RESEARCH   **Open Access**

Round robin comparison study on the form measurement of optical freeform surfaces

Ines Fortmeier[1]* , Reyko Schachtschneider[2], Vit Ledl[3], Ondrej Matousek[3], Jens Siepmann[4],
Antonia Harsch[5], Rolf Beisswanger[5], Youichi Bitou[6], Yohan Kondo[6], Michael Schulz[1] and Clemens Elster[2]

Abstract

A round robin comparison of freeform form measurements was carried out by the project partners and stakeholders of a European metrology research project. Altogether six measuring instruments were considered: five different (pointwise and areal) optical devices and one tactile device. Three optical freeform surfaces were used for the comparison measurements, where two specimens were measured by five instruments and one specimen by four instruments. In this paper, the evaluation methods and results of this round robin are presented for the three freeform surfaces made from a temperature-stable material, Super Invar®. The freeforms had diameters of 40 mm, 50 mm and 100 mm and best-fit radii of 39.75 mm (convex), 40.9 mm (convex) and 423.5 mm (concave). For comparison, the bilateral pointwise differences between the available measurements were calculated. The root-mean-square values of these differences ranged from 15 nm to 110 nm (neglecting spherical contributions) and provided an insight into the status of typical freeform measurement capabilities for optical surfaces.

Keywords: Freeform optical surfaces, Metrology, Interlaboratory comparison

Introduction

Aspherical and freeform optics are essential parts of many modern optical imaging systems [1]. These "optical" surfaces (i.e. surfaces suitable for optical imaging) are smooth and have form accuracy specifications which are typically in the range of some ten nanometers. These specifications distinguish this class of surfaces from others - for example, from illumination-forming surfaces whose tolerances are much higher. Several high-accuracy techniques are applied to manufacture these optical surfaces, examples of which include ultra-precise diamond turning [2], magneto-rheological finishing [3], ion beam figuring [4] and other local polishing techniques [5]. Accurate manufacturing is impossible without accurate measurement for form control during the manufacturing process. Concerning manufacturing machines for optical surfaces which can remove material of single-nanometre thickness,

measurement accuracy needs to be improved in order to improve manufacturing accuracy. Currently, no general reference measurement system with the required accuracy exists [6].

In this situation, comparisons are a key factor in gaining new insights. Comparisons [6–9] have shown that, while some progress has been achieved, further improvements are also necessary. Most of the specimens used in these comparison studies were conventional spherical or aspherical surfaces made of glass. The study with the largest variety of instruments and measurement data was published anonymously [6].

To address the lack of reference freeform measurement systems and surfaces, the European metrology project EMPIR *15SIB01: FreeFORM* [10] was launched in 2016 and dedicated to improving the manufacturing and measurement capabilities of national metrology institutes (NMIs), other institutes and industrial firms concerning aspherical and freeform standards. Several NMIs, other institutes and stakeholders from industry were involved in this project.

*Correspondence: ines.fortmeier@ptb.de
[1]Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany
Full list of author information is available at the end of the article

One particular focus of the project was to design, manufacture and investigate several freeform (reference) surfaces made from temperature-stable material. While glass-like materials (e.g. Zerodur) can be chosen for simple surface forms, and conventional polishing techniques exist, for complex surface forms, diamond turning of metal optics such as mirrors is suitable but challenging. The metrology of surfaces with complex surface forms is investigated in this paper. In a round robin, the surface forms of the freeform specimens manufactured were measured by the project partners and stakeholders. After the project, the specimens will be available for institutions interested in investigating their measuring capabilities and comparing their measurement results to the results published in this paper.

The purpose of this round robin was to investigate the state of the art of optical and tactile freeform measuring instruments, and to develop and apply evaluation techniques for such measurement comparisons.

In this paper, the round robin results for the form measurements of three freeform surfaces are shown. The paper is structured as follows: the specimens are described in "Specimens" section. The partners and the instruments used are listed and described in "Partners and Instruments" section. In "Data analysis" section, the data analysis methods used in the comparison of the round robin measurements are presented. The comparison results are shown and discussed in "Results and discussion" section. Finally, some conclusions are drawn in "Conclusions" section.

Specimens

Three specimens were investigated: a two-radii specimen, a convex toroid and a concave toroid (see Fig. 1).

All three specimens were designed at PTB and manufactured by IPP using single-point diamond turning (SPDT). The specimens are made of Super Invar˚, which was chosen because of its good thermal properties. Such properties are important within the scope of the comparison, as they allow significant form deviations caused by temperature effects to be avoided. For example, when the temperature is changed by 1 K, if a spherical mirror made of aluminium has a 40 mm radius, the radius will change by about 1 μm, while the radius of the same mirror made of Super Invar˚ will only change by about 40 nm. For this reason, the lower machinability of Super Invar˚ compared to other material has been accepted. The plots of the design forms and the deviations from the best-fit sphere are shown for each of the three specimens in Fig. 2. The basic data of the specimens are shown in Table 1; a short description of the specimens is given in the following:

Two-radii specimen The two-radii specimen consists of two pairs of spherical segments with differing radii (r_1 =39.5 mm, r_2 =40 mm) but the same vertex in the centre of the specimen. The transitions between the spherical segments are defined by cosine functions, leading to smooth transitions. The mathematical description of the surface form is complex and can be found in [11]. The specimen has a clear aperture of 40 mm and a best-fit radius of 39.75 mm. The peak-to-valley deviation from the best-fit sphere amounts to 79 μm (see Fig. 2, lower row, left). More details on this specimen design can be found in [11].

Convex toroid This specimen has a convex toroidal surface form. It has two different radii in two orthogonal directions. The mathematical description of the surface form is given by

$$z(x,y) = \sqrt{(\sqrt{r_v^2 - y^2} + r_h - r_v)^2 - x^2} - r_h. \qquad (1)$$

Fig. 1 Photos of the three specimens: The two-radii specimen (left), the convex toroidal surface (middle) and the concave toroidal surface (right)

Fig. 2 Design form and deviation from the best-fit sphere for the three specimens: The top row shows (from left to right) the design form of the two-radii specimen (left), the convex toroidal surface (middle) and the concave toroidal surface (right). The bottom row shows (from left to right) the corresponding deviations from the best-fit sphere for the two-radii specimen (left), the convex toroidal surface (middle) and the concave toroidal surface (right)
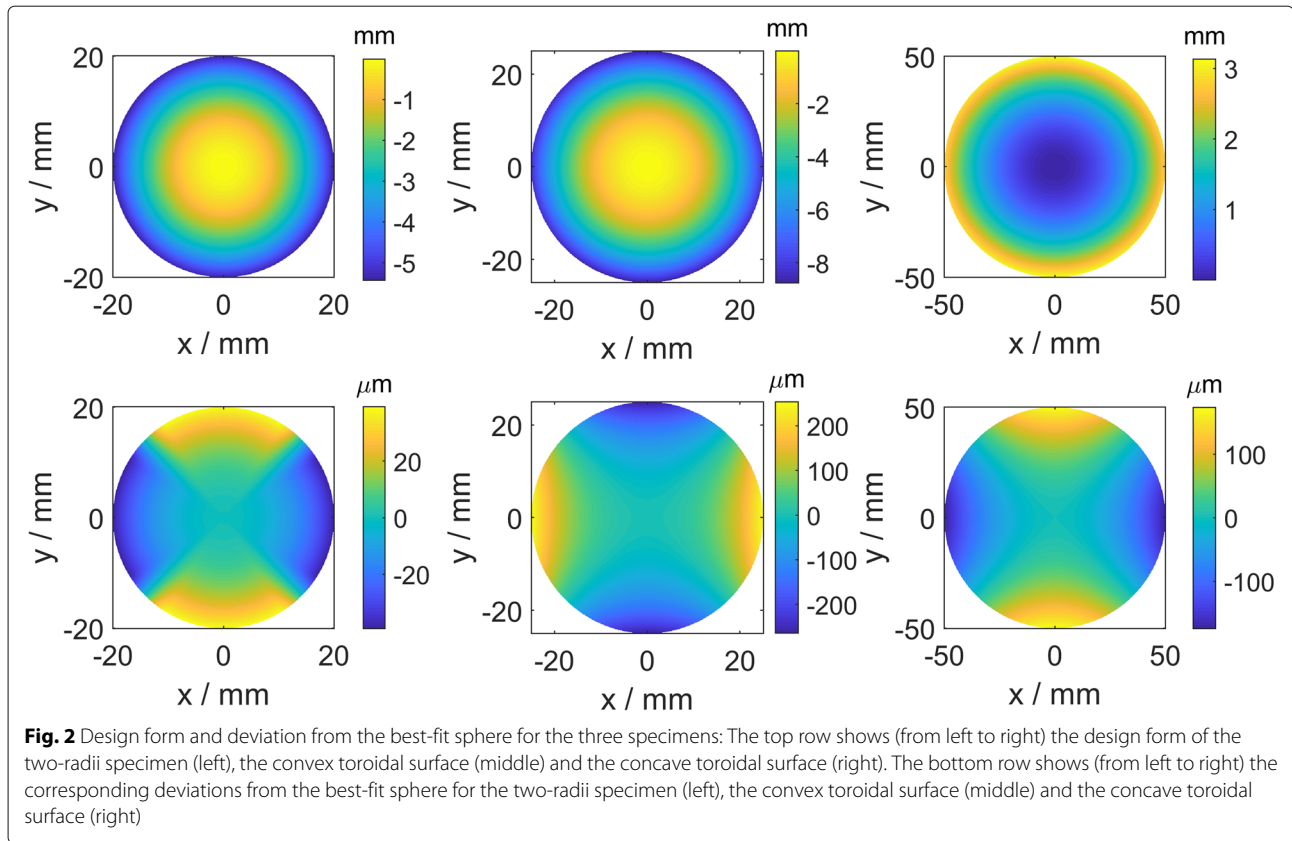
The radii are $r_v = 40$ mm and $r_h = 42$ mm, respectively. The specimen has a clear aperture of 50 mm and a best-fit radius of 40.9 mm. The peak-to-valley deviation from the best-fit sphere amounts to 518 μm (see Fig. 2, lower row, middle). Additionally, this specimen has four Gaussian peak markers of different depths of 0.5 μm, 0.75 μm, 1.0 μm and 1.25 μm, which are helpful for comparing the measurements and create a unique orientation between the different measurements.

Concave toroid This specimen has a concave toroidal surface form. It has two different radii along the two orthogonal axes. The mathematical description of the surface form is given by

$$z(x,y) = -\sqrt{(\sqrt{r_v^2 - y^2} + r_h - r_v)^2 - x^2} + r_h. \qquad (2)$$

The radii are $r_v = 400$ mm and $r_h = 450$ mm, respectively. The specimen has a clear aperture of 100 mm and a best-fit radius of 423.5 mm. The peak-to-valley deviation from the best-fit sphere amounts to 348 μm (see Fig. 2, lower row, right).

Partners and Instruments

In this section, the instruments used by the partners in this comparison study are described. We present measurements provided by the Czech Institute of Plasma Physics (IPP), the German company Mahr GmbH (Mahr), the German Institute of Technical Optics at the University of Stuttgart (USTUTT), the German national metrology institute, the Physikalisch-Technische Bundesanstalt (PTB) and the National Metrology Institute of Japan (NMIJ). Table 2 shows a list of all participating institutions, the instruments they used and which specimens were measured.

Table 1 Parameters of the round robin specimens

Property	Two-radii	Convex toroid	Concave toroid
Material	Super Invar®	Super Invar®	Super Invar®
Diameter	40 mm	50 mm	100 mm
Markers	none	4 Gaussian peaks; different depths	none
Radius of curvature	$r_1 = 39.5$ mm	$r_v = 40$ mm	$r_v = 400$ mm
	$r_2 = 40$ mm	$r_h = 42$ mm	$r_h = 450$ mm
Designed by	PTB	PTB	PTB
Manufacturer	IPP	IPP	IPP

Table 2 Partners and instruments used

Partner	Two-radii	Convex toroid	Concave toroid
IPP	MFU	MFU	MFU
Mahr	MFU, TWI	MFU, TWI	MFU, TWI
USTUTT	TWI	TWI	–
PTB	TWI	TWI	–
NMIJ	–	–	UA3P

TWI: Tilted-Wave Interferometer (product name: TWI-60), MFU: coordinate measuring machine (CMM) with optical point sensor (product name MarForm MFU 200 Aspheric 3D), UA3P: CMM with tactile probe (product name UA3P-4000)

While other partners of the European EMPIR project also performed measurements during the round robin, they are not considered in this analysis, as they measured a significantly smaller aperture. In order to make a reasonable comparison, we have to reduce all measurements to the largest common aperture.

The tilted-wave interferometer (TWI) is an optical areal measuring instrument using non-null interferometry by illuminating the surface under test (SUT) using several differently tilted wavefronts [12, 13]. In this round robin, three different setups of this measurement device were used: USTUTT hosted a laboratory breadboard setup which is flexible and can be changed or adapted to outer constraints. In this setup, the optical axis was horizontal and the specimen was mounted vertically. At PTB, the first prototype of the Mahr instrument was used. At Mahr, the current instrument version (TWI-60) was used for taking measurements. In both instruments at PTB and Mahr, the optical axis was vertical and the specimen was mounted horizontally.

The MarForm MFU 200 Aspheric 3D used by IPP and Mahr is a coordinate measuring machine (CMM) which can be equipped with a tactile probe or with an optical point sensor [8]. In this case, the optical sensor was used.

The Ultrahigh Accurate 3D Profilometer (UA3P) is a profilometer with a tactile probe. The instrument version used in this round robin is the UA3P-4000 [14].

The participating partners were allowed to follow their own measurement strategies which resulted, among other things, in different point densities and measurement times. The number of raw data points (N_{raw}) is given for each measurement in the corresponding result section. Please note that the number of raw data points does not directly correlate to the point density, since the measurement points are not equally distributed for all measurement devices.

The participating partners recorded the measurement temperature and ensured that temperature conditions were kept stable.

It should be emphasized that the instruments used are particular realizations of their type. On the accuracy scale discussed here, other instruments of the same type at other laboratories may differ in the state of their development or in their performance.

Data analysis

In this section, we describe the data analysis method we applied for evaluating the comparison.

The participants were advised to deliver the form measurement data of each specimen in Cartesian coordinates by submitting an x-, y-, z- point cloud, where the z-coordinate describes the absolute measured form of the SUT (the z-value of the design function plus the z-value describing the deviation from the design). The analysis we conducted was two-fold. In the first processing step, we handled each data set separately. Thereafter, we performed bilateral comparisons of the processed measurements obtained in the first step. The procedure for obtaining processed measurements is similar to that described in [6].

This first stage of the analysis consisted of the following steps:

- Restricting the data set to the largest common aperture of all data sets compared
- Subtracting the design function from the data
- Removing outliers
- Removing a best-fit sphere

The analysis steps are described in more detail in the following.

To subtract the design function, the measurement data were aligned with the design topography by minimizing the differences between the measurement and the design in a least-squares sense (describing the design deviations in the z-direction), thus allowing the measurement point cloud to be shifted and rotated along the three Cartesian axes. The z-values of the design were then subtracted from the measurement data. For this purpose, a functionality from the SimOptDevice software library [15] was used.

To remove measurement data points which likely correspond to outliers generated by dust particles or similar unwanted effects, we applied the following procedure. A Zernike fit up to degree 18 was fitted to the difference of the raw data and the design function. We then considered the residuals obtained after subtracting the design function and the obtained Zernike polynomial from the raw data. After removal of the median from these residuals, all residuals whose absolute value exceeded f_d times their median absolute deviation (MAD) [16] were then used to identify outliers in the raw data. The deviation factor f_d was set to 11 after close inspection of the raw data set and of points visually identified as outliers. For further processing, the difference of the raw data and the design function was taken after removal of the identified outliers.

In the next step, we removed an individual best-fit sphere from the resulting data set. To this end, a sphere was fitted to the data (describing the deviations in the *z*-direction) in a least-squares sense and then subtracted from the data. This step was performed because many instruments cannot determine the spherical component of the measured surface unambiguously, resulting in the spherical contributions dominating the processed measurements [6]. The subtraction of a best-fit sphere at this point is a common option in the evaluation of optical surfaces because, in most applications, a small spherical contribution/mismatch can be compensated in the mounting process of the complete optical system.

Although the design forms of the specimens are not rotationally symmetric and therefore belong to the group of so-called freeform surfaces, the two specimens still have symmetry via 180 degree rotation of the design around the *z*-axis, where no marker structures have been added to the design. To ensure that all data sets of these two specimens had the same orientation for further analysis, the correlation between one reference data set and all other data sets was calculated for a rotation of 0 degrees and 180 degrees around the *z*-axis. If a data set rotated by 180 degrees led to the maximum correlation, the rotated data set was chosen for further analysis.

In the second stage of the analysis, we performed bilateral comparisons between all data sets of the same specimen. After the processed measurements from step one were resampled on a regular 1000×1000 point Cartesian grid using spline interpolation, maps of pointwise differences between two processed measurements were computed. The root-mean-square (RMS) values were determined for each pairwise difference. Because the variability of the different measurement methods in this round robin was not large enough to ensure that the systematic errors inherent to the given measurement system would be reduced significantly when calculating a reference topography from all measurements, we chose to perform bilateral comparisons between all available data sets instead of deriving a reference topography from all measurements.

Results and discussion

In this section, the results of the data analysis described above are presented for the three freeform specimens investigated. All measurements shown here were recorded at room temperatures with variations of approximately 1.5 *K* between the different sites. These temperature differences mostly affected the radius components of the measurements, which were eliminated for the bilateral comparisons. Since the material of the specimens has good thermal properties, even the effect on the radius is small. For each specimen, we first present the processed measurements after subtracting the design form and an individual best-fit sphere which had been interpolated on a common grid. These measurements are shown in Fig. 3

Fig. 3 Processed measurements after subtracting the design form and an individual best-fit sphere of each data set for the two-radii specimen. The processed measurements were interpolated on a common grid (1000 pixels x 1000 pixels)
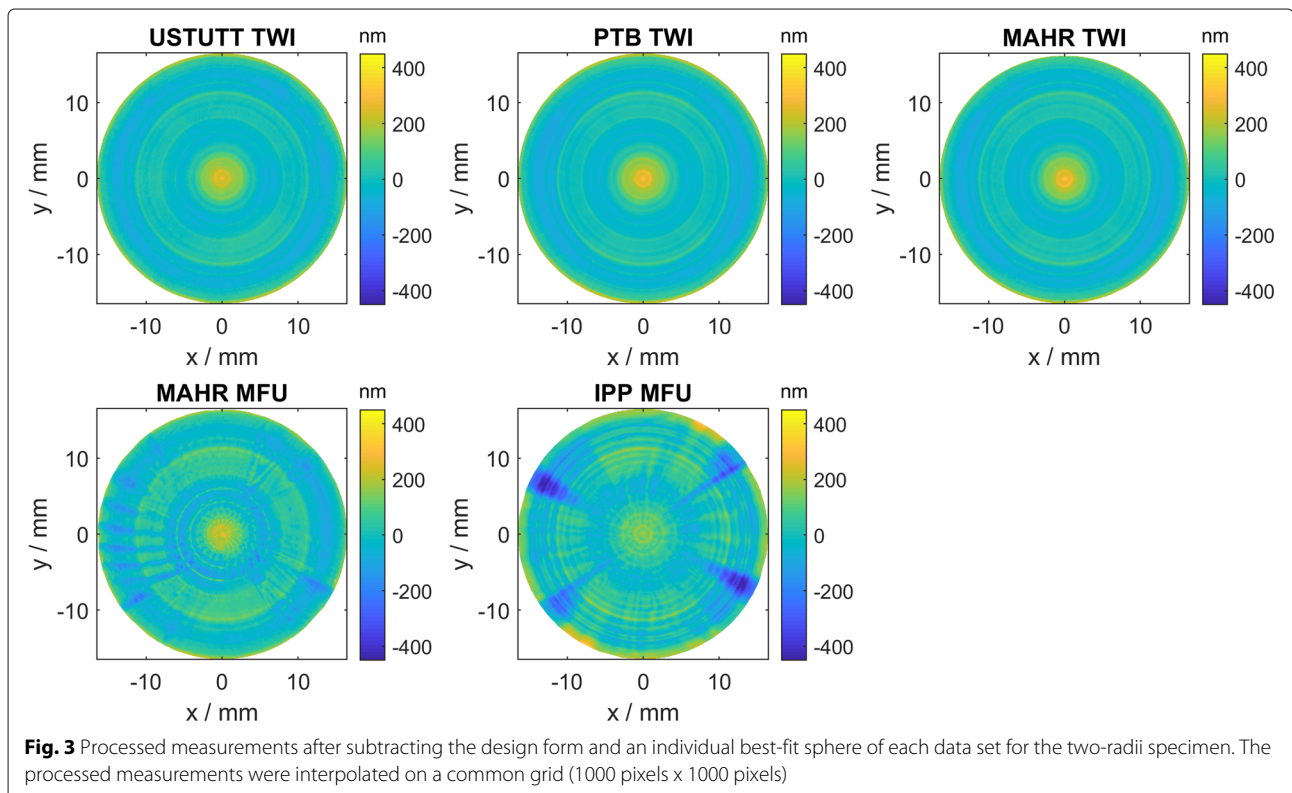
for the two-radii specimen, in Fig. 7 for the convex toroidal surface and in Fig. 11 for the concave toroidal surface.

The $x-z-$ profiles (at $y=0$) and the $y-z-$ profiles (at $x=0$) after subtracting the design form and an individual best-fit sphere are shown in Fig. 4 for the two-radii specimen, in Fig. 8 for the convex toroidal surface and in Fig. 12 for the concave toroidal surface.
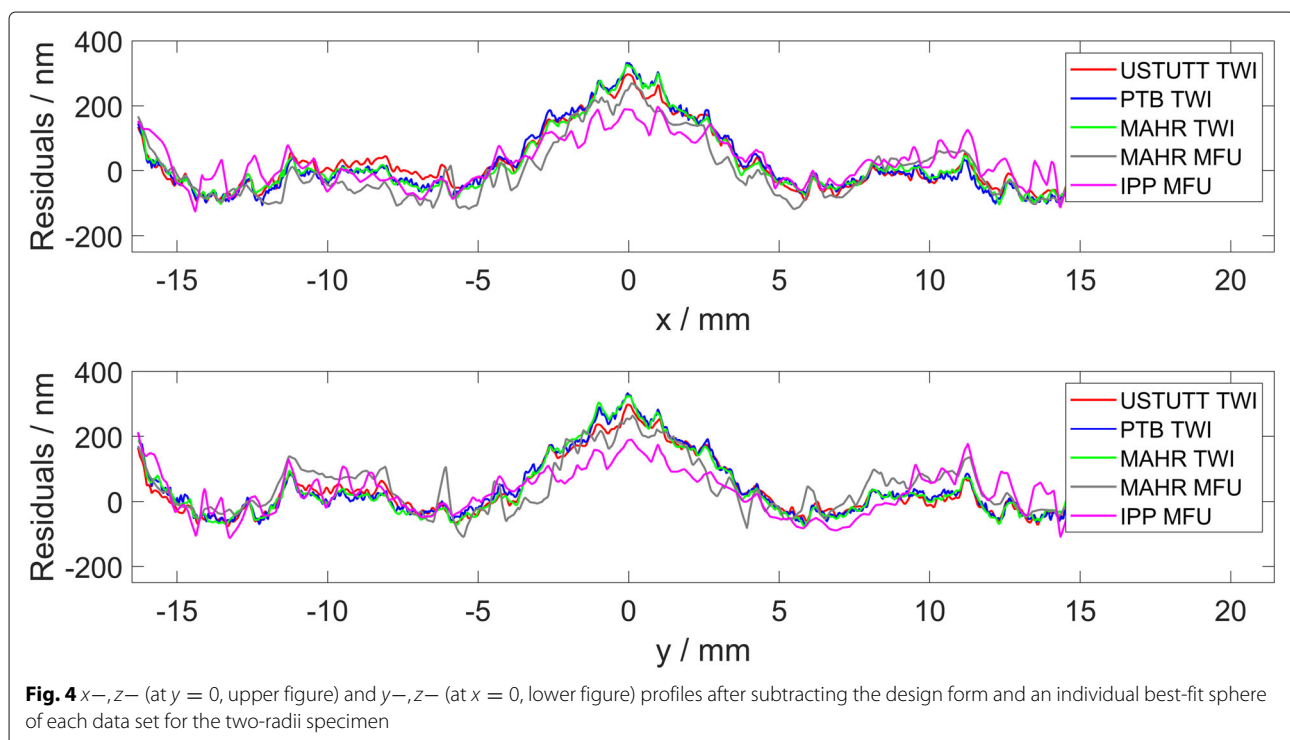
The processed measurements after subtracting the design form and an individual best-fit sphere formed the input data for the bilateral differences shown in Fig. 5 for the two-radii specimen, in Fig. 9 for the convex toroidal surface and in Fig. 13 for the concave toroidal surface.
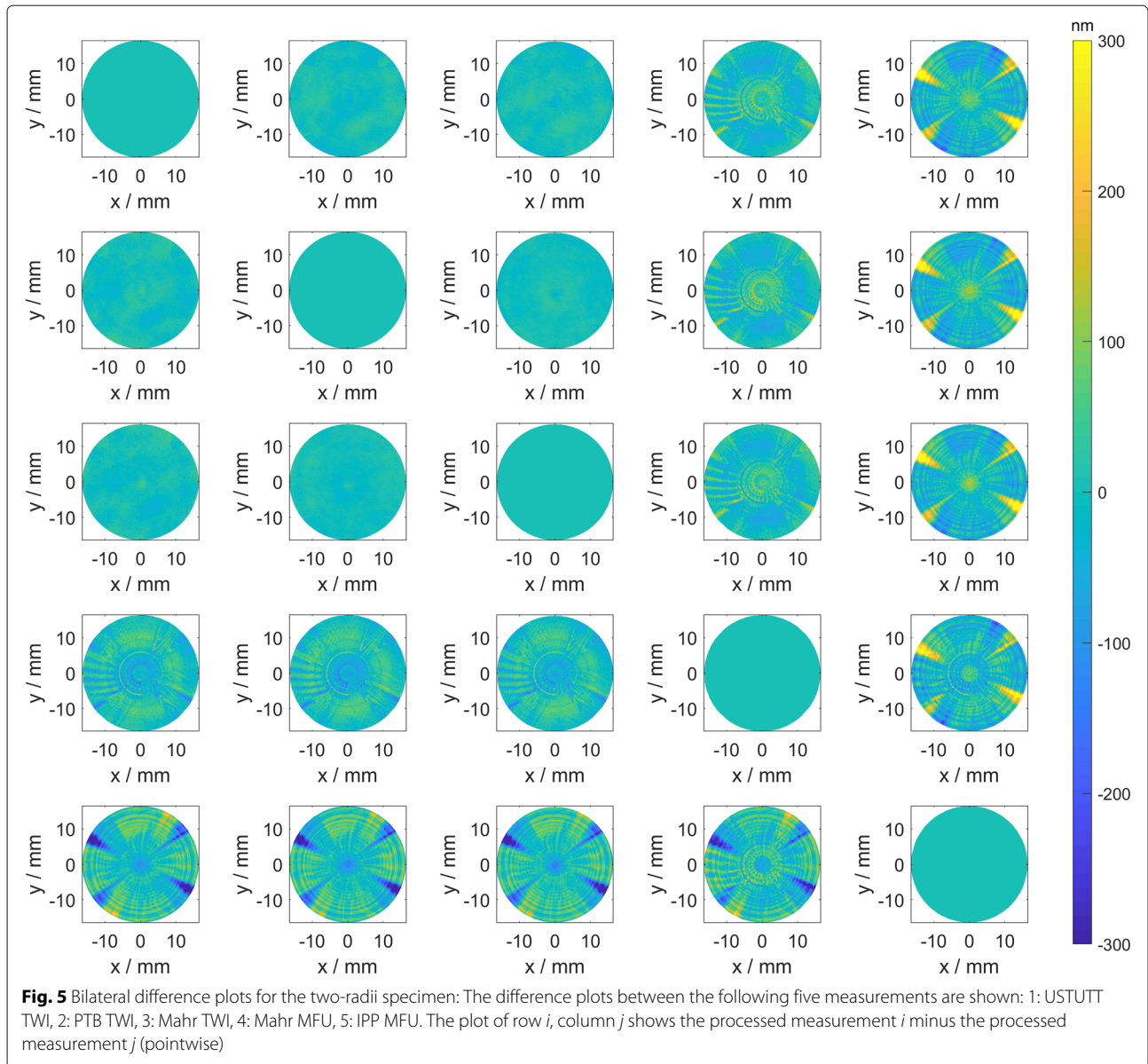
Two-radii specimen
For the two-radii specimen, the data were evaluated for a common aperture with a radius of 16.5 mm. For this measurement aperture, five data sets were available: 1: USTUTT TWI ($N_{\mathrm{raw}} = 777229$), 2: PTB TWI ($N_{\mathrm{raw}} = 470629$), 3: Mahr TWI ($N_{\mathrm{raw}} = 109225$), 4: Mahr MFU ($N_{\mathrm{raw}} = 28410$) and 5: IPP MFU ($N_{\mathrm{raw}} = 29829$). The curvatures of the subtracted best-fit spheres were 0.0056/m (USTUTT TWI), 0.0029/m (PTB TWI), 0.0035/m (Mahr TWI), 0.0027/m (Mahr MFU) and 0.0007/m (IPP MFU).

The processed measurements after subtracting the design form and an individual best-fit sphere had peak-to-valley values below 1 μm (see Fig. 3), while for the bulk of the data sets, the processed measurements had peak-to-valley values below 500 nm. The results of the

MFU at IPP have some large deviations at the cosine transition zones between the different spherical segments of the specimen located at the diagonals between the x- and y- axes (cf. design deviation from a best-fit sphere shown in Fig. 2, lower row, left). Some minor deviations in this area can also be seen in the results of the MFU at Mahr. The processed measurements of the three different TWI setups seem to be in good agreement. This is an excellent result when one considers that these three setups of the same measurement principle are very different. The good agreement between the three TWIs is also reflected in their profiles (see Fig. 4). Of course, this does not guarantee that these measurements are correct, since all TWI measurements might suffer from the same systematic errors.

The pointwise differences between the processed measurements are shown in Fig. 5 for the two-radii specimen and reinforce the similarities between the three TWI measurements. The largest deviations occur in the transition zones of the spherical segments of the specimen for the MFU measurement at IPP. Furthermore, several small high-frequency characteristic deviations between the MFU and the TWI measurements are visible. However, because these deviations are also visible in the bilateral comparison between the two MFU measurements, this seems to be more of a systematic influence of the MFU measurements than a feature of the surface which was not measured by the TWI measurements.

Fig. 4 $x-,z-$ (at $y=0$, upper figure) and $y-,z-$ (at $x=0$, lower figure) profiles after subtracting the design form and an individual best-fit sphere of each data set for the two-radii specimen

Fig. 5 Bilateral difference plots for the two-radii specimen: The difference plots between the following five measurements are shown: 1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU. The plot of row *i*, column *j* shows the processed measurement *i* minus the processed measurement *j* (pointwise)

To evaluate the bilateral deviations in a quantitative manner, the RMS values of the pointwise difference plots are illustrated in Fig. 6. Each element (row *i*, column *j*) represents the RMS value of the difference plot between partners *i* and *j* (1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU). While the RMS values between the three TWI measurements are in the range below 20 nm, the RMS values between the Mahr MFU measurement and all three TWI measurements are below 50 nm. The RMS values between the IPP MFU measurement and all other measurements are around 80 nm.

Convex toroid

For the specimen which has a convex toroidal surface form, the data were evaluated for a common aperture with a radius of 21.8 mm. For this measurement aperture, five data sets were available: 1: USTUTT TWI ($N_{raw} = 783260$), 2: PTB TWI ($N_{raw} = 427119$), 3: Mahr TWI ($N_{raw} = 124296$), 4: Mahr MFU ($N_{raw} = 40641$) and 5: IPP MFU ($N_{raw} = 30062$). Since the centres of the four Gaussian peak markers were located at ($x = 0$ mm, $y = \pm22$ mm) and ($x = \pm22$ mm, $y = 0$ mm), the markers were not completely visible in the measurement data sets. Nevertheless, since several partners measured a larger area of the specimen, and since the Gaussian markers had a full width at half maximum of 0.5 mm, at least some parts of the markers were visible for all partners, allowing the markers (or parts of them) to be used to align the measurement data sets.
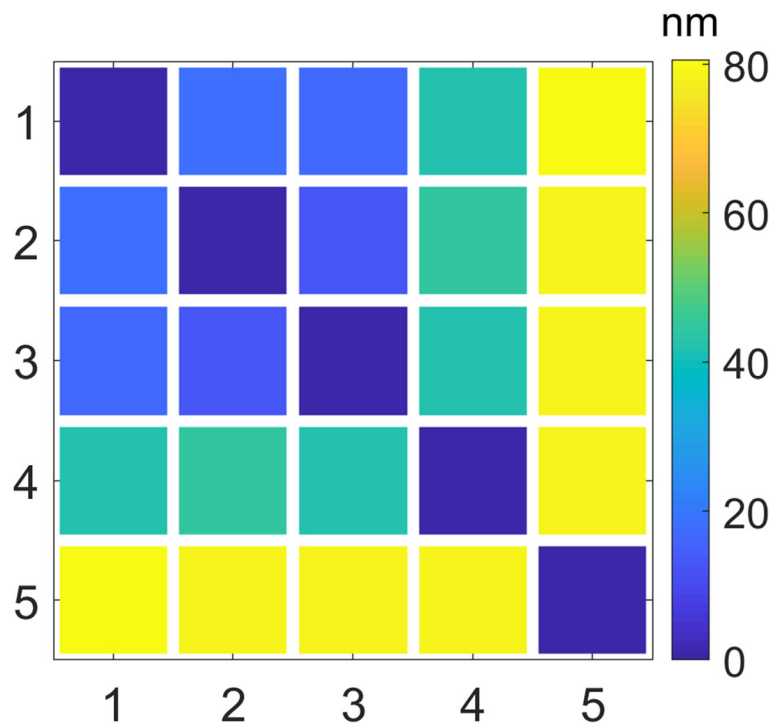
Fig. 6 RMS values of difference maps for the two-radii specimen. The RMS values of the pointwise differences between the following five measurements are shown: 1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU. The value of row *i*, column *j* shows the RMS value for the pointwise differences between measurement *i* and measurement *j*
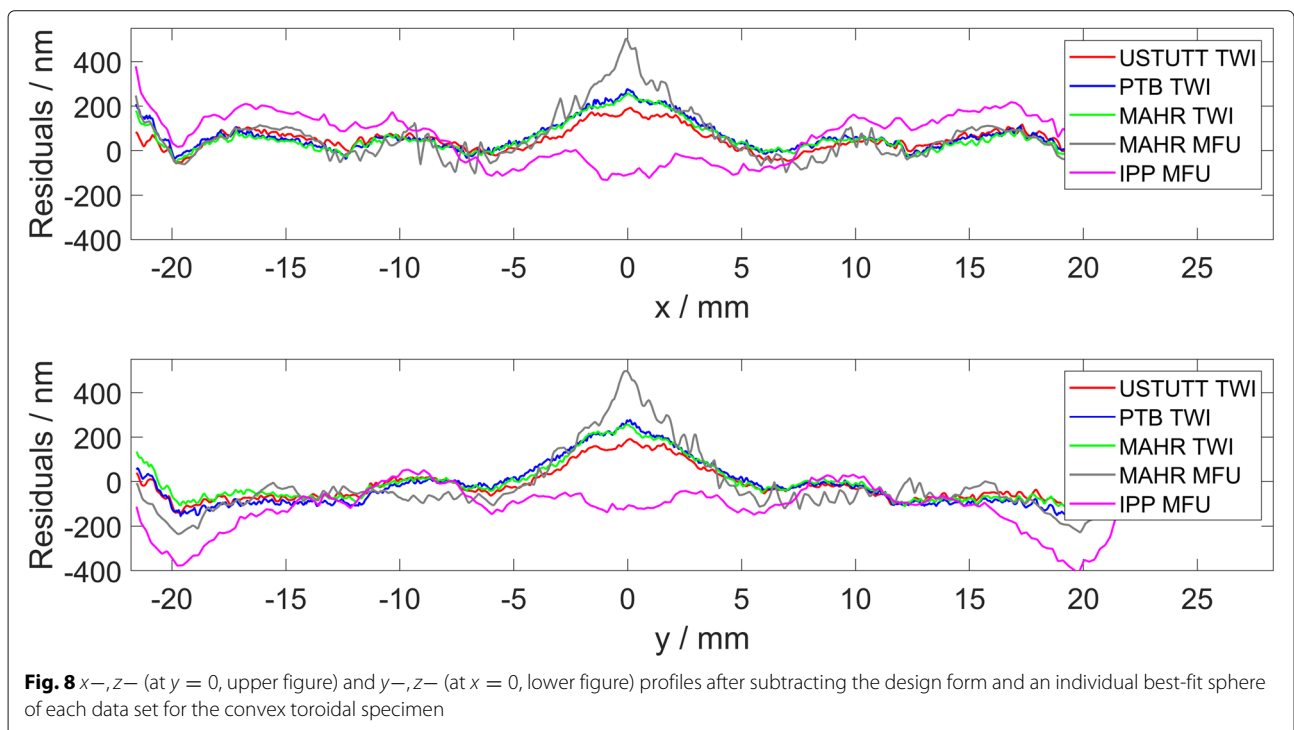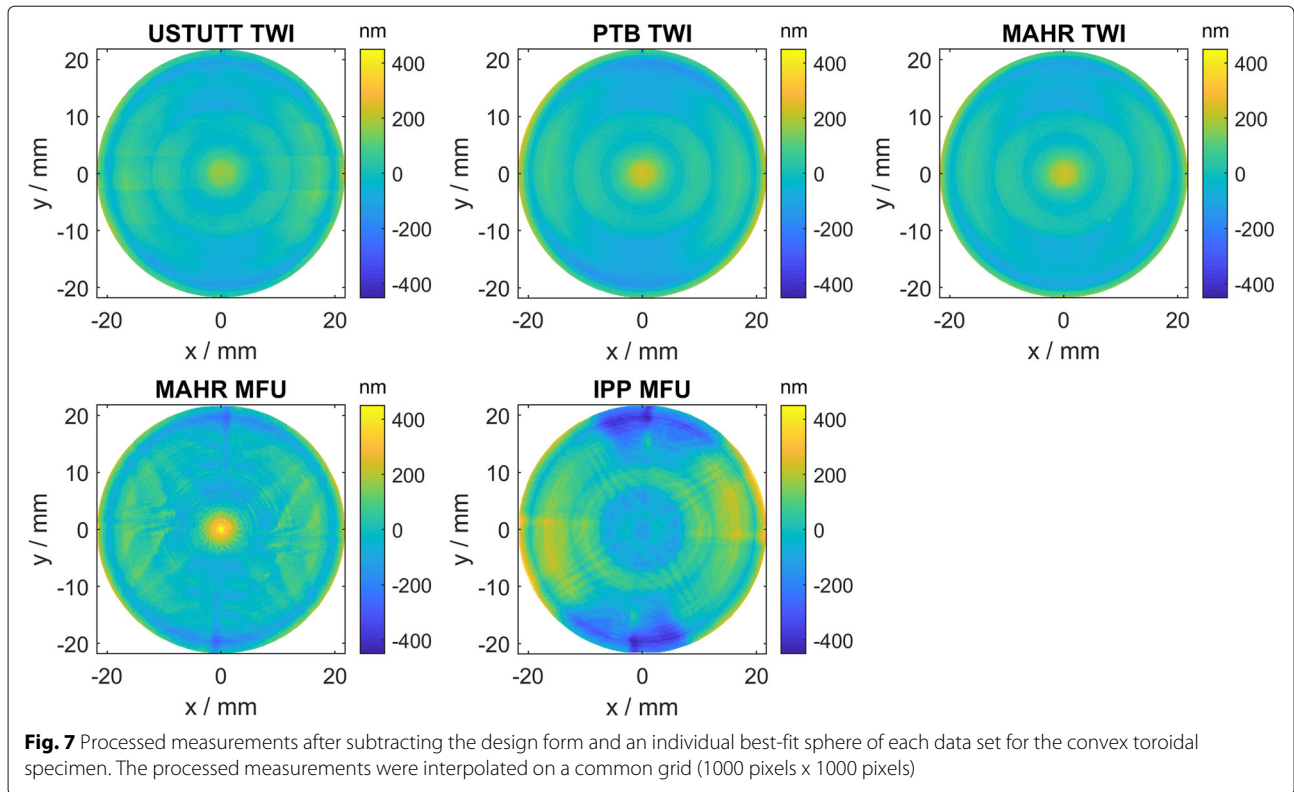
The curvatures of the subtracted best-fit spheres were 0.0005/m (USTUTT TWI), 0.0041/m (PTB TWI), 0.0038/m (Mahr TWI), 0.0013/m (Mahr MFU) and 0.0097/m (IPP MFU).

The processed measurements, after subtracting the design form and an individual best-fit sphere, have peak-to-valley values below 1 μm (see Fig. 7), while the bulk of the data sets have peak-to-valley values below 400 nm. The processed measurements show characteristic features between the different measurement instruments. The results from the MFU at IPP contain some larger deviations. A similar structure can be seen in the results from the MFU at Mahr, but to a much smaller extent. For this specimen as well, the processed measurements of the three different TWI setups seem to be in good agreement. For all processed measurements, apart from the measurement results of the MFU at IPP, there was a larger peak in the centre of the specimen. This was also the case for the two-radii specimen and may be an artifact from the manufacturing process in which SPDT was used.

The profiles of the processed measurements (see Fig. 8) show very good agreement between the results from the TWI at PTB and those from the TWI at Mahr. The measurement of the TWI at USTUTT is also close to these results. While the measurement results of the MFU at Mahr mainly differ from the TWI profiles in the centre,

the MFU results from IPP also contain larger deviations on the outer part of the specimen. Note that the two MFU measurements do not contain the same deviations from the three TWI measurements. Instead, they show deviations with opposite signs, especially in the centre of the specimen.

The pointwise differences between the processed measurements are shown in Fig. 9 for the convex toroidal surface; once again these bilateral comparisons emphasize the similarities between the three TWI measurements. Furthermore, the pointwise differences between the MFU measurement at Mahr and the three TWI measurements are much smaller than the pointwise differences between the MFU measurement at IPP and all other measurements. To evaluate this in a quantitative manner, the RMS values of the pointwise difference plots are illustrated in Fig. 10. Each element (row *i*, column *j*) represents the RMS value of the difference plot between partners *i* and *j* (1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU). While the RMS values between the three TWI measurements are around 30 nm, the RMS values between the Mahr MFU measurement and all three TWI measurements are around 50 nm. The RMS values between the IPP MFU measurement and all other measurements are around 110 nm.

Fig. 7 Processed measurements after subtracting the design form and an individual best-fit sphere of each data set for the convex toroidal specimen. The processed measurements were interpolated on a common grid (1000 pixels x 1000 pixels)

Fig. 8 $x-, z-$ (at $y = 0$, upper figure) and $y-, z-$ (at $x = 0$, lower figure) profiles after subtracting the design form and an individual best-fit sphere of each data set for the convex toroidal specimen

Fig. 9 Bilateral difference plots for the convex toroidal specimen: The difference plots between the following five measurements are shown: 1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU. The plot of row *i*, column *j* shows the processed measurement *i* minus the processed measurement *j* (pointwise)

Concave toroid

For the concave toroidal specimen, the data were evaluated for a common aperture with a radius of 34.9 mm. For this measurement aperture, four data sets were available: 1: Mahr TWI (N_{raw} = 64117), 2: Mahr MFU (N_{raw} = 137438), 3: IPP MFU (N_{raw} = 567709) and 4: NMIJ UA3P (N_{raw} = 313634).

The curvatures of the subtracted best-fit spheres were 0.0020/m (Mahr TWI), 0.0023/m (Mahr MFU), 0.0016/m (IPP MFU) and 0.0020/m (NMIJ UA3P).

The processed measurements after subtracting the design form and an individual best-fit sphere have peak-to-valley values around 500 nm (see Fig. 11) and show characteristic features between the different

measurement instruments. The results of the MFU measurements at IPP and Mahr show some structures which were also slightly visible in the measurements of the other two specimens. These structures may be calibration artifacts of the MFU which occur mainly in regions where the gradient in the azimuthal direction is high (cf. design deviation to a best-fit sphere shown in Fig. 2, lower row). For the measurement of the UA3P at NMIJ, several data points had to be masked out, since a large deviation from the design was visible in this area which was not present in the other/previous measurements. This deviation from the design may be due to damage to the specimen during transport. Therefore, we decided to mask out these data points to prevent them from contributing to
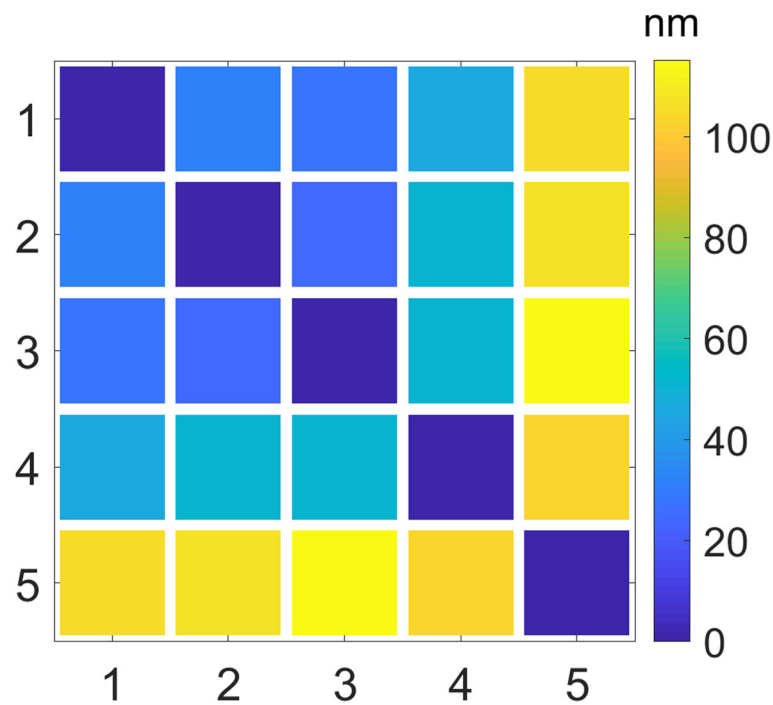
Fig. 10 RMS values of difference maps for the convex toroidal specimen. The RMS values of the pointwise differences of the following five measurements are shown: 1: USTUTT TWI, 2: PTB TWI, 3: Mahr TWI, 4: Mahr MFU, 5: IPP MFU. The value of row *i*, column *j* shows the RMS value for the pointwise differences of measurement *i* and measurement *j*
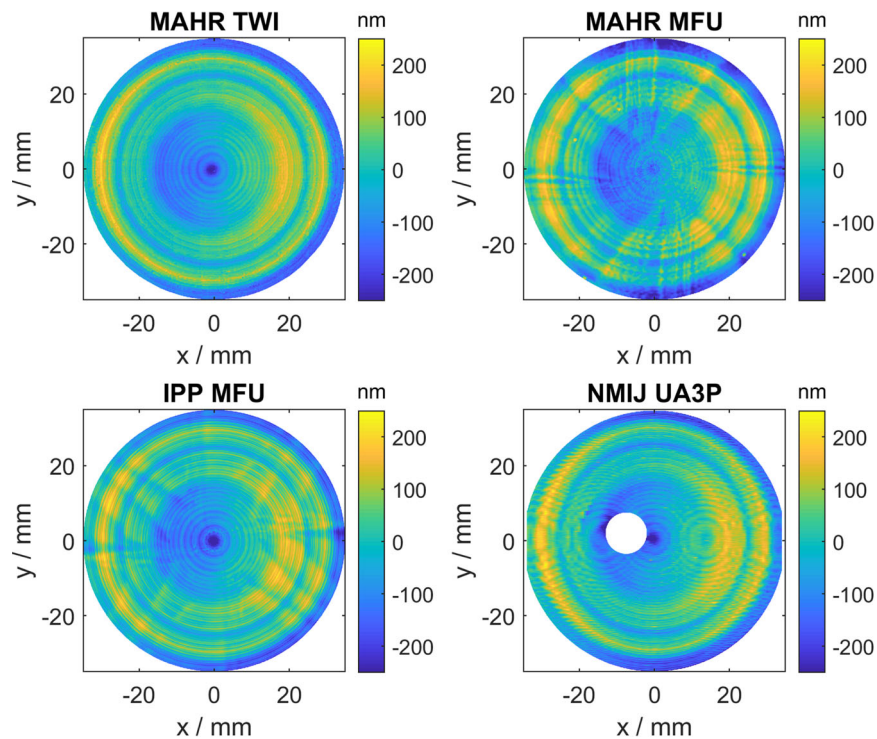

Fig. 11 Processed measurements after subtracting the design form and an individual best-fit sphere of each data set for the concave toroidal specimen. The processed measurements were interpolated on a common grid (1000 pixels x 1000 pixels). For the measurement of the UA3P at NMIJ, several data points had to be masked out, since a large deviation from the design was visible in this area which was not present in the other/previous measurements. This deviation from the design may be due to damage to the specimen during transport

the following analysis. The four processed measurements show good agreement in the basic form measured. This is also reflected in the profiles shown in Fig. 12, which also show high-frequency ripples detected by all measurement devices. These ripples may be caused by vibrations of the tool during manufacturing. In the centre of the specimen, a likely explanation for why the MFUs are closer to the other instruments is that the form of this concave toroidal surface is much flatter (see radii of curvature in Table 1) compared to the other two specimens.

The pointwise differences between the processed measurements are shown in Fig. 13 for the concave toroidal surface. All pointwise differences are in the same order of magnitude. To evaluate this in a quantitative manner, the RMS values of the pointwise difference plots are illustrated in Fig. 14. Each element (row i, column j) represents the RMS value of the difference plot between partners i and j (1: Mahr TWI, 2: Mahr MFU, 3: IPP MFU and 4: NMIJ UA3P). The RMS values between all bilateral difference measurements range from 43 nm to 53 nm. The smallest RMS value has the pointwise difference between the measurement of the TWI at Mahr and the MFU at IPP.
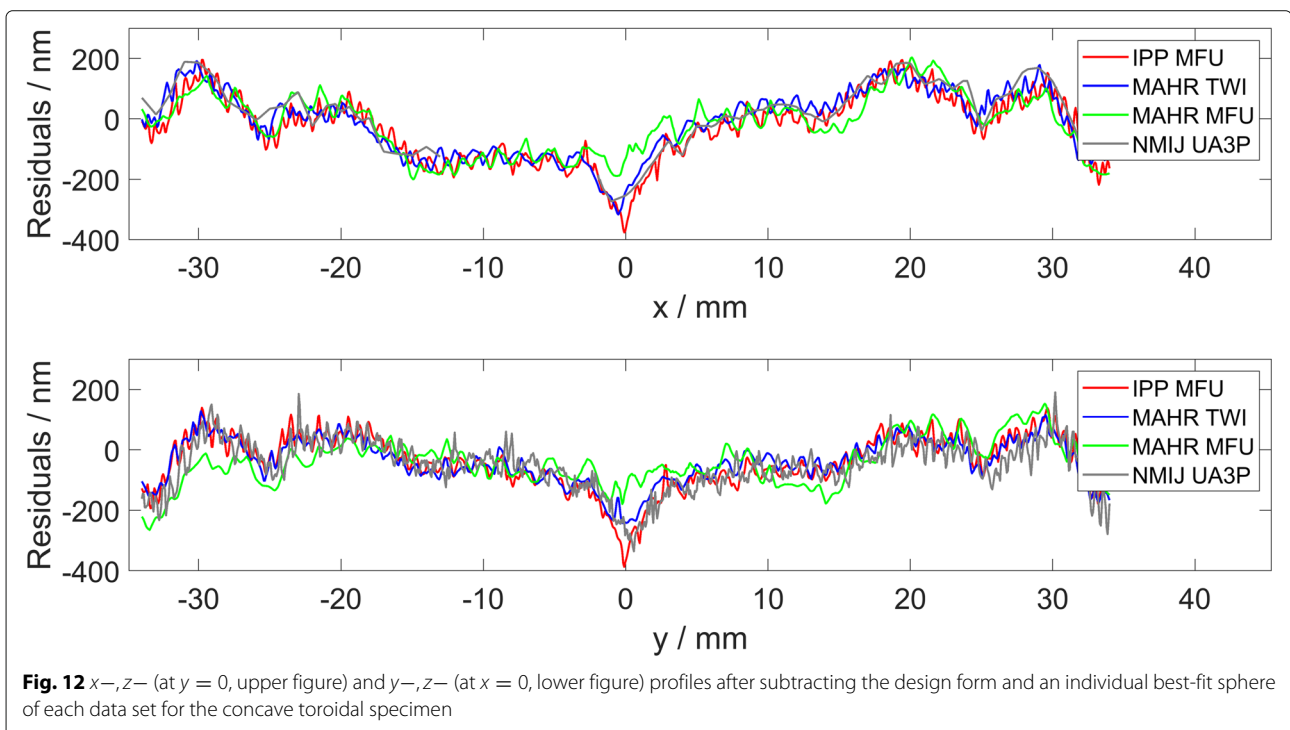
Conclusions

In this paper, we have presented measurement results obtained by means of optical and tactile form measuring instruments for three different freeform surfaces manufactured from the material Super Invar®, as well as bilateral comparisons of these results. Two specimens

(a two-radii specimen and a convex toroidal surface) have been measured with three different TWI setups at USTUTT, PTB and Mahr and with two commercially available MFU instruments at Mahr and at IPP. The results obtained via the three TWIs show very good agreement. The bilateral difference data have RMS values below 20 nm for the two-radii specimen and below 30 nm for the convex toroidal surface (neglecting spherical contributions). The differences between the TWIs and the Mahr MFU have RMS values around 50 nm, while for the IPP MFU, the RMS values amount to about 100 nm. Differences can be observed between the two MFU instruments. One possible reason for these differences may be different operation of the instrument, differently chosen measurement strategies or different calibration routines. The concave toroidal surface was measured with the commercially available TWI setup at Mahr, with the two MFU instruments at Mahr and IPP and with a tactile UA3P at NMIJ. The comparisons of the results of this specimen show that the RMS values of the bilateral differences are around 50 nm.

The results presented in this paper demonstrate that all instruments used in this round robin can be used to measure complex freefrom surfaces of different kinds. The differences between the results obtained using the different measuring principles are less than 110 nm peak-to-valley (neglecting spherical contributions).

The curvatures of the individually subtracted best-fit spheres show some deviations. Therefore, the differences when comparing the absolute form are expected to be
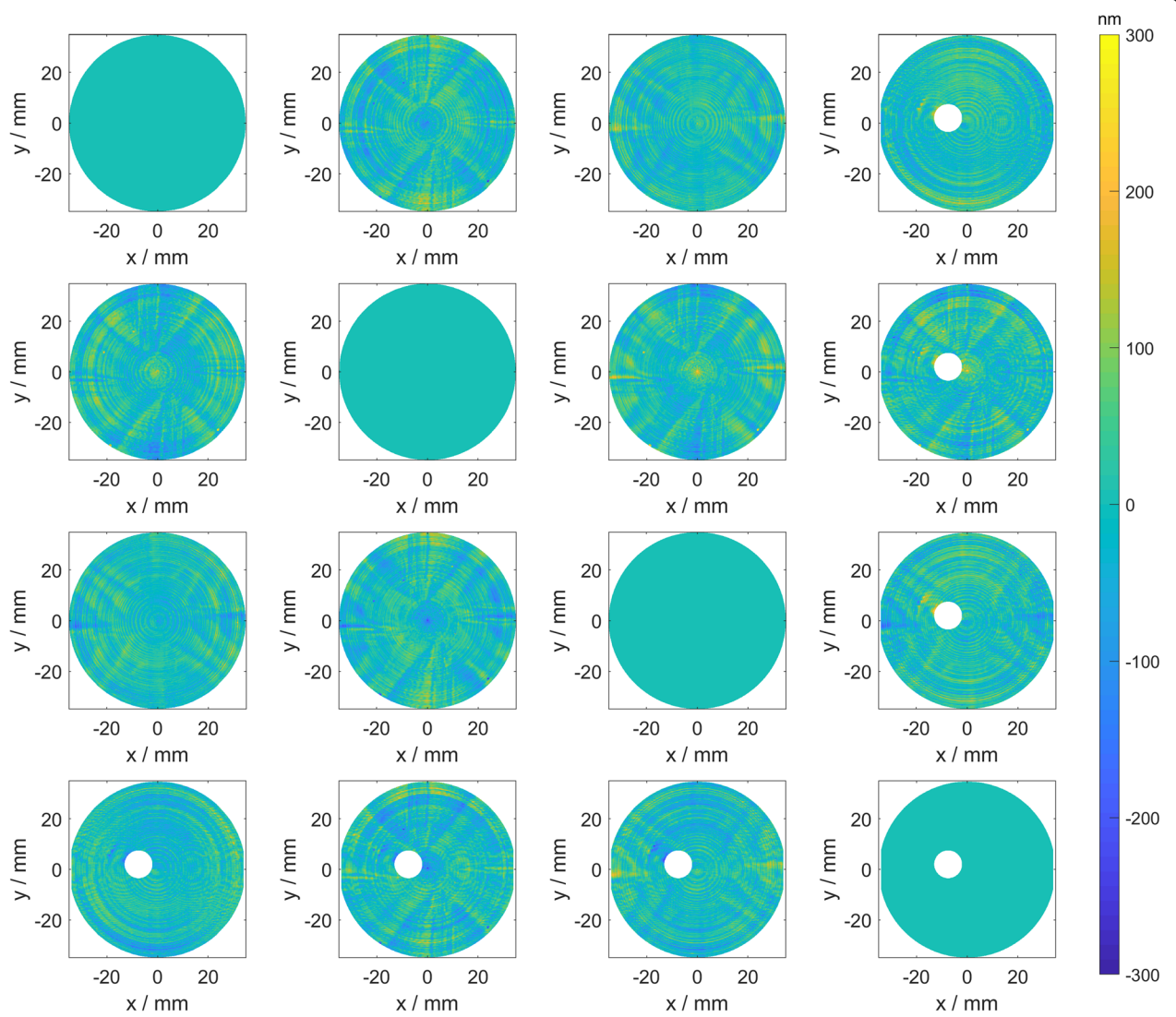

Fig. 12 $x-$, $z-$ (at $y = 0$, upper figure) and $y-$, $z-$ (at $x = 0$, lower figure) profiles after subtracting the design form and an individual best-fit sphere of each data set for the concave toroidal specimen

Fig. 13 Bilateral difference plots for the concave toroidal specimen: The difference plots between the following four measurements are shown: 1: Mahr TWI, 2: Mahr MFU, 3: IPP MFU, 4: NMIJ UA3P. The plot of row *i*, column *j* shows the processed measurement *i* minus the processed measurement *j* (pointwise)
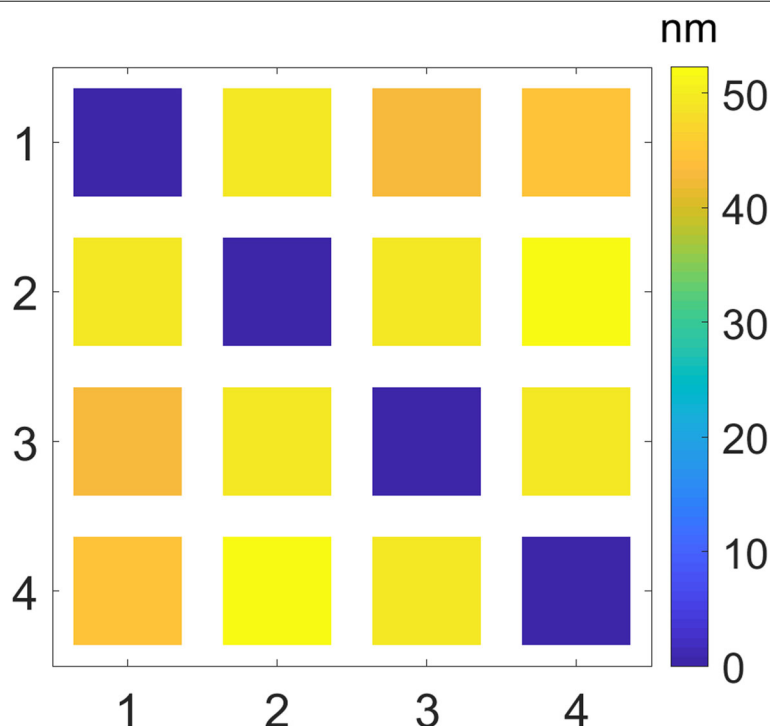
Fig. 14 RMS values of difference maps for the concave toroidal specimen. The RMS values of the pointwise differences of the following four measurements are shown: 1: Mahr TWI, 2: Mahr MFU, 3: IPP MFU, 4: NMIJ UA3P. The value of row *i*, column *j* shows the RMS value for the pointwise differences of measurement *i* and measurement *j*

significantly larger than the differences after subtracting the best-fit spheres.

In this comparison study, we have chosen to perform bilateral comparisons between all available data sets instead of deriving a reference topography from all measurements. The reason for this is that the variability of the different measurement methods in this round robin was not large enough to ensure that systematic errors inherent to a given measurement system would be reduced significantly when calculating a reference topography from all measurements.

Finally, the specimens manufactured and used in this round robin have two special features: First, the freeform surfaces are made from a very temperature-stable material in order to avoid significant temperature effects when performing comparison measurements at different sites. Second, the freeform specimens will be made available to institutions interested in investigating their measuring capabilities and comparing their measurement results to the results presented in this paper.

Abbreviations
EMPIR: European Metrology Programme for Innovation and Research; IPP: Institute of Plasma Physics; NMI: National metrology institute; MAD: Median absolute deviation; NMIJ: National Metrology Institute of Japan; PTB: Physikalisch-Technische Bundesanstalt; RMS: Root-mean-square; SPDT: Single-point diamond turning; SUT: Surface under test; TWI: Tilted-wave interferometer; UA3P: Ultrahigh Accurate 3-D Profilometer; USTUTT: University of Stuttgart

Authors' contributions
IF designed the freeform specimens, performed the measurements with the TWI at PTB, evaluated the data, discussed the results and wrote the manuscript. RS developed the data evaluation algorithms, evaluated and discussed the data and contributed to the manuscript. OM performed the measurements with the MarForm MFU 200 Aspheric 3D at IPP. VL contributed to the design, manufacturing and measurements of the freeform specimens at IPP. JS performed the freeform measurements with the TWI at Mahr. AH prepared the TWI at USTUTT for performing the measurements and postprocessed the TWI data of USTUTT. RB performed the measurements with the TWI at USTUTT. YB was involved in the discussion of the evaluation and results of NMIJ and contributed to the manuscript. YK performed measurements with device UA3P at NMIJ. MS was involved in the discussion of the evaluation and results and contributed to the manuscript. CE suggested the format of bilateral comparisons, supported the data analysis, discussed the results and contributed to the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The data sets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 38116 Braunschweig, Germany. [2] Physikalisch-Technische Bundesanstalt (PTB), Abbestrasse 2-12, 10587 Berlin, Germany. [3] Regional Centre for Special Optics and Optoelectronic Systems (TOPTEC) Institute of Plasma Physics of the Czech Academy of Sciences, Za Slovankou 1782/3, 182 00 Praha 8, Czech Republic. [4] Mahr GmbH, Carl-Zeiss-Promenade 10, 07745 Jena, Germany. [5] Institut für Technische Optik, Universtiät Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany. [6] National Institute of Advanced Industrial Science and Technology (AIST), National Metrology Institute of Japan (NMIJ), Tsukuba, 305-8563 Ibaraki, Japan.

References
1. Pfeffer, M.M.: Aspheric optics: from design to manufacturing and aspheric metrology. Adv. Opt. Technol. **5**(3), 199–200 (2016)
2. Lee, W.B., Cheung, B.C.F.: Surface Generation in Ultra-precision Diamond Turning: Modelling and Practices. John Wiley & Sons, Inc. (2003)
3. Saraswathamma, K.: Magnetorheological Finishing: A review. Int. Conf. Adv. Mech. Sci. 2014 Spec. Issue-2, 168–173 (2014). https://doi.org/10.14741/Ijcet/Spl.2.2014.30
4. Schaefer, D.: Basics of ion beam figuring and challenges for real optics treatment. Proc. SPIE. **10829** (2018). https://doi.org/10.1117/12.2318572
5. Kelm, A., Boerret, R., Sinzinger, S.: Modeling of the polishing process for aspheric optics. Proc. SPIE. **7102** (2008). https://doi.org/10.1117/12.796388
6. Schachtschneider, R., Fortmeier, I., Stavridis, M., Asfour, J., Berger, G., Bergmann, R.B., Beutler, A., Blümel, T., Klawitter, H., Kubo, K., Liebl, J., Löffler, F., Meeß, R., Pruss, C., Ramm, D., Sandner, M., Schneider, G., Wendel, M., Widdershoven, I., Schulz, M., Elster, C.: Interlaboratory comparison measurements of aspheres. Meas. Sci. Technol. **29**(5), 055010 (2018). https://doi.org/10.1088/1361-6501/aaae96
7. Bergmans, R.H., Nieuwenkamp, H.J., Kok, G.J.P., Blobel, G., Nouira, H., Küng, A., Baas, M., Tevoert, M., Baer, G., Stuerwald, S.: Comparison of asphere measurements by tactile and optical metrological instruments. Meas. Sci. Technol. **26**(10), 105004 (2015). https://doi.org/10.1088/0957-0233/26/10/105004
8. Beutler, A.: Comparison of 2D and 3D measurements of aspheres with a tactile and optical sensor on one measuring instrument. In: Classical Optics 2014, p. OTu4A.1. Optical Society of America, (2014). http://www.osapublishing.org/abstract.cfm?URI=OFT-2014-OTu4A.1. Accessed 26 June 2019
9. Beutler, A.: Metrology for the production process of aspheric lenses. Adv. Opt. Technol. **5**(3), 211–228 (2016). https://doi.org/10.1515/aot-2016-0011
10. EMPIR project 15SIB01: FreeFORM (2016). https://www.ptb.de/emrp/freeform-home.html. Accessed 26 June 2019
11. Fortmeier, I., Schulz, M., Meeß, R.: Traceability of form measurements of freeform surfaces: metrological reference surfaces. Opt. Eng. **58**(9), 092602 (2019). https://doi.org/10.1117/1.OE.58.9.092602
12. Garbusi, E., Pruss, C., Osten, W.: Interferometer for precise and flexible asphere testing. Opt. Lett. **33**(24), 2973–2975 (2008)
13. Baer, G., Schindler, J., Pruss, C., Siepmann, J., Osten, W.: Calibration of a non-null test interferometer for the measurement of aspheres and free-form surfaces. Opt. Express. **22**(25), 31200–31211 (2014). https://doi.org/10.1364/OE.22.031200
14. Panasonic (2019). https://www.panasonic.com/global/business/ua3p.html. Accessed 26 June 2019
15. Schachtschneider, R., Stavridis, M., Fortmeier, I., Schulz, M., Elster, C.: SimOptDevice: a library for virtual optical experiments. J. Sensors Sensor Syst. **8**(1), 105–110 (2019). https://doi.org/10.5194/jsss-8-105-2019
16. Huber, P.J.: Robust Statistics. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg, (2011). https://doi.org/10.1007/978-3-642-04898-2_594

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.