u_{\mathrm{cut-off}}, \quad \mathrm{for}\, u(f_n^{\mathrm{norm}}) < u_{\mathrm{cut-off}} \end{aligned} \tag{13}$$

which is calculated as the following:

$$\begin{aligned} u_{\mathrm{cut-off}} &= \mathrm{mean}\{u(f_n^{\mathrm{norm}})\}, \\ &\quad \mathrm{for\ all}\, u(f_n^{\mathrm{norm}}) \leqslant \mathrm{median}\{u(f_n^{\mathrm{norm}})\}. \end{aligned} \tag{14}$$

Then, a new reference value is calculated by repeating the process from equations (8) to (11) to check whether the global consistency is achieved with slightly larger $u_{\mathrm{cut}}(f_n^{\mathrm{norm}})$ for some participants. If the consistency condition is still not fulfilled, the uncertainties are further adjusted by taking into account an additional interlaboratory variance s^2 [23]:

$$u_{\mathrm{mp}}(f_n^{\mathrm{norm}}) = \sqrt{u_{\mathrm{cut}}^2(f_n^{\mathrm{norm}}) + s^2}. \tag{15}$$

The Mandel-Paule adjusted uncertainties $u_{\mathrm{mp}}(f_n^{\mathrm{norm}})$ now replace the uncertainties $u(f_n^{\mathrm{norm}})$ in equations (8)–(11) with iteratively increasing s until the χ^2 test passes, and the new DoE is calculated.
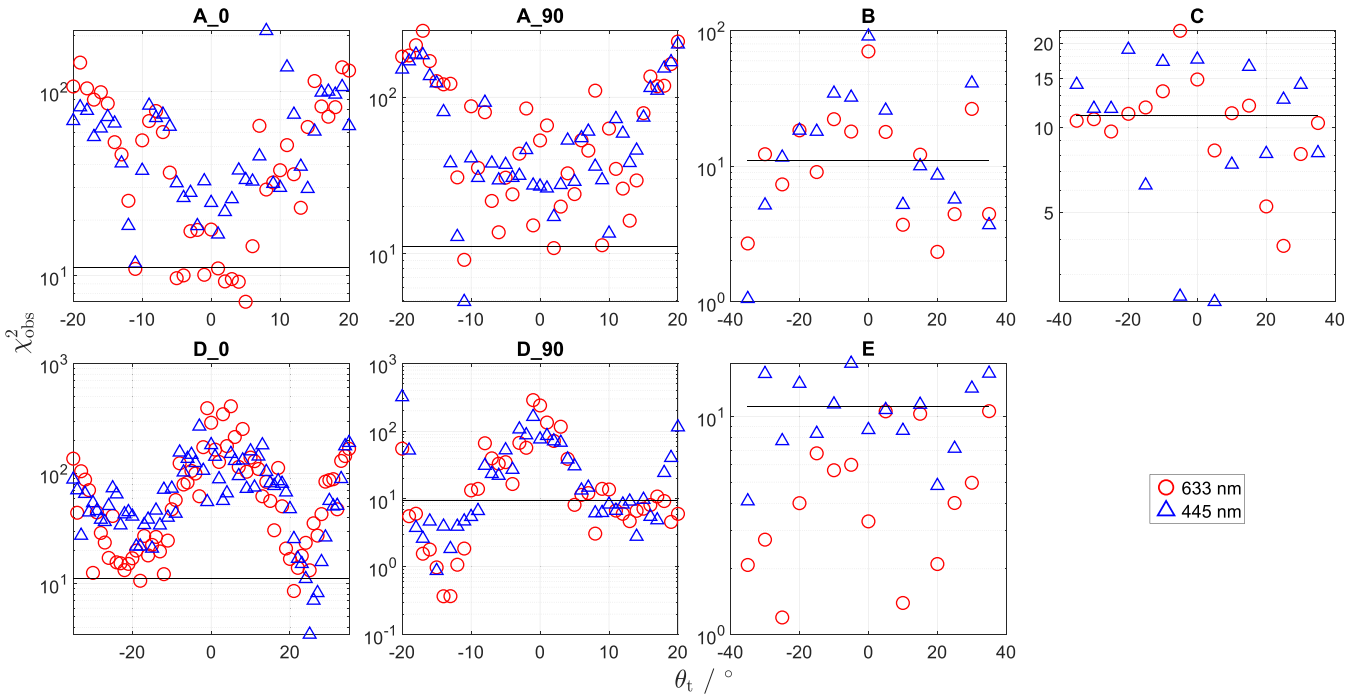
3. Results of comparison one

In this section, the results of comparison one are presented and discussed. Note that the measured BTDFs for the same sample type may deviate in both comparisons, this could be caused either by different measurement conditions of the instruments involved, or by inter-sample variation. The normalised reported data can be found in section S2 in the SM.

3.1. Consistency of reported data

Figure 5 illustrates the χ^2_{obs} values of every sample type in every geometry at both wavelengths for the initially reported measurement data. The black line in the subfigures indicates the 95 % quantile of the $\chi^2(N-1)$ distribution, resulting from the corresponding degrees of freedom. Note that there are only 5 data sets for the narrower BTDF distribution of sample D and 6 data sets for the other samples. The uncertainty components which have been taken into account by each participant are summarised in table 4. '+' and '−' sign indicate whether the component has been considered or not in the uncertainty budget by that participant and 'N/A' stands for 'not applicable' since some uncertainty components do not concern all participants. The main reasons for the observed inconsistencies for each of the samples are discussed in the following paragraphs.

3.1.1. Samples C and E.
The measurement data are in general more consistent at the longer wavelength than at the shorter one, especially for the two quasi-Lambertian samples C and E. For the latter, global consistency is achieved in every geometry at 633 nm. At 445 nm, the data are also consistent in more than half of the geometries, even though the measurements at 445 nm are more difficult due to a lower optical transmitted signal and detector response. Observations showed that sample E, being a thin foil, exhibited some mounting issues due to non-flatness. Even better congruence could be achieved if all participants had reacted on this, either by increasing the uncertainty contribution to the detection solid angle or by applying a correction on the measured BTDF value. In light of the flatness issue, it is surprising that the data for sample E show better consistency than for the rigid bulk type sample C. Possible explanations for the lower consistency of sample C could be a spatial inhomogeneity of the sample (already described in [24] for natural SiO_2 volume diffusers, more details in section S5 of the SM) and a strong lateral diffusion of the incident beam. Both lead to an enhanced dependence of the BTDF on the measurement parameters, especially the area sizes of irradiation A_{i} and measurement A_{M}.

The stronger lateral scattering by sample C compared to sample E can be observed by evaluating the spatial intensity distribution on the sample back surface taken by a high-resolution camera, aligned perpendicularly to the incident beam path. The photos of both samples irradiated with a laser beam of 5 mm diameter are shown in figure 6, with red circles indicating the incident beam area A_{i}. Due to the large difference in the thickness and the material's scattering magnitude

Figure 5. χ^2_{obs} values for all sample types at 633 nm and 445 nm for the initially reported results (details see text) in comparison one. The black line in each subfigure indicates the 95 % quantile of χ^2 distribution for the corresponding degrees of freedom. χ^2_{obs} values below the black line imply a global consistency for the results measured in those geometries by all participants, whereas χ^2_{obs} values above the black line indicate an inconsistency in the reported data and further procedures to reduce the inconsistency are performed as described in the text.

of these two sample types, the size of the irradiation spot increased to about 7.5 mm in diameter on the back surface of sample C, whereas on sample E the size of the spot almost does not change after being transmitted.

On the instruments of some participants, A_M is well-defined on the sample back surface by an imaging system. For measurements with an undefined A_M, the risk of detecting stray light apart from the scattered light is high. But even for a well-defined A_M, if it is only slightly larger than A_i, some loss in the measured BTDF is unavoidable as the diffusely transmitted radiation could no longer be fully collected within A_M. In this case the measurement scheme could no longer be regarded as under-irradiation, nor as over-irradiation, since the intensity distribution of the diffused spot may not be spatially homogeneous. Therefore, it is necessary to use an area A_M being sufficiently larger than A_i when measuring a thick bulk scattering sample, so that the whole diffused spot can be measured. On the other hand, A_M should not exceed the physical boundary of the sample as it increases with higher detection polar angle θ_t. This makes the selection of proper measurement parameters important. Their influence is illustrated in more detail in section S4 in the SM by a measurement on sample C with different A_M. While all participating labs in comparison one performed the measurement in the under-irradiated scheme, the relation between A_i and A_M varies for each facility. However, only PTB considered the lateral diffusion of the irradiation spot in the uncertainty budget and provided larger uncertainties for this sample than for the others.

3.1.2. Sample B. The χ^2 test for sample B shows consistency in only half of the geometries. At both wavelengths the inconsistency increases at smaller θ_t angles and becomes the largest at the peak of the BTDF distribution. The latter observation can be explained by the different detection angular resolution used on different instruments (see table 2). In this geometry, participants applying a high angular resolution tend to measure higher BTDF values, which more closely reproduce the true BTDF value, whereas a low resolution results in smaller values. This convolution effect can be regarded in analogy to the influence of the instrumental bandpass function in spectral measurements, as described in [25]. For the measurement facilities involved in comparison one, the detection angular resolution can almost be represented by the FWHM of the instrument function, which is determined by the size and divergence of the irradiation beam and the detector solid angle. It can be measured as the angular spread of the incident beam without the presence of a sample. The measured BTDF distribution is always the true BTDF distribution convolved with this instrument function [7]. For peak-shaped distributions (such as sample types B and D), the instrument function lowers the peak and broadens the width. Its impact is proportional to $\frac{\partial^2 f_t}{\partial \theta_t^2}$. If experimentally feasible, the instrument function should be adapted to the expected angular width of the true distribution of the sample under test. For Gaussian-shaped distributions, the FWHM of the instrument function can be recommended to be about 1/20 of the FWHM of the sample's true BTDF distribution, so that the relative difference in the peak value can be under 0.2 %, as long as the sensitivity of

Table 4. Uncertainty components considered by each participant in comparison one. '+' means 'considered' and '−' means 'not considered', 'N/A' means 'not applicable'.

Participant	Irradiation			Detection			Signal acquisition								Lateral diffusion of irradiation		
	Wavelength	Beam spatial nonuniformity	Angle positioning	Solid angle	Angular resolution	Signal readings[a]	Detector linearity	Digitisation error	Polarisation	Attenuation filter[b]	Sample inhomogeneity	Inter-reflection	Stray light	Temperature	Lateral diffusion of irradiation beam[c]	Wavelength correction[d]	Sample azimuth[e]
PTB	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DFM	−	−	+	+	+	+	+	−	−	+	−	−	+	−	−	+	−
CMI	+	−	+	+	−	+	+	−	+	+	−	+	+	+	−	N/A	−
MSL	+	+	+	+	+[f]	+	+	−	+	+	+	+	+	+	−	N/A	+
CNAM	−	−	+	+	−	+	+	−	N/A	N/A	−	−	−	−	−	N/A	−
Aalto	+	−	+	+	−	+	+	−	+	−	+	−	+	−	−	N/A	−

[a] Including fluctuation, drift, and noise for incident, transmitted, and dark signals.
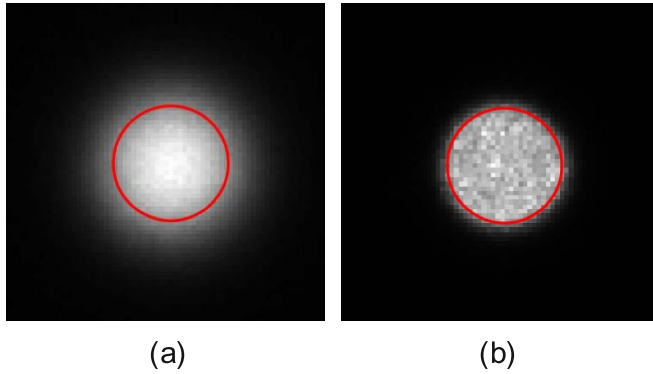[b] For CMI and MSL, the gain ratio of the amplifier for signal acquisition.
[c] Only observed on sample type C due to thick bulk volume.
[d] Only concerns PTB and DFM, other participants measured at the nominal wavelengths.
[e] Only concerns sample types A and D.
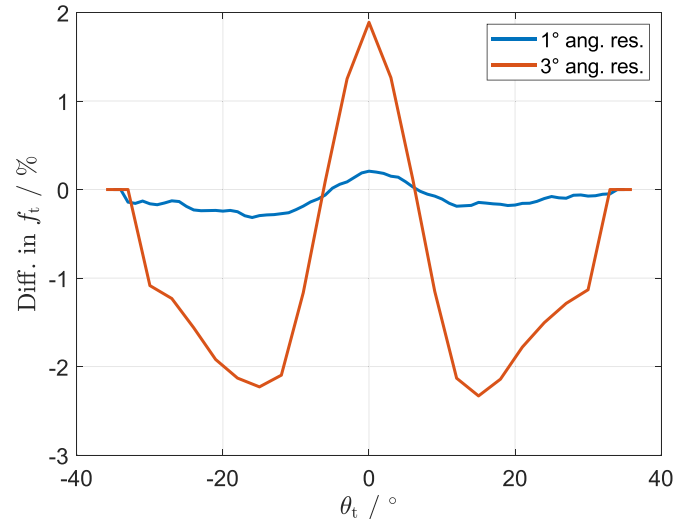[f] Only applied for sample type A.

Figure 6. Photos of the back side of (a) sample C and (b) sample E irradiated with laser beam of 5 mm in diameter (indicated by a red circle). It can be seen that the irradiation beam was transmitted with strong lateral scattering due to the thick bulk volume of sample C, resulting in an obvious size difference between the spot on the front and the back side of the sample.

Figure 7. Difference between the measured and the corrected BTDF distribution of sample B regarding the applied instrument function of 1° (blue) and 3° (red) FWHM. This highlights the importance of taking into account the influence of the applied instrument function for a proper comparison.

the instrument is high enough for an acceptable signal-to-noise ratio [26].

In comparison one, sample B was measured with detection angular resolutions ranging from 0.8° to 3°. The FWHM of the sample's true BTDF distribution is approximately 16°, corresponding to a recommended instrument function of approximately 0.8° FWHM for a moderate measurement uncertainty. If the angular resolution is too low, a correction on the measured distribution or increased measurement uncertainties would be necessary for a proper comparison. The impact of doing so is exemplified in figure 7 by performing such a correction on the measured BTDF distribution of sample B with an instrument function of 1° and 3° FWHM, respectively. Compared to the originally measured BTDF distribution, the deconvolved BTDF shows about 2 % difference at $\theta_t = 0°$ and $\theta_t = 18°$ when measured with the lower angular resolution. On the other hand, if the measurements are made with the higher resolution, the difference stays reasonably small for every θ_t angle. This highlights the importance of taking the instrument function into account when evaluating the measurement uncertainty. However, not all participants included the influence of the instrument function in their uncertainty budgets or corrected their measurement results accordingly, which contributed to the large deviations observed for the smaller θ_t angles for sample B.

3.1.3. Sample D. For sample D, the instrument function influences the measurement in a similar way. In the wide orientation ($\varphi_t = 0°$) the impact is less pronounced than in the narrow orientation ($\varphi_t = 90°$). This is discussed in detail in the SM (section S6). Another aspect regarding the applied angular resolution, which was observed in all PTB's measurements at the wide orientation, is the spiky behaviour in the central angular area of the BTDF distribution (see figures S12 and S13 in the SM). It is assumed that the scattering by structures on the holographic layer, causing this spiky behaviour, was resolved only by PTB's highest applied detection angular resolution

of 0.5°. This again emphasises the necessity of having well-defined measurement parameters for a future key comparison.

Apart from the instrument function, a limited accuracy in realising the azimuthal orientation can also strongly contribute to the measurement uncertainty of sample D. As outlined in section 2.2, the orientation of samples was pre-determined and marked. If the uncertainty in the sample azimuth alignment is approximately 1°, the associated uncertainty due to this misalignment can be up to 0.5 % to the BTDF. However, this uncertainty source was only considered by PTB and MSL in the uncertainty analysis. This might explain why the inconsistency of sample D is larger than that for sample B, because sample D is much more sensitive to uncertainties regarding the azimuthal angle.

3.1.4. Sample A. Orientation-related uncertainties might also explain the data inconsistency observed with sample A. It is supposed that the BTDF of sample A, which exhibits a square-shaped top-hat distribution, should be independent of φ_t for θ_t angles close to 0°. On the other hand, the width of the top-hat part of the BTDF distribution would be larger for φ_t deviating from the correct value. Thus, the inconsistency is large at the falling edges ($9° < |\theta_t| < 11°$). In the central angle range, the reported data are only consistent for $\varphi_t = 0°$ at 633 nm, but not for $\varphi_t = 90°$. This could be attributed to the micro-lens units on the sample's functional surface, which generate an imperfect top-hat BTDF distribution with slight variation in the central angular range. Therefore, the measured BTDFs in this range do depend on φ_t. Again, the uncertainty regarding the sample azimuth was only considered by two participants, so this likely contributed to the inconsistency that was observed for $\varphi_t = 90°$. More detail about this is given in section S7 of the SM.
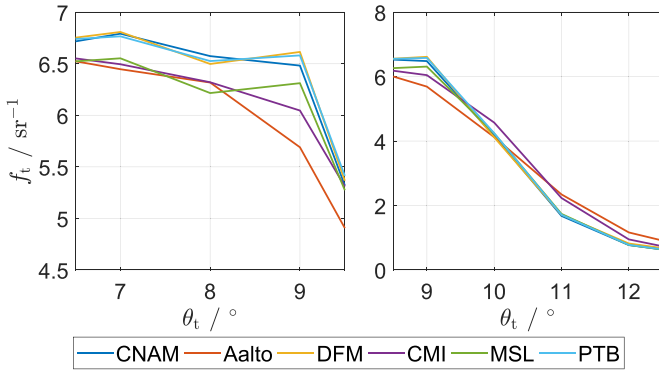
Figure 8. Different angular ranges of normalised BTDF distribution of sample A for $\varphi_t = 0°$ at 633 nm measured by all participants in comparison one. It can be observed that participants who applied a high detection angular resolution could measure the sharp variation in the BTDF distribution of sample A, indicated by the similar shape of the DFM, MSL and PTB curves in both angular ranges.

The instrument function influences the measurement of sample A in the way that the sharp variation ranging from $\theta_t = 7°$ to $9°$ ('batwing') and from $9°$ to $12°$ (falling edge) in the BTDF distribution is modified (figure 8). The variations in the batwings were resolved by CNAM, DFM, MSL, and PTB, who all applied similarly high angular resolutions. CMI and Aalto used a lower angular resolution. Thus, their measured BTDF distributions do not exhibit these structures. The observations for the falling edge of sample A's BTDF distribution also underline this issue, as the curves measured by CMI and Aalto exhibit a flatter slope than the other four participants. Therefore, it is not surprising that the global consistency is not achieved in these angular ranges.

3.2. Degrees of equivalence

Since the χ^2_{obs} calculated from the reported data does not fulfil the requirement $\chi^2_{\text{obs}} \leqslant \chi^2_{0.05}(N-1)$ for almost all of the sample types (except for sample E at 633 nm), the normalised uncertainties of all participants are first adjusted using the cut-off. After applying the cut-off, the data becomes consistent for sample C at both wavelengths, but for other samples the inconsistency is only slightly reduced. This implies a general underestimation of uncertainties, for which some reasons were given in the previous section. Therefore, the uncertainties were further increased using the Mandel-Paule method as described in section 2.4, although the additional interlaboratory-variance only accounts for shortcomings in the determination of uncertainty in an unspecified way. The procedure is performed until the χ^2 test passes for each geometry. For sample D at $\varphi_t = 0°$ the data points at $|\theta_t| > 35°$ are excluded in the Mandel-Paule adjustment since the BTDF in these geometries is almost zero. The development of each participant's uncertainties regarding the cut-off and the Mandel-Paule adjustment is plotted for each sample in section S8 in the SM.

The DoEs are then calculated using the metric $d_n/U(d_n)$ (simplified as d/U in the following text) for each participant after the uncertainty adjustment, and results are presented in

figures 9 and 10 for the comparison at 633 nm and 445 nm, respectively. For measurement data with reasonable uncertainties, the d/U metric of each sample should be scattered between -1 and 1. For some participants, their d/U metric shows this behaviour after the uncertainty adjustment, while the d/U of other participants still indicates either a poor measurement, an under- or overestimated uncertainty, or a wrong basic assumption, for example, when distributions measured with considerably different detection angular resolution are compared.

In general, we can see that the d/U values for DFM are close to 0 for all samples at both wavelengths, suggesting that DFM overestimated all of their measurement uncertainties. In the following paragraphs, we consider the d/U metrics for each sample in turn.

3.2.1. Sample C. For sample C at 633 nm, the d/U metric of Aalto's data splits from the other participants, with the Aalto values being lower than the others. Since Aalto submitted low uncertainties for this sample, they would have had a large weight in the calculation of the reference value, dragging the reference value closer to their measurements. The deviation between Aalto and the other participants suggests that it is likely that Aalto underestimated their uncertainty.

3.2.2. Sample B. For sample B at 633 nm, the d/U values of all participants are scattered between -1 and 1, with only a few lying outside this range. This suggests that the adjusted uncertainties are adequate to describe the deviations observed between the participants. The inequivalence of CNAM's data in the central angular range, with the d/U values above 1, might suggest a measurement problem, as the d/U metric for CNAM's measurements at other θ_t angles stays close to -1. At 445 nm, PTB's data show inequivalence for negative θ_t angles but are equivalent for $\theta_t \geqslant 10°$. This shows some asymmetry in the measurement, together with possibly overlooked uncertainty components. The influence of detection angular resolution is observed by Aalto's generally close to -1 d/U values at both wavelengths, due to Aalto's measured values being lower than other participants.

3.2.3. Sample D. For sample D, the equivalence is only achieved for DFM, CNAM, and CMI at the wide orientation even after the uncertainty adjustment. PTB tended to measure higher BTDF values in the central angular range, whereas MSL measured lower values. This may be explained by the difference in the detection angular resolution applied on different instruments. However, the uncertainties provided by both participants could not cover the difference between their data, leading to the poor consistency. Similar to PTB's data of sample B at 445 nm, the difference in the d/U metric of Aalto's data between positive and negative θ_t angles might imply some asymmetry in their measurement setup, along with generally understated uncertainties.

For the data of the narrow sample D orientation, inequivalence in PTB's data is observed in the central detection
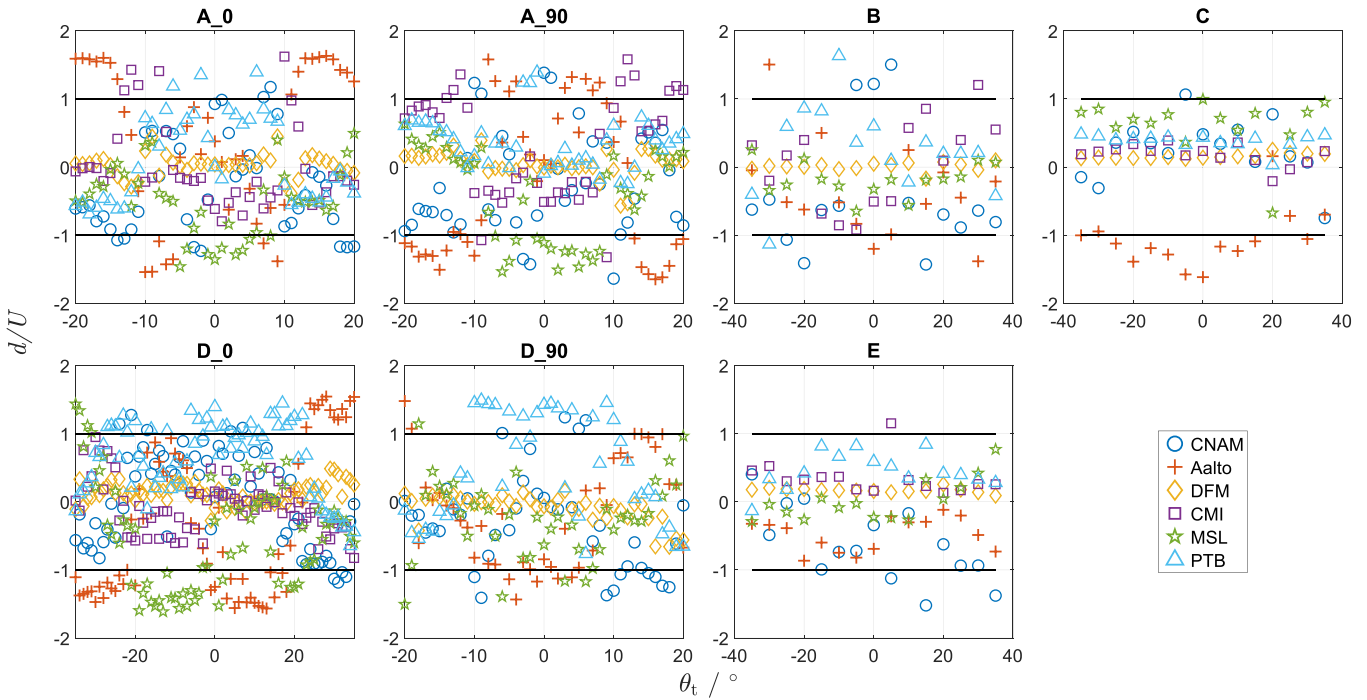
Figure 9. d/U metric for all participants in comparison one at 633 nm, after uncertainty adjustment using cut-off and Mandel-Paule method. Points randomly scattered within the −1 to 1 interval indicate equivalent measurements of the sample with adequate uncertainty analysis.
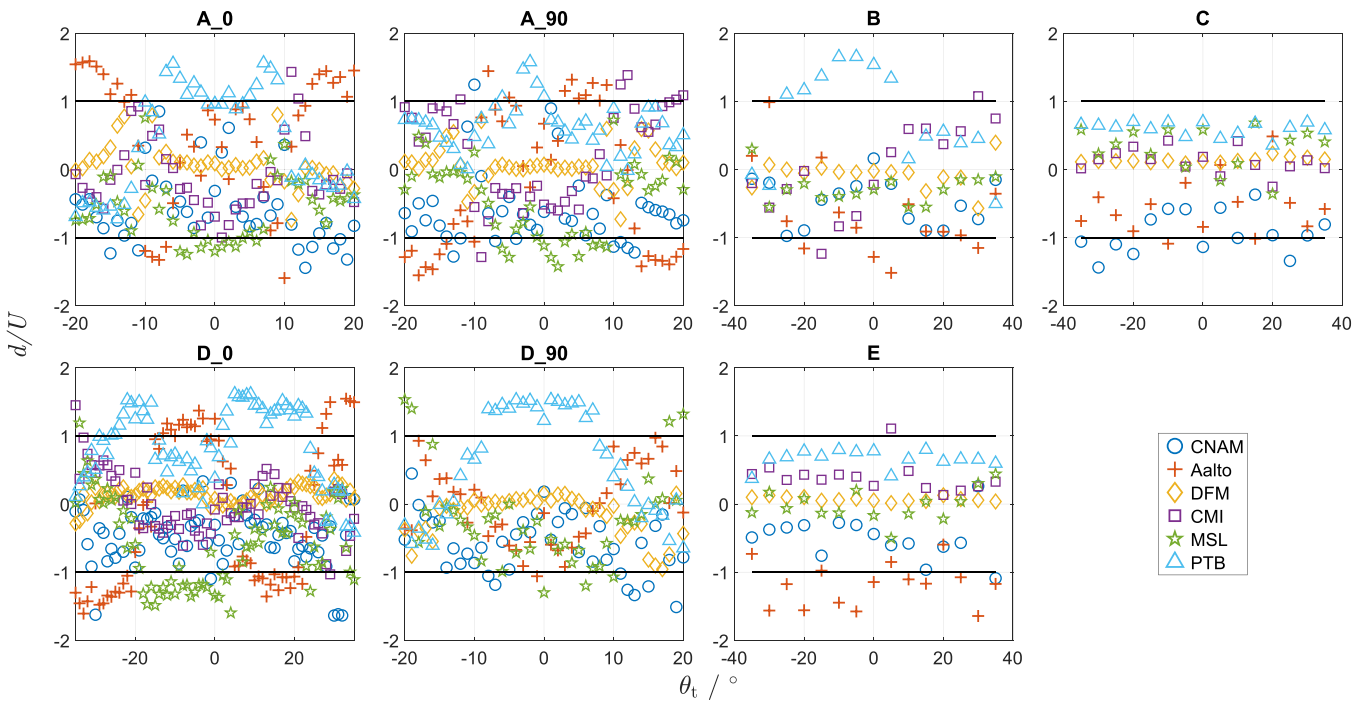
Figure 10. d/U metric for all participants in comparison one at 445 nm, after uncertainty adjustment using cut-off and Mandel-Paule method. Points randomly scattered within the −1 to 1 interval indicate equivalent measurements of the sample with adequate uncertainty analysis.

range, as PTB reported the highest BTDF values in this measurement. Apart from the influence of the instrument function as discussed in the previous section, another reason is that the measurements of the inter-pilot sample D failed to connect the lab scales between the two pilots and thus, the two

sub-groups of the comparison. Only for this specific sample, the BTDF value reported by DFM deviates from other samples of the same type and is much lower than the value reported by PTB, whereas PTB's measurements of different sample D did not reveal large inter-set variation. As a result, the
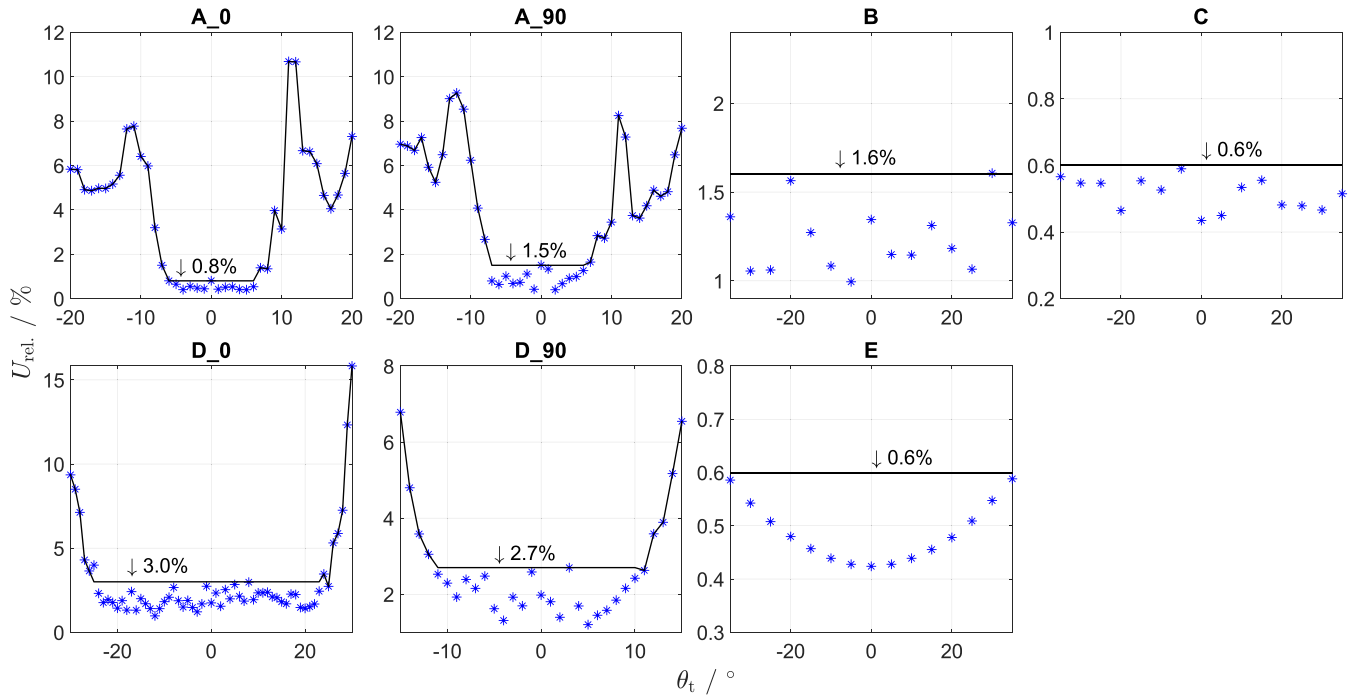
Figure 11. Relative standard ($k = 1$) uncertainty of every sample type for a fictive mean participant after applying cut-off and Mandel-Paule method. The fictive mean participant is taken on that number n of participants, who possess the smallest uncertainty for a specific sample type. In order of occurrence $n = 3, 3, 3, 2, 3, 4, 2$. In this way, the participants' potential to perform high quality BTDF measurements could be assessed despite that the agreement in comparison one is not sufficiently good.

normalisation incorrectly lowered the data from those participants in DFM's sub-group. In PTB's sub-group, on the other hand, CMI did not supply data for sample D at narrow orientation and MSL reported lower BTDF values. Therefore, only the data from PTB are left much higher while calculating the DoE.

3.2.4. Sample A. For sample A in both principal azimuths, measured at 633 nm, the equivalence of PTB's and CNAM's data is generally good. The inequivalence in CMI's and Aalto's data again indicates the problem of applying low detection angular resolutions to measure the sharp variations in the distribution. On the contrary, MSL's data show inequivalence at angles where the BTDF only exhibits slight variation. The d/U metric below -1 indicates generally low BTDF values. The tendency that MSL measured lower values whereas PTB did the opposite can only be understood as a systematic difference between their measurement scales. The type of irradiation source may be identified as a possible influence factor. For BTDF measurements, narrowband CW lasers were used at PTB, while a pulsed supercontinuum light source was used for irradiation in the longer wavelength range at MSL. With the latter source, ps-pulsed irradiation with high peak power is applied. However, possible influence on, for example, filter transmittance or other optical components was not studied.

For sample A at 445 nm, inequivalence is observed in PTB's data while CMI's data become equivalent. A possible explanation would be a problematic normalisation process caused by the absence of DFM measurements for this

wavelength. This is described in more detail in section S3 of the SM.

3.3. Discussion

In the previous sections (and in the SM), the measurement results and related uncertainty contributions are analysed sample-, instrument-, and geometry-specifically. The main effects that led to the observed inconsistencies between the involved participants are differences in the irradiation and measurement area sizes A_i and A_M, differences in the applied detection angular resolution, and misalignment regarding the samples' azimuthal orientation. These identified effects can be considered by participants in the future to perform improved BTDF measurements and to set up more reliable uncertainty budgets.

Even though the agreement in comparison one is not sufficiently good, it is possible to use the results from this comparison to assess the participants' potential to perform high quality BTDF measurements. This can be done by using only a subset of data to generate a so-called 'fictive mean participant', as shown in figure 11. The relative standard uncertainty of this fictive mean participant was calculated by taking the mean of the uncertainties of the labs with the smallest uncertainty values after applying the processes described in section 2.4. The sample-specific relative uncertainties for the fictive mean participant plotted in the subfigures range from 0.6 % to 3 %.

The lowest uncertainties can be achieved for samples C and E. The uncertainties for these samples are close to the

uncertainties reported in [2], indicating that at least a subset of participants, applying adequate experimental settings, is capable of performing equivalent measurements.

Another aspect for future improvement concerns the procedure of the comparison. Although some work was invested in setting up technical protocols, some measurement parameters were only specified as a recommendation, so as to not exclude project partners with less flexible setups from participating in this comparison. This has led, in some cases, to measurement results which are not comparable by principle, if resulting effects are not corrected. A prominent example is the detection angular resolution, which must be stated precisely in future comparisons.

Despite the observed shortcomings, the results from this comparison allow us to identify potential reference samples for BTDF measurements. A thin PTFE foil like sample E, being highly insensitive to varying experimental settings, may serve as a good candidate for a Lambertian laboratory standard. Azimuthally independent samples like sample B are capable of serving as narrow-angle references. Mie-scatter samples like sample C find frequent use in aerospace measurements due to their good physical properties and stability [27, 28]. However, the mentioned lateral diffusion effect should be taken into account when referring to the true BTDF value of this sample type or measurements must be performed under similar conditions.

Not surprisingly, the largest inconsistencies were observed for the samples with high azimuthal dependence. Besides the mentioned instrument function issue, the results imply that the involved participants must step up their efforts in defining and aligning to the principle-axes of the samples for future comparisons.

4. Results of comparison two

In this section the results of comparison two are presented and discussed. In order to distinguish between the errors introduced by angular evaluation and those from the absolute value assigned to the measurement, all the reported measurement results were further adjusted for having exactly the same reference value calculated for $\theta_t = 0°$ geometry. This way, the deviations observed in other geometries are independent of the absolute calibration, but only include the other typical sources of uncertainties in goniospectrophotometric measurements. The absolute value of the measurements, required for transferring the scale, was provided by using a reflectance standard as reference by most of the participants. The method used here avoids including the deviations between standards. In addition, it excludes the issue with the different size of irradiated area that is explained later in this section.

The results of the BTDF measurements, after adjustments explained above and the evaluation procedure described in section 2.4, can be found in section S2 in the SM. In general, although there are large differences among the participants, it can be stated that the relative uncertainties are generally higher at steeper slopes in the BTDF distribution. The DoEs for the measurement of each participant are given in figure 12.

Measurements are consistent for smaller θ_t, and they become more inconsistent as θ_t increases. It is hard to draw a conclusion on the impact of the angular distribution of the BTDF. For instance, there is no significant difference between samples B (narrow distribution) and C (wide distribution). In general, we might say that the reported uncertainties are well estimated. The exception might be the measurements from Temicon. The large inconsistency might be due to an underestimation of the impact of the straylight in their measuring system. It is also noticeable that the data from Covestro and Saint-Gobain are extremely consistent for sample C, and slightly less for other samples. It could point to an overestimation of the reported uncertainty. But it might also be that the reported uncertainty at any detection angle is almost completely dominated by the uncertainty from the scale transfer at 0°. Note that Saint-Gobain and Covestro generally reported the largest uncertainties (see figure S15 in the SM).

As explained above, the error committed by transferring the scale at a known geometry was excluded in the previous analysis. This partially avoids considering the differences due to the different irradiated areas in the comparison. Its impact is explained and shown in the following lines.

According to the considerations in [29], the radiance of a translucent object, with high contribution of scattering within the bulk, is proportional to the incident radiant flux. When measuring the BTDF, the evaluated radiance L_t per unit of irradiance E_i increases for larger irradiated areas A_i because more radiant flux is entering the material. As a consequence, the BTDF is higher for larger A_i, and this effect is more significant when the contribution of the scattering in the bulk with respect to the scattering at the surface is larger.

Since each participant in this comparison kept a constant A_i for all measurements, and, in addition, each participant used a different value for this parameter, it is possible to study the dependence of the BTDF on it. This is shown in figure 13, where, for the sake of a clear visual comparison, the ratio of the measured BTDFs at $\theta_t = 0°$ with respect to the BTDF measured with the largest A_i (at CSIC) is shown versus the irradiated area size used on the participating measuring systems (excluding Temicon, who did not report absolute BTDF values), for all samples involved in comparison two. It is observed that the ratio decreases consistently towards lower irradiated areas, as expected. In addition, the results seem coherent with the hypothesis that this effect is larger for samples possessing stronger scattering in the bulk. The lowest variation is observed for sample B, which presents mainly surface scattering and seems rather transparent. Samples E and Co1 show a larger and similar variation (around 15 %), as they are both volume scattering samples with small thicknesses (0.25 mm and 1 mm, respectively). The largest variation (almost 30 %) is observed for sample Co3, which has the same material as Co1 but with a larger thickness (3 mm), and also for sample C, which also has a thick bulk scattering volume (2 mm). These results show the importance of an adequate irradiated area for having a well-defined BTDF measurand. For over-irradiated measurements, the irradiated area has to be large enough to avoid variation in the measurement caused by different sizes of A_i. For under-irradiated measurements, the measurement area A_M needs to
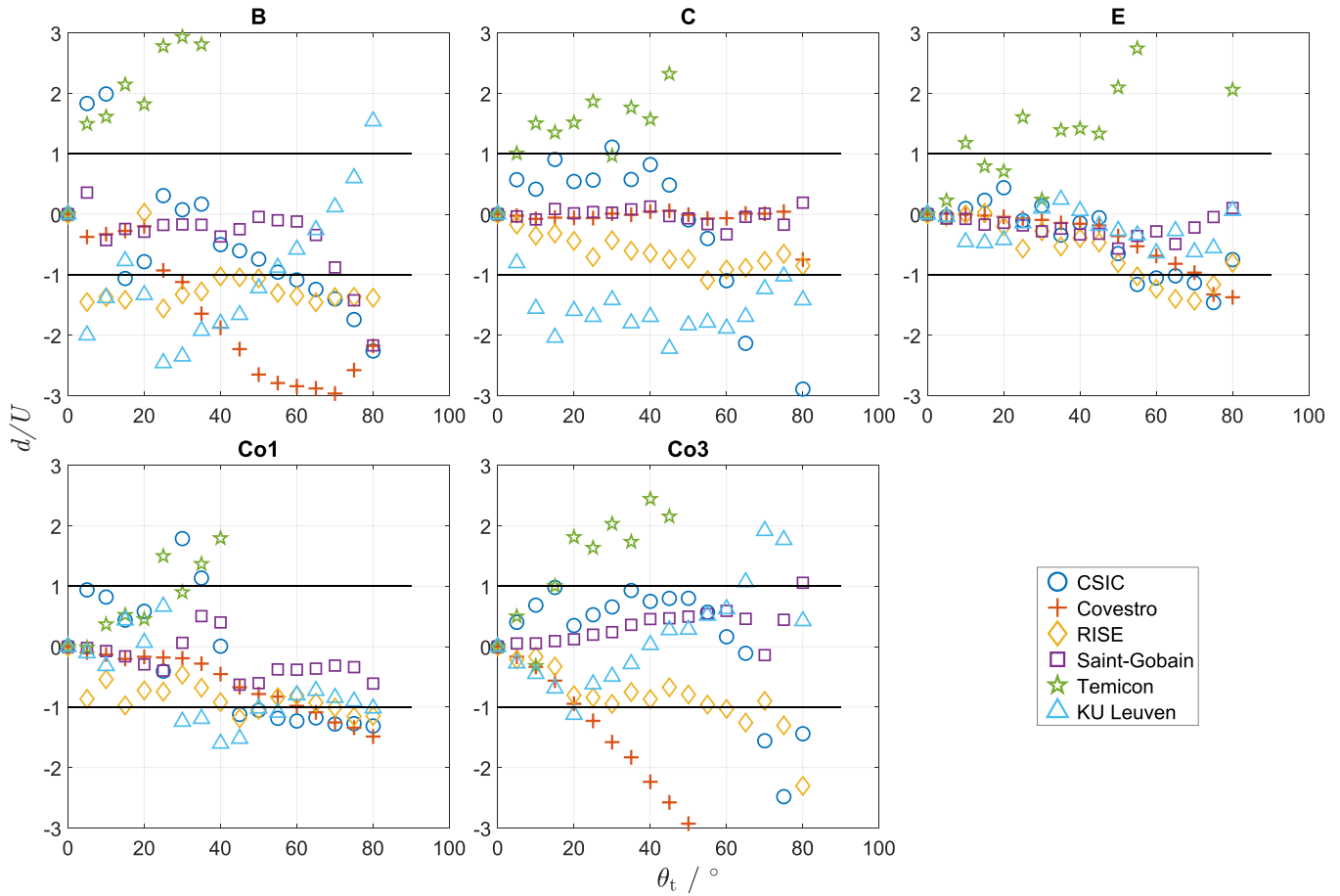
Figure 12. d/U metric for participants in comparison two. Points randomly scattered within the -1 to 1 interval indicate equivalent measurements of the sample with adequate uncertainty analysis.
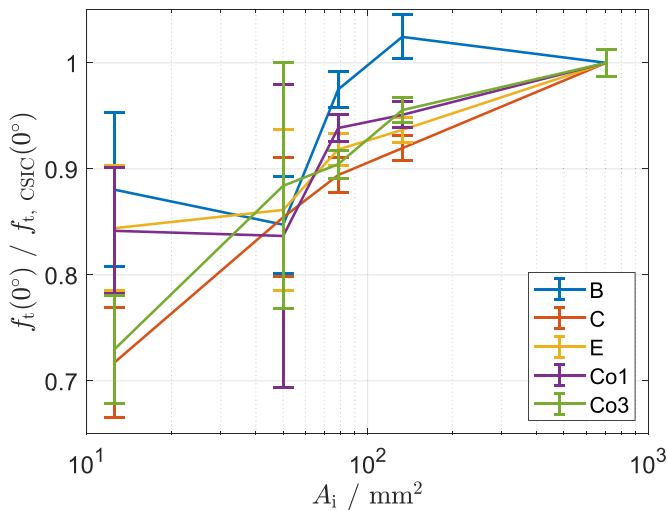
Figure 13. Dependence of the measured BTDF ($\theta_t = 0°$) on the applied irradiated area size A_i in comparison two. For a clear visual comparison, all BTDF values are shown relative to the ones measured by CSIC, who applied the largest A_i in all measurements.

be large enough, as discussed in section 3.1. The inconvenience of this second measurement scheme is that for thick and translucent samples A_M should be very large, and the measured

BTDF is only an average within A_M, which might exhibit high spatial variation. It makes the results measured from different measurement schemes not completely comparable.

5. Conclusion

In this paper, the capability of performing accurate BTDF measurements on the facilities of six NMIs, two DIs, one university, and three industrial partners was thoroughly studied by means of two multilateral scale intercomparisons. In comparison one, the NMIs and the DI could perform consistent measurements on the quasi-Lambertian samples regardless of the different measurement conditions of each facility. However, for samples possessing a narrow peak or rapid change in their BTDF distribution, inconsistency was observed between the participants, mainly due to the difference in the applied detection angular resolution. The results of samples with dedicated azimuthal dependence implied several aspects that require more attention for improved future measurements. One important aspect was a better determination and realisation of the sample azimuth angle. Despite some inconsistency in the results, the potential to perform high-quality BTDF measurement equivalent to the state of the art was demonstrated by a subset of the participants.

In comparison two, the generally consistent measurement results successfully connected the scale between metrology institutes and industrial partners. There was one common issue observed in both comparisons when measuring highly translucent samples with large contribution of scattering coming from the bulk volume. The lateral diffusion of the irradiation spot made the measurement performed in the under-irradiated scheme, as most of the participants did in both comparisons, strongly dependent on the measurement conditions. However, even in the over-irradiated scheme the measured BTDF still exhibited dependence on the size of the irradiated area. This behaviour needs further investigation for more consistent results.

For future comparisons, a more precise definition of measurement parameters is crucial and a thorough study on the sample characteristics would allow a better congruence in the participants' results.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

ORCID iDs

Jinglin Fu ⬤ https://orcid.org/0000-0001-8260-8017
Alejandro Ferrero ⬤ https://orcid.org/0000-0003-2633-3906
Tatjana Quast ⬤ https://orcid.org/0009-0000-6165-0195
Pablo Santafé-Gabarda ⬤ https://orcid.org/0000-0001-6319-8091
Néstor Tejedor ⬤ https://orcid.org/0000-0003-1577-5549
Joaquín Campos ⬤ https://orcid.org/0000-0001-6527-1440
Lou Gevaux ⬤ https://orcid.org/0000-0001-9070-9467
Gaël Obein ⬤ https://orcid.org/0000-0002-2577-6361
Robin Aschan ⬤ https://orcid.org/0009-0005-3033-0777
Farshid Manoocheri ⬤ https://orcid.org/0000-0003-3935-3930
Erkki Ikonen ⬤ https://orcid.org/0000-0001-6444-5330
Geiland Porrovecchio ⬤ https://orcid.org/0000-0002-4765-5675
Ellie Molloy ⬤ https://orcid.org/0000-0001-5563-8237
Annette Koo ⬤ https://orcid.org/0000-0001-9390-345X
Jan Audenaert ⬤ https://orcid.org/0000-0001-6260-1950
Youri Meuret ⬤ https://orcid.org/0000-0002-2815-5915
Stefan Källberg ⬤ https://orcid.org/0009-0003-8259-3400
Iryna Gozhyk ⬤ https://orcid.org/0000-0002-8287-307X
Alfred Schirmacher ⬤ https://orcid.org/0000-0002-2228-7853

References

[1] Bureau international des poids et mesures (BIPM) Pavillon de Breteuil, F-92312 Sèvres Cedex France (available at: www.bipm.org)

[2] Butler J J, Georgiev G T and Cooksey C C 2019 *Metrologia* **56** 065008

[3] EMPIR JRP 18SIB03 New quantities for the measurement of appearance (available at: https://bxdiff.cmi.cz/)

[4] Bartell F O, Dereniak E L and Wolfe W L 1981 *Proc. SPIE* **0257** 154–60

[5] Nicodemus F E 1965 *Appl. Opt.* **4** 767–75

[6] Nicodemus F E, Richmond J C, Hsia J J, Ginsberg I W and Limperis T 1977 Geometrical considerations and nomenclature for reflectanc *Natl Bur. Stand. Monogr.* **160** 4 (available at: http://graphics.stanford.edu/courses/cs448-05-winter/papers/nicodemus-brdf-nist.pdf)

[7] Stover J C 2012 *Optical Scattering: Measurement and Analysis* 3rd edn (SPIE Press)

[8] Germer T, Zwinkels J and Tsai B 2014 *Spectrophotometry: Accurate Measurement of Optical Properties of Materials* 1st edn (Elsevier)

[9] Fu J, Frisvad J R, Esslinger M, Quast T and Schirmacher A 2022 *11th Colour and Visual Computing Symp.* vol 3271 (available at: https://ceur-ws.org/Vol-3271/Paper13_CVCS2022.pdf)

[10] Fu J, Quast T, Velke E, Esslinger M, Pastuschek M, Schirmacher A, Aschan R, Manoocheri F and Ikonen E 2023 *Rev. Sci. Instrum.* **94** 083102

[11] Schirmacher A and Hesse S 2013 Neuer Referenzmessplatz für gerichtete spektrale Transmission und Reflexion *DGaO-Proc* (available at: https://www.dgao-proceedings.de/download/114/114_p3.pdf)

[12] Schirmacher A, Hesse S and Velke E 2014 *12th Int. Conf. on New Developments and Applications in Optical Radiometry (NEWRAD 2014)* pp 221–2 (available at: http://newrad2014.aalto.fi/Newrad2014_Proceedings.pdf)

[13] Ouarets S, Ged G, Razet A and Obein G 2012 A new gonioreflectometer for the measurement of the bidirectional reflectance distribution function (brdf) at LNE-CNAM *Proc. of CIE* pp 687–91

[14] Molloy E 2023 Metrology of Scattering Distributions *PhD Thesis* (https://doi.org/10.26686/wgtn.23735001) Victoria University of Wellington

[15] Nevas S, Manoocheri F and Ikonen E 2004 *Appl. Opt.* **43** 6391–9

[16] Rabal A M, Ferrero A, Campos J, Fontecha J L, Pons A, Rubiño A M and Corróns A 2012 *Metrologia* **49** 213

[17] Bernad B, Ferrero A, Pons A, Hernanz M L and Campos J 2015 *Proc. SPIE* **9398** 93980E

[18] Turbil C, Yoo T S H, Simonsen I, Teisseire J, Gozhyk I and Garcia-Caurel E 2019 *Appl. Opt.* **58** 9267–78

[19] Leloup F B, Forment S, Dutré P, Pointer M R and Hanselaer P 2008 *Appl. Opt.* **47** 5454–67

[20] Cox M G 2002 The evaluation of key comparison data *Metrologia* **39** 589–95

[21] CCPR WG-KC 2019 Guidelines for CCPR key comparison report preparation (BIPM) (available at: www.bipm.org/documents/20126/30126060/CCPR-G2+%28rev.+4%29+Guidelines+for+CCPR+key+comparison+report+preparation/ad4d6e1a-bbdb-e272-7d37-73ac80005805)

[22] Paule R C and Mandel J 1982 *J. Res. Natl Bur. Stand.* **87** 377–85

[23] Ongoing discussion suggests that the interlaboratory variance *s* should be added to the reported uncertainty instead of to the cut-off-adjusted uncertainty. Here we follow the instruction in [21]

[24] Heath D F and Georgiev G 2011 *Proc. SPIE* **8153** 321–34

[25] Zwinkels J, Woolliams E, Gardner J, Baribeau R and Bialek A 2014 *CIE 214:2014 Effect of instrumental bandpass function and measurement interval on spectral quantities* (Beuth Verlag)

[26] EMPIR JRP 18SIB03 BxDiff Project 2023 *Good Practice Guide for the Measurement of BTDF* (available at: https://bxdiff.cmi.cz/wp-content/uploads/2023/03/Good-practice-guide-BTDF.pdf)

[27] Hilsenrath E, Williams D and Frederick J 1988 *Proc. SPIE* **0924** 215–22

[28] Dobber M *et al* 2006 *IEEE Trans. Geosci. Remote Sens.* **44** 1209–38

[29] Ferrero A, Frisvad J R, Simonot L, Santafé P, Schirmacher A, Campos J and Hebert M 2021 *Opt. Express* **29** 219–31