17NRM03 EUCoM

Evaluating Uncertainty in Coordinate Measurement

VALIDATION REPORT DESCRIBING RESULTS OF UNCERTAINTY EVALUATIONS (AGAINST GUM) OF THE DEVELOPED METHODS (A POSTERIORI AND A PRIORI)

DELIVERABLES D3, D4

Lead partner: PTB

Josef Frese, Ulrich Neuschaefer-Rube, Markus Bartscher, Alessandro Balsamo

and all other project partners

Deliverable Due Date: **January 2021**

Actual Submission Date: 2022-02-01

This document combines the Deliverables D3 (on prismatic geometry) and D4 (on freeform) in a single comprehensive report. It has been edited accordingly without changes of the content.

Table of Contents

1 Abstract

This is the validation report for the uncertainty evaluation methods developed over the course of the EUCoM project. A set of workpieces was selected and measured by project participants to generate experimental data. Using the data or related information, measurement uncertainties were estimated according to EUCoM methods A or B. Measurement results and associated uncertainties were compared to reference measurements and uncertainties obtained by other, already established methods.

2 Introduction

2.1 Project background

EUCoM – Evaluating Uncertainties in Coordinate Measurement – is an EMPIR/Euramet-supported project to develop new methods for estimating the uncertainties of tactile measurements. There are two basic approaches:

- **A posteriori** (Method A): Determine uncertainties using experimental data from repeated measurements in four different orientations. A length and a sphere standard must also be measured.

- **A priori** (Method B): Two approaches relying on, for example, expert knowledge and performance characteristics of or prior experience with the CMM (coordinate measuring machine) being used.

In essence, method A is an empirical approach to measurement uncertainty. While it requires a great deal more measurement work to be done, it does not depend on any modelling of the particular measurement system used to obtain it. Method B is the exact opposite, requiring no data other than information that would usually be available from prior use of the system. A detailed explanation of these methods can be found in deliverable reports D1 (method A) [1] and D2 (method B). [2]

2.2 Objectives

The project is split into several work packages. Work package 1 and 2 were dedicated to developing the two respective methods. Work package 3, which is the topic of this deliverable report, deals with the validation of the two methods, specifically:

- Selection of suitable artefacts with prismatic or freeform geometries
- Measurement of the artefacts by all project partners to obtain data for the validation
- Data collection
- Data evaluation according to methods A and B
- Evaluation of the results to validate both methods

3 Artefact selection and measurement strategy

Eight artefacts were originally selected for validation measurements. Due to delays and other problems arising from the COVID-19 pandemic, only six were ultimately circulated for measurement. This chapter introduces the artefacts and their measurands or features.

3.1 Selection criteria

As a group, artefacts had to meet a set of requirements to ensure their suitability for method validation:

Geometry	At least one prismatic and one freeform artefact *Method validation for both prismatic parts and freeforms*
Availability	Each workpiece must remain available for the duration of the project
Shipping	Workpieces should be conducive to easy shipping, e.g., not too heavy or awkward to handle during shipping or customs inspection, robust and stable, … *Each artefact would have to be sent to several partners across international (customs-) borders, over a long period of time*
Surface quality	Different quality artefacts should be used *Include "industrial grade" workpieces and avoid using only high quality/small tolerance test artefacts*
Reference data	Current reference or calibration data must be available for each measurand *Validation requires independent (uncertainty) values*
Dimensions	Artefacts should cover an approximate size range (100 – 500) mm *Some uncertainties depend on size of artefacts or measurement volume*
Features	Together the selected objects should include at least five distinct types of feature or measurand (e.g.: length, angle, parallelism, roundness, …). Individual parts should include datum features. The targeted features must be relevant to the task (prismatic or freeform). The targeted features must be accessible and measurable using CMMs of all partners. *Validate methods for as many different types of features/measurands as possible to demonstrate their general applicability*

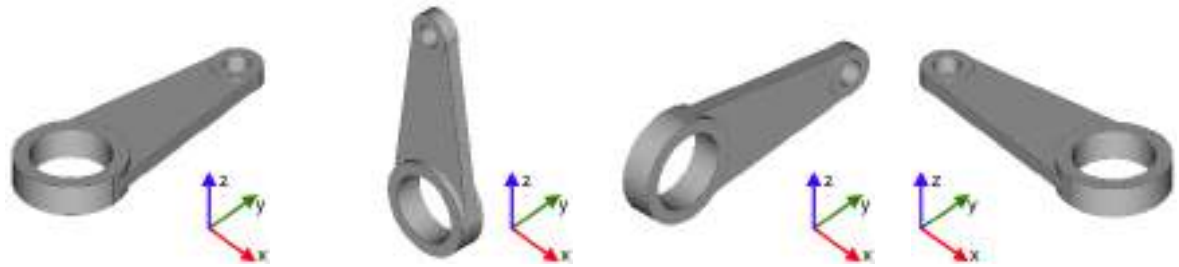
3.2 General measurement strategy

The basic measurement strategy is the same for all artefacts. A reference coordinate system was defined in which the artefact is registered using suitable features on the part itself. Each target measurand was acquired by collecting individual points or scanning across the surface using one or more styli, as necessary.

Method A requires the artefact to be measured in four distinct orientations, which were defined as part of the measurement strategy. This was done to ensure consistent results and is illustrated in **Figure 3-1**. Note that the re-oriented artefacts were also re-registered, keeping the reference frame the same for all measurements and orientations. (The graphic only shows the position 1 coordinate system). Each measurement (in each orientation) was repeated up to five times.

Method B does not use measurement data but system specifications and similar information to derive estimates. Examples include MPE-values, size of the measurement volume and, for method B1, a host of other estimated parameters. Method B2 relies primarily on results from the standardised length error measurement *E* from ISO 10360-2. [3]

For each artefact, a reference or calibration measurement was carried out by one of the project partners to provide a ground truth for the target measurands. The measurement uncertainties of these reference measurements were determined using already established methods to act as an independent estimate to which the EUCoM methods could be compared. For the final evaluation, measured values were compared to these reference values. Uncertainties of the measured values were determined using EUCoM methods and compared to the reference uncertainties and to each other.



Position 1 (start) Position 2 (x+90°) Position 3 (y+90°) Position 4 (z+90°)

Figure 3-1: Illustration of the four orientations from method A.

3.3 The artefacts

3.3.1 Connecting rod

A connecting rod manufactured for a car was selected as one of the prismatic artefacts. (**Figure 3-2**) It is an industrial workpiece roughly 185 mm × 75 mm × 20 mm in size. The cylinders (eyes) are 50 mm and 20 mm in diameter. Target measurands are cylinder and circle diameters, distance and axis parallelism between the cylinders.

Figure 3-2: The connecting rod.

During the final data evaluation for method A problems with the measurands became apparent. In brief, all measurands except distance between the two eyes displayed significant deviations and spread. This in turn affected the comparison results and validation. (Refer to chapter 6.1 for the detailed results). An investigation revealed several problems or potential issues relating to this workpiece.

Design: The expected and actual design of the connecting rod differed, leading to two problems, illustrated in **Figure 3-3**.

First, the part is not monolithic: the large eye consists of two segments joined together. Changes to the joint due to creep, clamping, general handling, etc. will affect the cylinder. Re-verification showed a 3 µm change in diameter after the round robin. A shift of the two segments along the cylinder axis would introduce a step tilt of the datum plane. The severity would depend on where the two segments are sampled to generate the datum. This in turn would affect the coordinate system and introduce cosine errors or affect the sampling height for the cylinders, particularly, the small one, which is farthest from the origin.

The second problem was that the eyes were not cylindrical. The walls follow specific profiles. Hence, the diameter changes depending on height and sampling location. Unknown at the time, proper measurement as done by the manufacturer would require specialised equipment and strategies. The strategy used assumed regular cylinders and sampled them along three circles at pre-defined heights relative to the datum.

Software: One concern is the implementation and evaluation of measurands. The majority – not all – of measurements were done with CMMs running two software suites widely available on the market. Not all participants were able to use the pre-prepared measurement programmes, e.g., due to different software or software versions being used, which may result in discrepancies between partners. Another potential error source is missing, inaccessible or incompatible settings and parameters, which were not or could not be transferred when programmes from one version or language were translated to another. This may have been particularly severe in the case of the connecting rod because delicate evaluations were involved. However, given

that other artefacts don't have similar issues, the software itself can probably be ruled out as a problem.

Clamping: Method A has somewhat unusual requirements for mounting the workpiece, which must be done in four orientations while all features remain accessible. Participants had to make their own arrangements subject to their system and the available means. Hence, the rod may have been subject to and potentially distorted by different clamping forces in each measurement and orientation. The observed drift may also have been caused by an unsuitable fixture. In at least one case, clamps pressed down on the small segment of the large eye and could have introduced a shift along the cylinder axis. (Step offset in **Figure 3-3**)

Temperature: Temperature as a factor was ruled out. Even with the largest reported temperature deviations, thermal expansion was negligible compared to the observed variations.

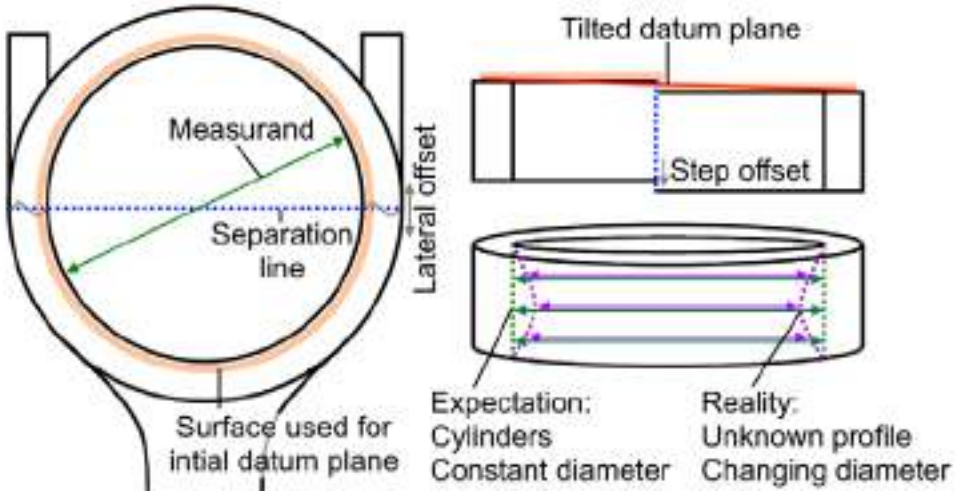


Figure 3-3: Schematic diagram of the connecting rod, the unexpected design deviations, and their consequences.

3.3.2 Multi-feature check

The prismatic multi-feature check (MFC) is designed as a complex test piece for CMMs (200 mm × 100 mm, length × diameter). (**Figure 3-4**) It includes a large variety of features and measurands and is manufactured to high specifications. Two MFCs were used, one being an older example which was expected to show some wear. Hence the two pieces were considered separately and labelled “HQ” and “LQ” for “high” and “low quality”, respectively. Measuring an MFC in full using all features would have been far too time-consuming, hence a subset of the available features was chosen for the validation. (**Figure 3-5**, **Table 3-1**)

Figure 3-4: A multi-feature check being measured.

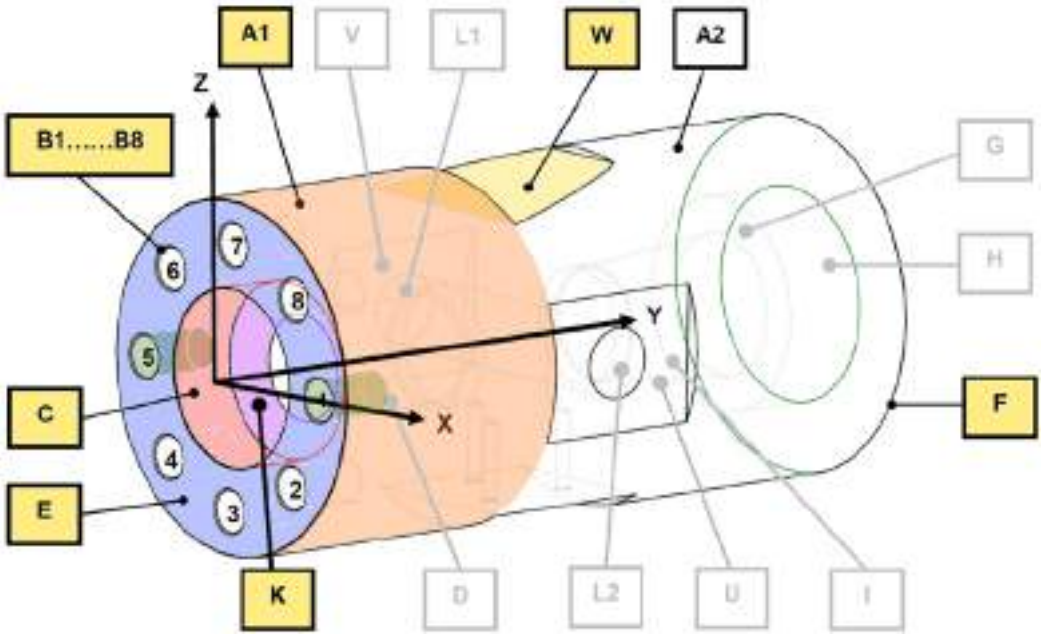


Figure 3-5: MFC diagram highlighting the selected features. (Original diagram by eumetron)

Table 3-1: Target measurands for the MFC.

Name	Geometry	Measurands
A1	External cylinder	surface straightness, roundness, cylindricity, diameter
B1, B5	Two internal cylinders	distance between bores
C	Internal cylinder	diameter and form, perpendicularity to E, concentricity, radial and total radial runout to A1
E, F	Two planes	flatness, distance, parallelism
K	Internal cone	diameter, cone angle, parallelism to A1
W	Inclined plane	Angularity to A1

3.3.3 Hyperbolic paraboloid

The hyperbolic paraboloid (HP, **Figure 3-6**) is a freeform artefact designed by CMI, one of the project partners. The artefact (100 mm × 100 mm × 60 mm) is made from titanium but includes steel reference spheres for registration. The “freeform” part is the hyperbolic paraboloid surface on the top of the circular socket. The surface is defined by mathematical model. [4, 5] The freeform surface is sampled in a uniform grid of 52 points. The points are evaluated individually by calculating the 3D distance of the measured point to the nominal surface along the normal of the nominal surface. These point deviations are reported and compared to the corresponding reference values.

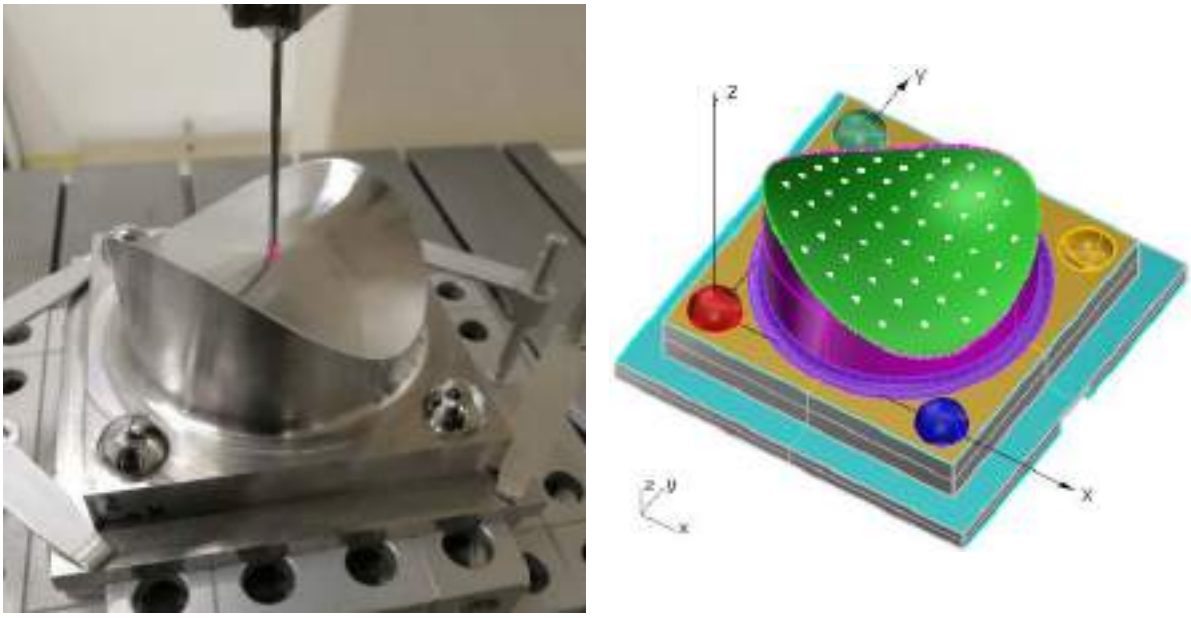


Figure 3-6: The hyperbolic paraboloid *[4, 5]* and the point grid to be measured.

3.3.4 Involute gears

The involute gears are two 10 mm steel plates with an outside involute profile about 370 mm long. On one of the gears a pattern of three superimposed sinewaves has been imprinted on the otherwise smooth involute. (**Table 3-2**) The involutes were circulated together.

The involutes are measured by scanning along the profile. The profile is interpolated and roughly 1000 points are compared to the nominal profile. The profiles are evaluated in 2D only, along x and y, z being along the plate height. Profile deviations are calculated by interpolating the measured profile and calculating the normal distance of the nominal/reference points to this profile. (**Figure 3-7** and **Figure 3-8**) This approach is often used in industry to evaluate profiles. To obtain the reference profile deviations, deviations from three independent datasets were combined and averaged.

During the evaluation, a consistent offset was observed in some of the datasets. The only common factor was the set of software instructions used – CMMs and CMM software versions changed from partner to partner. The exact cause is unknown but the systematic error (7.5 µm for the smooth involute, 3.7 µm for the sinusoidal involute) was corrected in the affected datasets. This should not affect the validity of the results with respect to method validation. The uncertainty estimates are not directly affected. An offset increases the deviations of all measurands – here, distances from the expected profile – and hence lead to increased E_N-values. This would not be due to the uncertainties or the estimation methods, however, and should therefore not have any bearing on method validation.

Table 3-2: Constituent sine waves of the second involute profile.

Sine component	**1**	**2**	**3**
Wavelength / mm	80	25	8
Amplitude / µm	10	7	5

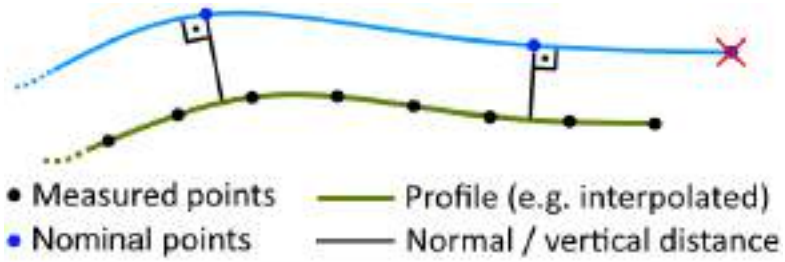


Figure 3-7: The involute profiles are evaluated against set reference points.

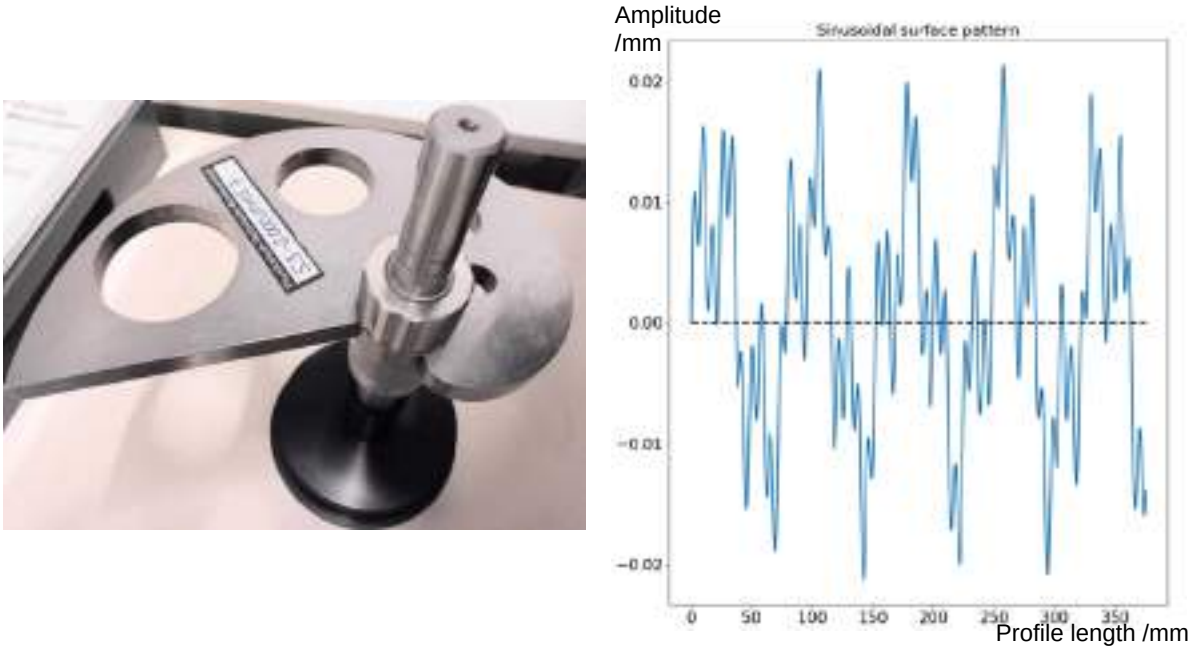


Figure 3-8: Involute gear (left) and the sinusoidal wave pattern added to one of two profiles.

3.3.5 Additional artefacts

Two additional workpieces had been selected: The steering knuckle and the NPL freeform plate. (**Figure 3-9**) The steering knuckle would have been another industrial workpiece; the plate would have introduced 3D freeform profiles. Unfortunately, the onset of the COVID-19 pandemic and knock-on effects repeatedly interfered with the preparation of the workpieces. Ultimately, they were never circulated, and no useful datasets could be collected for either part.

Figure 3-9*:* The steering knuckle (left) and freeform plate (right) were also selected for the validation study, but adverse conditions prevented them from entering circulation.

4 Data acquisition and collection

4.1 Organisation

Data was generated by project partners measuring one or more of the selected artefacts. The original project requirement was that every consortium member should measure one prismatic and one freeform artefact each to yield twenty-four datasets in total. The original plan also assumed that only two artefacts – one prismatic, one freeform – would be selected. Due to the additional requirements the consortium set itself with regard to the artefacts (see section 3.1), more were selected.

Measurements were planned in sequence similar to a round robin, albeit without the need for confidentiality regarding reference or measurement results, at least within the project consortium. A workpiece would be calibrated, or a reference measurement was made.[1] The workpieces were then sent to different partners according to a schedule drawn up during the planning stages. The outbreak of the COVID-19 pandemic, and in particular the early lockdowns and associated disruptions and uncertainty, scuttled any plans early on in EUCoM round robin.

- Reference measurements were delayed and, in two cases, could not be completed in a timely fashion. The artefact count was reduced from eight to six and some workpieces entered circulation later than planned. One artefact already in circulation was temporarily “trapped” when a partner was forced into lockdown.

- The schedule, which included a total of thirty-seven measurements, became unworkable due to nations’ very varied response to the pandemic. This included lockdowns, mandatory work from home or limited access to laboratories, as well as knock-on effects, e.g. delayed CMM maintenance, etc.

- The project was granted a six-month extension to help mitigate the impact of the pandemic.

A further issue was the separation of the UK from the EU. The repeatedly delayed exit agreement added more uncertainty, and ultimately created a new, initially somewhat chaotic, customs border between NPL and the EU-bound project partners.

Beneficial deviations from the project plan included two new participants joining the round robin as project collaborators. Cracow University of Technology (Poland) joined early in the starting phase of the project and was part of the schedule. MG Spa (Italy) joined much later but was able to contribute datasets as well.

Maintaining an organised, long-term schedule was impossible during the global pandemic. The organisational model therefore switched to an “ad hoc” approach: Partners would report when a measurement was complete and the artefact ready for shipment. The round robin organiser then polled the consortium to decide the next destination for this artefact. Beside availability and ability of a partner to measure the artefact in short order, a key consideration was the likelihood of a lockdown. Beyond that, artefacts had to be distributed such that all partners would have the opportunity to contribute at least the minimum amount of data required.

Another issue were customs, which added significant delays to the otherwise spontaneous shipment of artefacts. The original schedule had allowed generous time slots and additional buffer time. Partners outside the EU customs union had been scheduled to go last so that the return trip would not affect the round robin. These provisions were, of course, also negated.

[1] Calibration certificates could not be issued for every part. For simplicity, all these measurements are simply referred to as a “reference measurements” to distinguish them from the measurements made for EUCoM-type evaluations.

4.2 Collection

Datasets were shared openly within the consortium. They were uploaded to a shared online repository. Specially designed template files were prepared for each artefact, where the data was to be entered in a specified format. Depending on the method, the templates (spreadsheets, created in MS Excel 365, v2108, 32b) either executed the calculations directly or re-implemented methods developed and tested with other tools such as MATLAB. For context, “one dataset” includes all the data required for the uncertainty estimates using method A.

After initial discussion, it was decided to share the implementation of the uncertainty calculations, rather than having each partner implement it for themselves. Since the calculations are essentially the same, re-implementing them in different ways should yield no benefit. At most, this would be a test of how easy the methods are to realise. This is interesting with a view on future adoption of the methods, but not the objective of method validation. It could also lead to a lot of “unnecessary” validation work, tracking down any discrepancies between implementations. A single implementation must still be checked but avoids technical pitfalls, such as small numerical differences arising from the use of different software. Another benefit is that all data and results are forced into common formats. This makes reading, archiving, and evaluating datasets much easier as well.

4.2.1 Method A spreadsheet templates

The spreadsheet templates collect the data and calculate the uncertainty according to method A, as well as the normalised error, E_N, between a measurement (with method A uncertainty) and the reference measurement (with a non-EUCoM uncertainty estimate). The use of spreadsheets offers several benefits:

- Human-readable, interactive format in widespread use
- Not license-bound; free and open-source software kits are available
- Can handle the calculations for method A directly
- Clear structure for data input and output (tables and graphs)
- Can be parsed by other software
- Calculations and implementation require relatively little technical experience to follow
- Artefact-specific templates can be easily adapted to suit new test objects

The main disadvantages are scaling and speed when dealing with very large datasets – not the case here – and a set of functions that is somewhat limited compared to other tools. Thus, the spreadsheet implementation of some of the more complex evaluation steps was less direct and efficient than in MATLAB, for example.

A template is split into several worksheets. The introduction includes an illustrated step-by-step guide, informing users on how to use the template and where which data needs to go. (**Figure 4-1**)

The length and sphere standard data from method A is placed in a separate sheet. (**Figure 4-2**) This sheet calculates uncertainty contributions related to the scale and probes.

The main sheet is a large data entry table with one row per measurand. (**Figure 4-3**) Some contributions can selectively be included, excluded, or corrected in the uncertainty calculation. The table displays averaged measurand results, a corrected value if applicable, the expanded measurement uncertainty according to method A, and the normalised error. Graphs show a breakdown of the individual components of the uncertainties (see **Table 4-1**) and the differences between a measurement and the reference data.

The data is entered simply by copying or importing the numbers to the appropriate cells. Calculations are implemented as spreadsheet formulas. Many intermediate steps are hidden from view to keep the visible size more manageable. Anything besides “data entry” cells is set

to read only to avoid accidental overwriting of critical cells. However, the protections can be removed, and hidden calculations be revealed. Users can still inspect the inner workings of the template and modify it, for example, to adapt it to another workpiece with different measurands.

The templates for the two freeform objects are special cases, since both artefacts only have large numbers of points to evaluate. Since there are up to 1000 points, interactive settings are changed globally, not individually. In addition to the point-wise evaluation, each template also collects and evaluates minima and maxima for each repeat measurement. The involute gear template takes this further by calculating the uncertainties based on all minima and maxima and estimating the profile's form deviation from this.

The spreadsheets' flexibility was an important consideration, not just for the project but also for future use. Someone interested in testing the EUCoM methods for themselves can use these templates as a starting point, either using them directly or, for example, as a means of validating their own implementation. As both human- and computer-readable files, the templates may also double as a data archive.

Table 4-1*:* Method A uncertainty contributions.

Contribution	Description
E_S	Scale error from length standard measurement
u_S	Scale error uncertainty
E_D	Probe size error from diameter standard measurement
u_D	Probe size error uncertainty
$E_{D,Loc}$	Probe location error from diameter standard measurement
$u_{D,Loc}$	Probe location error uncertainty
u_{rep}	Uncertainty contribution from repeatability
u_{geo}	Uncertainty contribution from CMM geometry errors

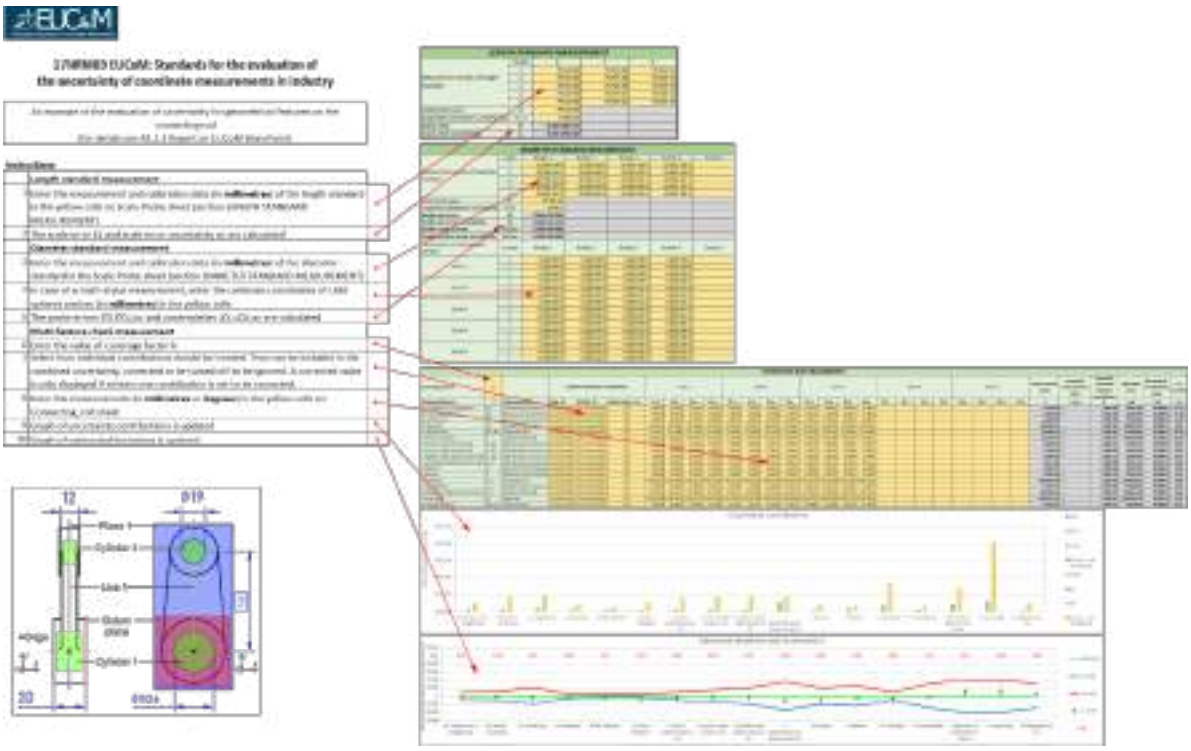


Figure 4-1: The “Introduction” sheet guides users through the template and highlights key elements.

DIAMETER STANDARD MEASUREMENT

	Cycle	Probe 1	Probe 2	Probe 3	Probe 4	Probe 5
Measurement results of diameter standard	1	25.976 150 0	25.976 550 0	25.976 480 0		
	2	25.976 140 0	25.976 550 0	25.976 480 0		
	3	25.976 150 0	25.976 560 0	25.976 480 0		
	4	25.976 180 0	25.976 550 0	25.976 480 0		
	5	25.976 140 0	25.976 540 0	25.976 450 0		
Calibrated value	Dc	25.976 32				
Expanded calibration uncertainty	U D,C	0.000 1				
Probe size error	**ED**	**0.000 517 555**				
Probe size error uncertainty	**uD**	**0.000 101 670**				
Probe location error	**E D,Loc**	**0.000 489 183**				
Probe location error uncertainty	**u D,Loc**	**0.000 071 880**				
Coordinates of LSM spheres centres	Coords	Probe 1	Probe 2	Probe 3	Probe 4	Probe 5
Cycle 1	x	0.000 190 0	0.000 170 0	-0.000 050 0		
	y	0.000 080 0	-0.000 090 0	-0.000 030 0		
	z	0.000 130 0	-0.000 100 0	0.000 020 0		
Cycle 2	x	0.000 180 0	0.000 200 0	-0.000 040 0		
	y	0.000 150 0	0.000 080 0	-0.000 050 0		
	z	0.000 150 0	0.000 510 0	0.000 050 0		
Cycle 3	x	0.000 140 0	0.000 240 0	0.000 030 0		
	y	0.000 120 0	-0.000 010 0	0.000 030 0		
	z	0.000 150 0	-0.000 120 0	0.000 010 0		
Cycle 4	x	0.000 140 0	0.000 200 0	0.000 000 0		
	y	0.000 110 0	-0.000 040 0	0.000 080 0		
	z	0.000 120 0	-0.000 130 0	0.000 020 0		
Cycle 5	x	0.000 130 0	0.000 200 0	0.000 040 0		
	y	0.000 080 0	-0.000 050 0	0.000 100 0		
	z	0.000 170 0	-0.000 140 0	0.000 020 0		

LENGTH STANDARD MEASUREMENT

	Cycle	X	Y	Z
Measurement results of length standard	1	199.827 45	199.827 58	199.827 67
	2	199.827 42	199.827 59	199.826 85
	3	199.827 40	199.827 41	199.826 80
	4	199.827 49	199.827 38	199.826 93
	5	199.827 41	199.827 55	199.826 93
Calibrated value	Lc	199.827 03		
Expanded calibration uncertainty	U Lc	0.000 01		
Scale error	**EL**	**0.000 225 335**		
Scale error uncertainty	**uL**	**0.000 150 025**		

Figure 4-2: The “Scale-Length” sheet accepts the sphere and length standard measurement data and calculates some of the uncertainty contributions for EUCoM method A.

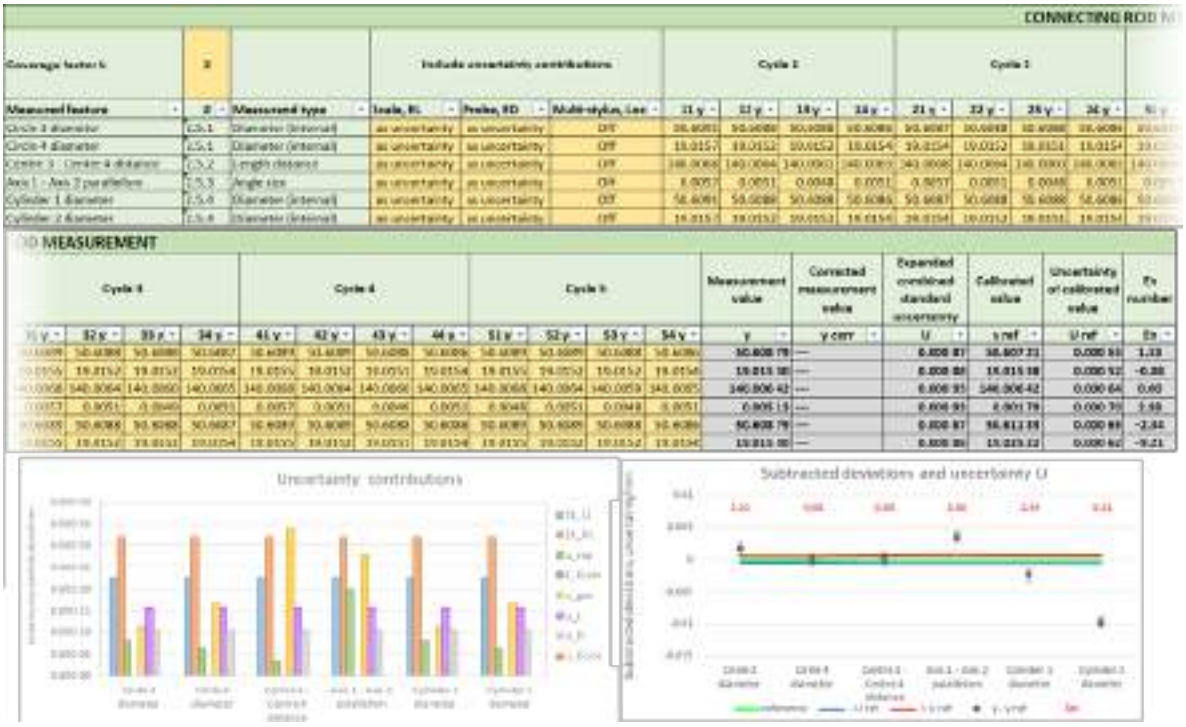


Figure 4-3: The artefact-specific sheet lists all the measurands and allows users to include, exclude or correct some of the Method A uncertainty contributions. Results are also displayed here.

4.2.2 Method B1 GUM methodology

Method B1 templates were also created as spreadsheets. The original implementation was created using MATLAB, which is more efficient but less accessible. A spreadsheet-type implementation offered the same benefits as discussed in the previous section. The downside is that many matrix evaluations are not as easy or efficient to implement. As a result, the template for the connecting rod template uses twenty-two sheets to run the calculation. A documentation sheet provides an overview.

Only two sheets are used for user input and to display results. The first accepts information about the CMM, e.g., MPE-values, the measurement volume and probe information. More parameters gauge the performance of the system. (**Figure 4-4**) Experienced users can estimate these themselves or use the suggested default values. Examples are repeatability, scale and squareness and different types of spatial correlations. The second input sheet accepts the nominal point cloud and probing vectors. Each point must also be associated with a particular probe which would be used to measure it. (**Figure 4-5**)

CMM information		CMM summary information					
Cert. date		Effect	Parameter	Unit	Value		Value Advice
Cert. number		MPE	A	mm	0.002		
MPE A0/microns	**2**		B	1	1.25E+05		
MPE B0/1000	**125**	**Repeatability**	sigR	mm	2.00E-04	<=	1.00E-03
Manufacturer		**Scale and squareness**	sigL	1	4.00E-06	<=	4.00E-06
Model			sigLa	1	2.00E-06	<=	4.00E-06
Serial Number			sigQ	1	2.00E-06	<=	4.00E-06
Measureing volume/	mm	**Probe qualification**	sigPQ	mm	5.00E-04	<=	1.00E-03
LX	500	**Spatially correlated location e**	sigET	mm	5.00E-04	<=	1.00E-03
LY	500		lamET	mm	1.00E+02	~	20 mm - 200 mm
LZ	400	**Spatially correlated rotation e**	sigER	radians	1.00E-05	~	a few arcseconds
			lamER	mm	2.00E+02	<=	50 mm - 400 mm
			Pmax	mm	2.00E+01		
		Spatially correlated probing e	sigP0	mm	2.00E-04	<=	1.00E-03
			sigP	mm	2.00E-04	<=	1.00E-03
			lamP	1	5.00E-01	~	0.2 - 0.6
		Expansion factor	K	1	2.00E+00	~	0.2 - 0.7
		Uncertainty calculations					
		Absolute component	sigA	mm	0.000906	<=	1.00E-03
			Lmax	mm	812.4038		

Figure 4-4: The input side of the “CMM_Summary” sheet parameterises the CMM for the simulation. The sheet also displays tables of the uncertainty contributions for individual characteristics.

Xx	Xy	Xz	Nx	Ny	Nz	ProbeInde	w
20.00000	0.00000	-10.00000	1.00000	0.00000	0.00000	1	1.00E+00
0.00000	20.00000	-10.00000	0.00000	1.00000	0.00000	1	1.00E+00
-20.00000	0.00000	-10.00000	-1.00000	0.00000	0.00000	1	1.00E+00
0.00000	-20.00000	-10.00000	0.00000	-1.00000	0.00000	1	1.00E+00
20.00000	0.00000	0.00000	1.00000	0.00000	0.00000	1	1.00E+00
0.00000	20.00000	0.00000	0.00000	1.00000	0.00000	1	1.00E+00
-20.00000	0.00000	0.00000	-1.00000	0.00000	0.00000	1	1.00E+00
0.00000	-20.00000	0.00000	0.00000	-1.00000	0.00000	1	1.00E+00
20.00000	0.00000	10.00000	1.00000	0.00000	0.00000	1	1.00E+00
0.00000	20.00000	10.00000	0.00000	1.00000	0.00000	1	1.00E+00
-20.00000	0.00000	10.00000	-1.00000	0.00000	0.00000	1	1.00E+00
0.00000	-20.00000	10.00000	0.00000	-1.00000	0.00000	1	1.00E+00
12.00000	150.00000	-5.00000	1.00000	0.00000	0.00000	2	1.00E+00
0.00000	162.00000	-5.00000	0.00000	1.00000	0.00000	2	1.00E+00

Figure 4-5: The “Coordinate_Data_All” sheet accepts the nominal point cloud, vectors and the probe associated with each point.

4.2.3 Method B2 sensitivity analysis

Method B2 templates were also implemented as spreadsheets following the same basic principles of the previous methods. The information required to apply the sensitivity analysis approach to uncertainty evaluation concerns CMM data from performance testing. In particular, MPE statements and the results of re-verification tests, i.e., length measurement error E_L according to ISO 10360-2. [3] The following figures present the spreadsheet templates which are used to calculate the necessary coefficients corresponding to particular CMMs. (**Figure 4-6**, **Figure 4-7**)

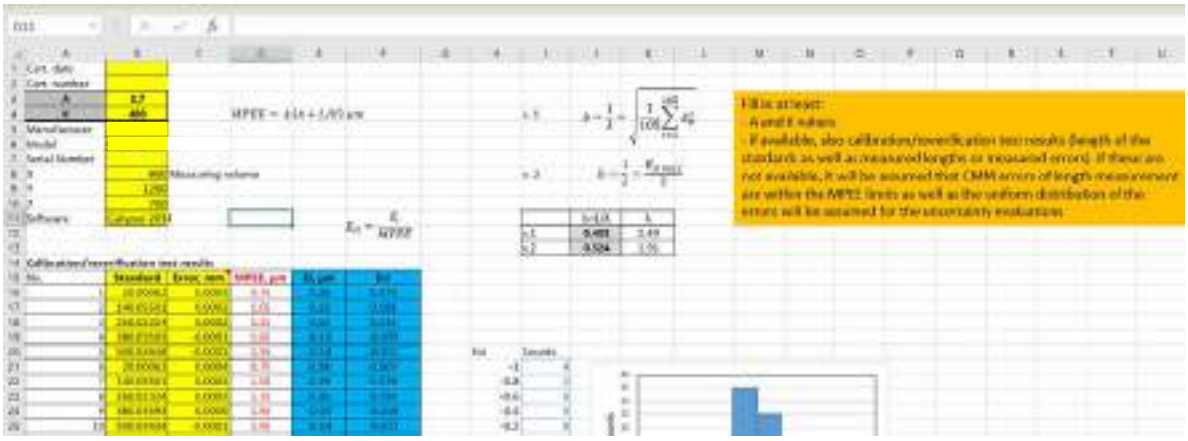


Figure 4-6: Datasheet concerning basic information about the CMM – yellow marked fields are to be filled in; the “b” coefficient is calculated automatically in two versions reflecting the actual performance of the CMM.

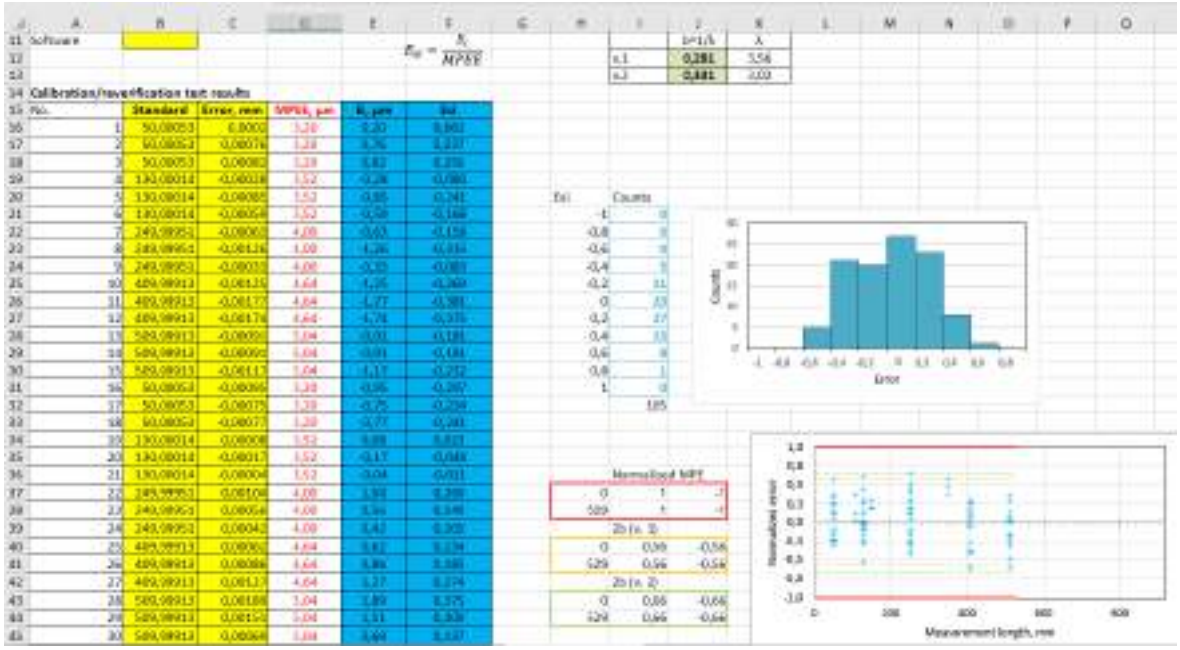


Figure 4-7: Example evaluation of a CMM’s “b” coefficient with additional graphical analysis.

For the workpieces used in the round robin, template spreadsheets were prepared for evaluating the measurement uncertainty and uncertainty budgets for each characteristic. This is done using the information and intermediate results calculated by the previous template. **Figure 4-8** depicts the spreadsheet for the connecting rod.

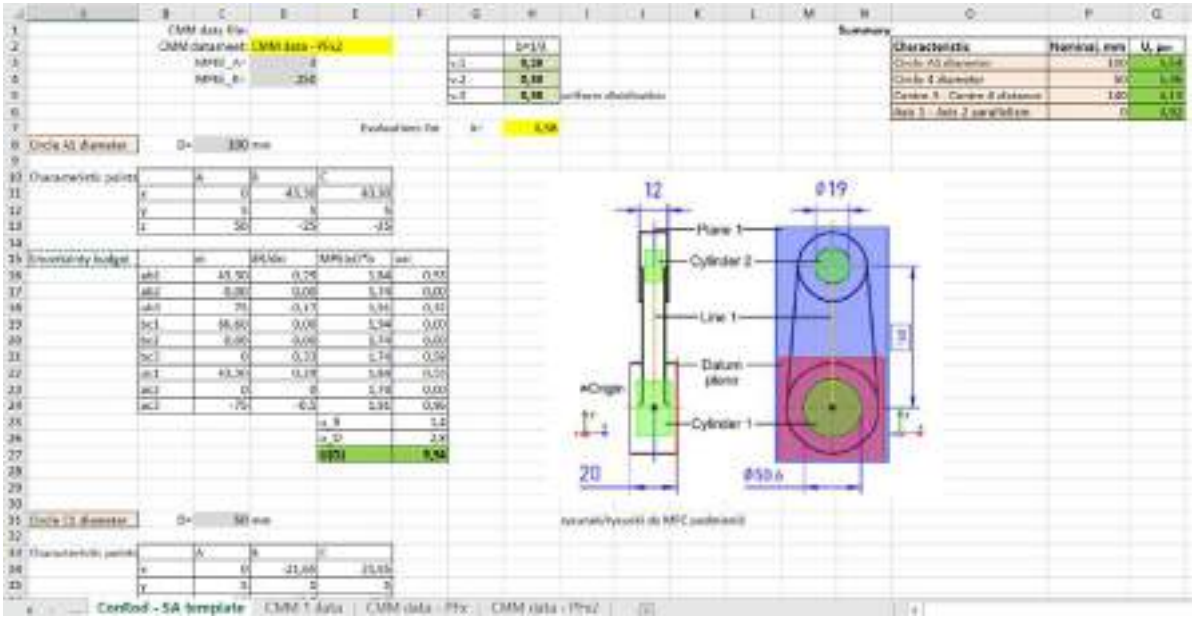


Figure 4-8*:* Example evaluation of measurement uncertainty for connecting rod's measurands with uncertainty budgets for each characteristic.

5 Validation method

The EUCoM methods were validated using the round robin data collected by the consortium: a total of forty-seven tactile datasets, not counting reference measurements for each artefact. The following results were considered:

- Direct comparison of a measured value (x_m) to the corresponding reference value (x_r)
- Direct comparison of a measurand's EUCoM uncertainty (U_A, U_{B1}, U_{B2}) to the reference uncertainty (U_r)
- Direct comparison of one EUCoM uncertainty to another (e.g. U_A vs. U_B)
- Normalised error for a measurand pairing a EUCoM result and corresponding reference ($E_{N,Ar}$, $E_{N,B1r}$, $E_{N,B2r}$)
- Normalised error for a measurand pairing consecutive EUCoM measurements or two EUCoM uncertainty methods. (e.g., $E_{N,AA}$ – same method – or $E_{N,B1A}$ – different methods, where the second method uses data from a previous measurement)

Measurement results should ideally agree with the reference value and other measurements. Discrepancies indicate either an issue with one or more of the measurements involved or are explained by the measurement uncertainty. Hence the normalised error, E_N, is calculated to test the conformity of two results. The normalised error considers the difference between two measurement results while taking their associated uncertainties into account. For instance, two very different results could still be compatible, if the uncertainties are very large. Conversely, results with very small uncertainties need to agree closely or will still fail the conformance test. E_N is a dimensionless number; to pass the conformance test it should be between -1 and 1. (**Equation 5-1**) [6] However, E_N-values close to zero may also indicate that at least one uncertainty has been grossly overestimated.

$$E_N = \frac{x_m - x_r}{\sqrt{U_m^2 + U_r^2}} \qquad \textbf{Equation 5-1}$$

Here, x_m is a measurement result from a round robin measurement and x_r the corresponding reference value for a specific measurand or feature. U_m is the expanded measurement uncertainty of x_m and evaluated using one of the EUCoM methods. U_r is the uncertainty of the reference value and is **not** evaluated by EUCoM methods, but by any **other accepted** and **independent** method.

The normalised error can be calculated for any of the EUCoM uncertainties when tested against the reference value. Conformity between different EUCoM results can only be calculated using different datasets. Otherwise, the difference between measurement values, and thus E_N, would be equal to zero. For consistency, these comparisons always use consecutive measurements. This minimises the potential for drift over time affecting the result but still pairs up partners at random, as measurement order was also randomised. For practical testing, the method B uncertainties used the method A measurement results (x_m) to calculate corresponding E_N-values. This works since the measurement results from method A are simply averaged values from many repeat measurements. Since the results are the same, Method A-B cross-comparisons must also use consecutive datasets.

The uncertainties can also be compared directly. While the uncertainty budget will differ slightly for different estimation approaches, we would expect the overall uncertainty to be similar in most cases. The EUCoM methods offer alternative ways to determine measurement uncertainty, but they are not adding or removing any major contributors to or from consideration, compared to other methods. Likewise, different partners should usually obtain similar uncertainty budgets for their measurements, although different systems will naturally differ based on their performance.

6 Results

Measurement data was collected with the templates described in chapter 4. Calculations were performed with the templates. Some additional processing was applied to obtain or visualise the different comparisons in chapter 5.

Note that all datasets have been anonymised by assigning partners random ID numbers. Furthermore, E_N -charts may truncate the y-scale (E_N) to better visualise the results, as with values beyond ±1, the precise magnitude makes no difference to the end result.

6.1 Connecting rod

Nine datasets are available for the connecting rod. The measurands are limited to distances, diameters and parallelism of the cylindrical bore holes.

6.1.1 Method A uncertainties

The uncertainties are shown in **Figure 6-3**. The measurand uncertainties are very similar within each dataset, except for set 2, which yields much higher uncertainties for diameters than for distance or parallelism. Across sets, uncertainties fall into groups. Some datasets (e.g. 8, 10) are comparable with reference uncertainties (derived from VCMM). Exceptions are diameter uncertainties from 2, and set 11, which has the highest consistent uncertainty. Dataset 8a was acquired on the same CMM as the reference measurement.

The internal consistency of uncertainties suggests that measurand type has limited impact on the method A estimate. This aspect will be re-visited with the MFCs, which include more feature types. It's not clear why set 2 does not follow this trend. Here, the geometry contribution (u_{geo}) dominates the budget for the diameters but not the other measurands. (**Figure 6-1**). The other contributions remain stable. Thus, the noticeably higher diameter uncertainties are only due to u_{geo}.

For the most part, method A estimates are higher than the reference estimate. Most interesting is the 8a result, which was collected on the same CMM as the reference measurement itself. 8a fits in well with other method A results, but we can rule out "quality differences" in the CMM setup as a source of the different uncertainties. The measurement setup from 8a is of very high quality. Most other measurements cannot match these conditions, yet many uncertainties are on similar levels. Other factors, e.g., different operators, part handling and shipping between measurements, etc., are also eliminated. This suggest that (empirical) method A estimates are slightly more pessimistic than (simulated) VCMM estimates.

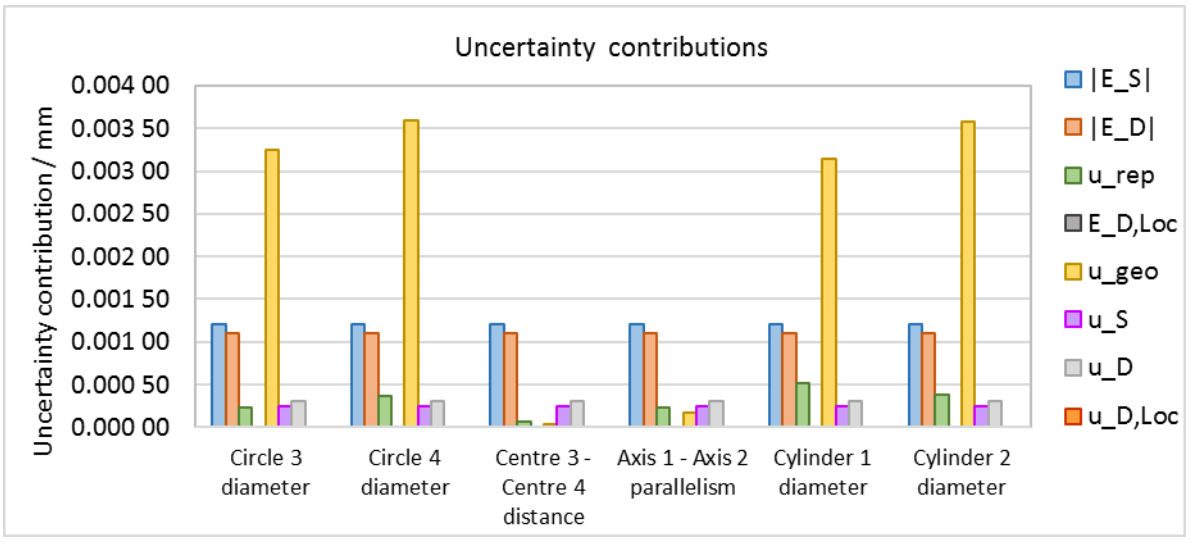


Figure 6-1: Method A uncertainty contributions to each measurand for dataset 2.

6.1.2 Method A conformity testing against the reference

The $E_{N,Ar}$ normalised errors are shown in **Figure 6-5**. The dataset results and Method A uncertainties are being tested against the reference data. The measurement results have a wide spread compared to the reference measurement and many values do not conform with the reference measurements. (**Figure 6-4**)

For example, the centre-centre distance yields very low E_N scores for all datasets as all results lie within the "tolerance zone" defined by the reference uncertainty. Method A uncertainties (indicated by error bars) extend far beyond the zone. Small measurand differences combined with large uncertainties yield very small E_N-values and may even indicate an overestimated uncertainty. Other measurands deviate more or have smaller uncertainties and thus fail the conformity test. This is most likely due to the problems with the connecting rod (see section 3.3.1), rather than an issue with the method itself. **Figure 6-2** shows some of the results and illustrates the drift of the larger, segmented cylinder as well as the inconsistency of the measured parallelism between the two "cylinder" axes.

The conformity test of 8a is somewhat special. Low deviations are expected for measurements acquired with the same setup. The deciding factor for a pass or fail are the separately determined uncertainties. However, because of the small deviations, the normalised errors are expected to low and too small for good conformance.

6.1.3 Method A conformity testing against other method A results

Comparing method A measurements to each other yielded slightly better results than the comparison to the reference. Note that the comparison could have paired up any measurements. However, measurements were paired in chronological order, as this might also reveal progressive degradation of the workpiece or individual measurands. There are fewer violations overall.

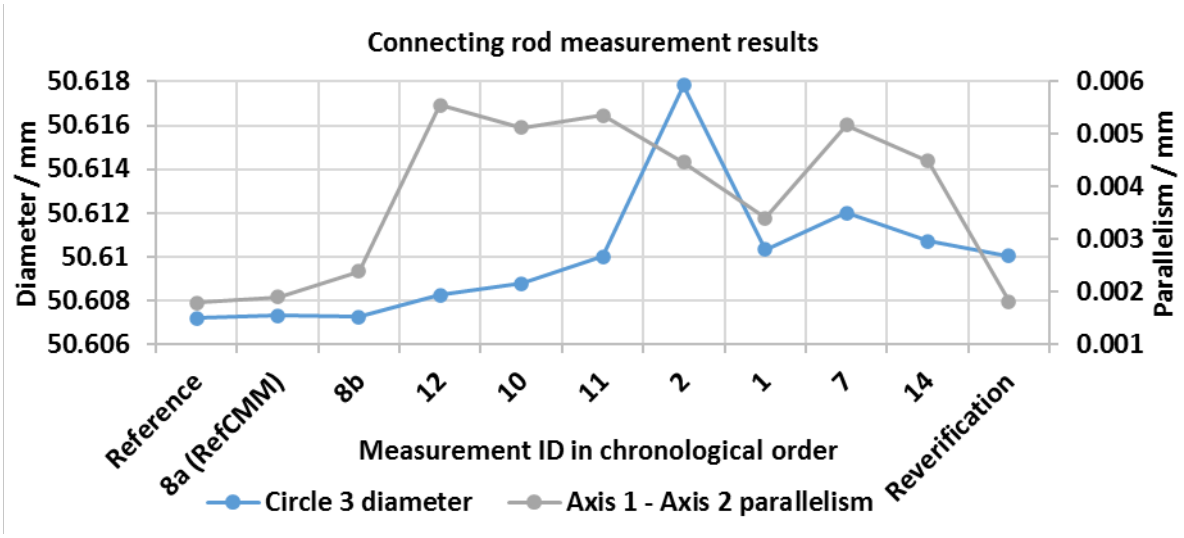


Figure 6-2: Excerpt from the connecting rod measurement results, showing drift over time on the larger ring and the inconsistent cylinder axes parallelism.

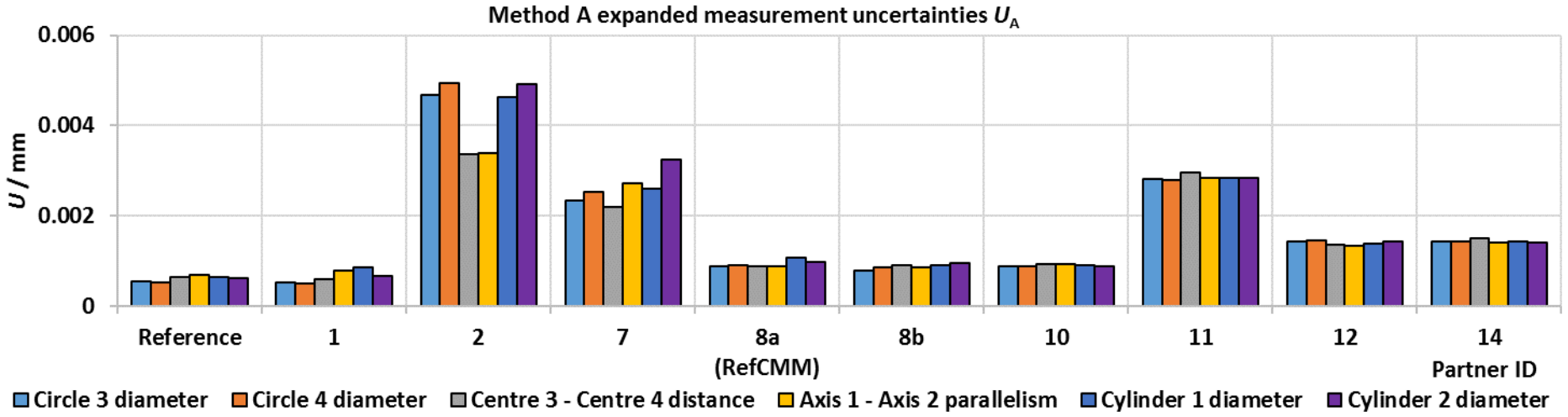


Figure 6-3: Connecting rod method A uncertainties. The reference measurement uncertainties were determined with the VCMM.

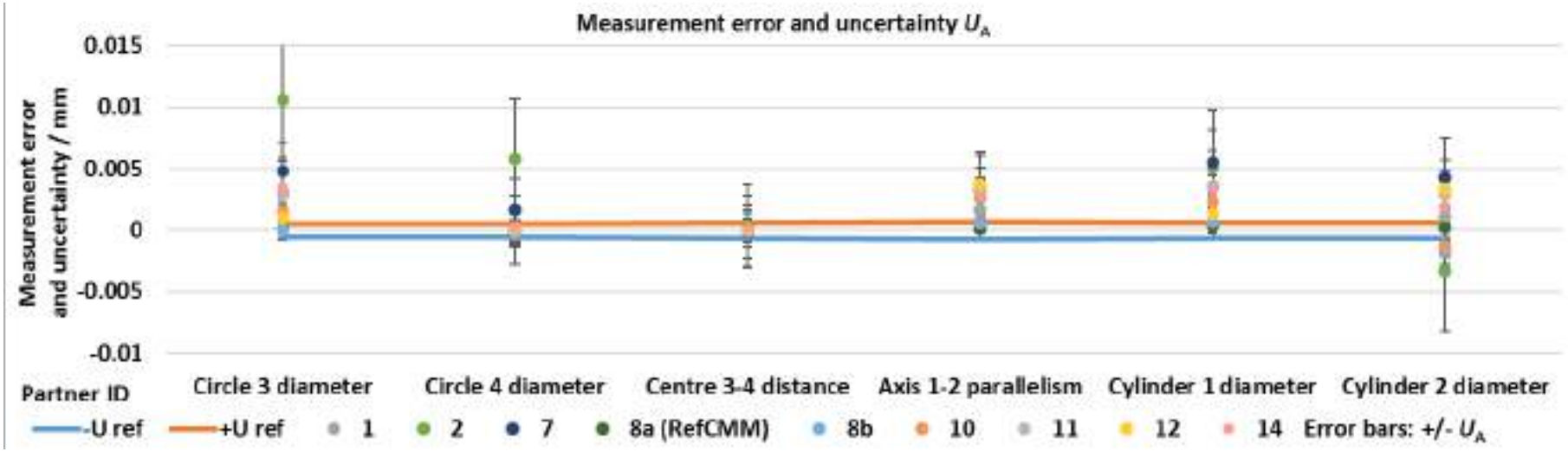


Figure 6-4: Connecting rod measurement deviations from the reference.

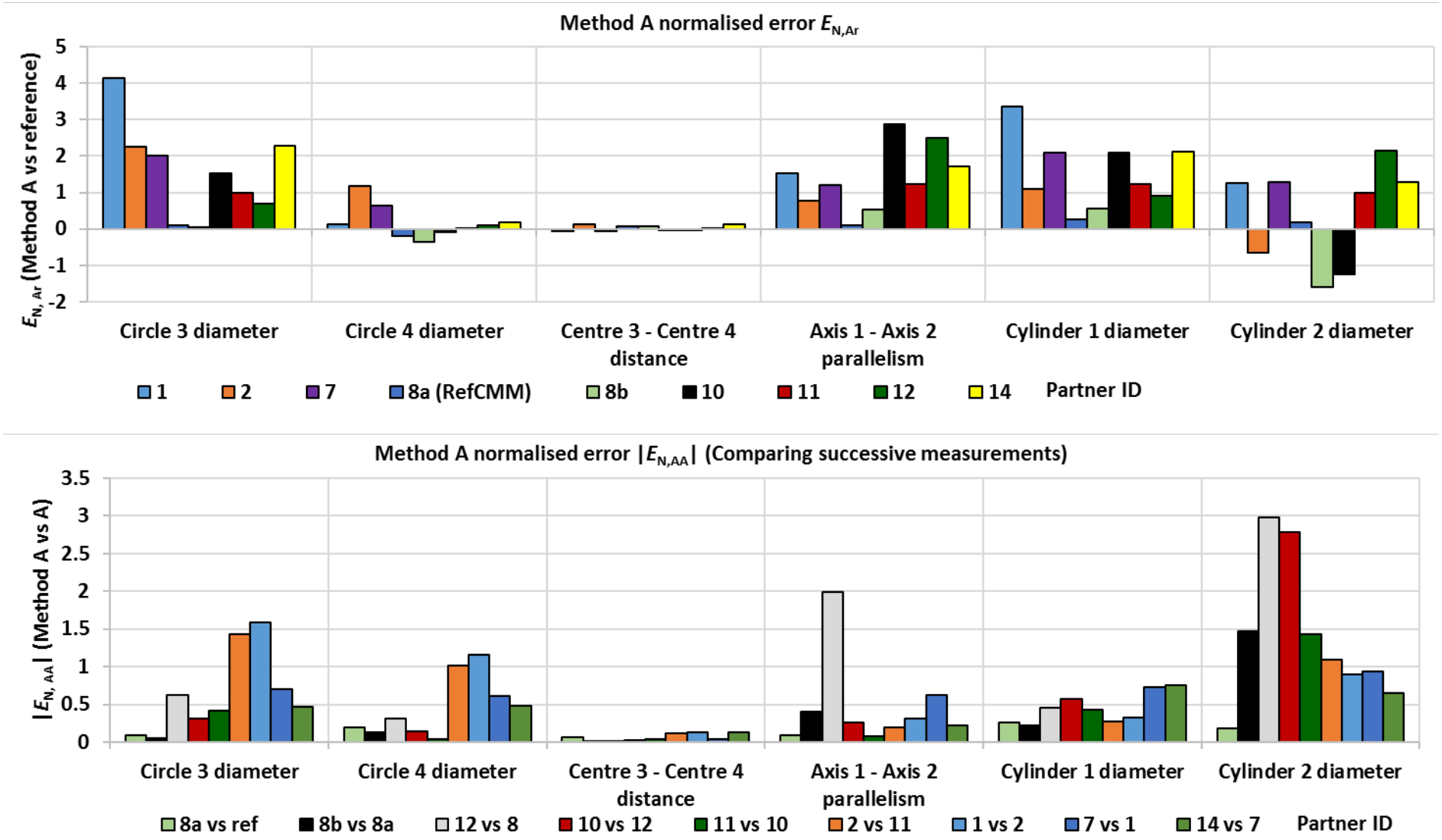


Figure 6-5: Connecting rod normalised errors calculated against the reference (top) or the previous measurement (bottom)

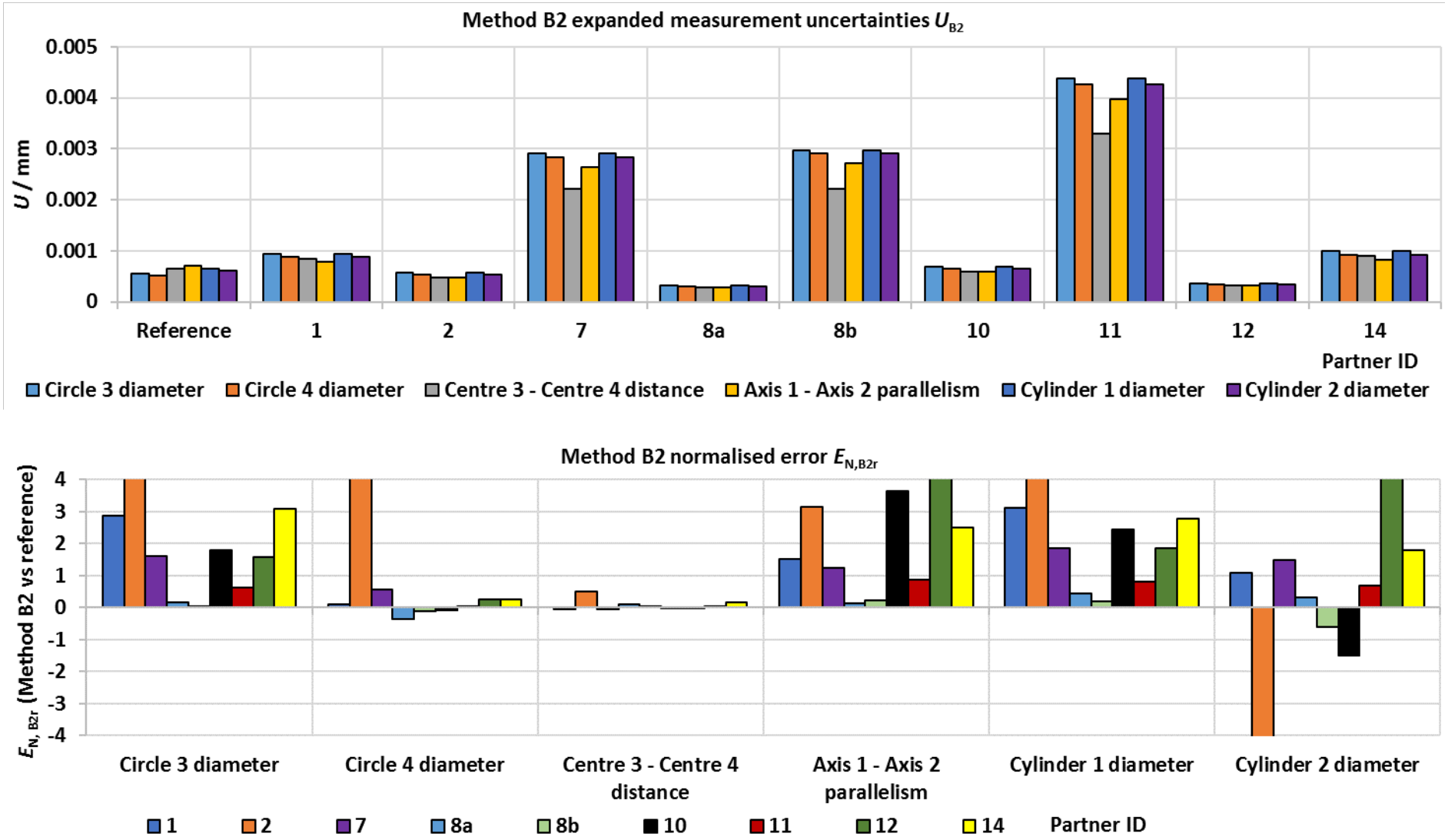


Figure 6-6: MPE-derived method B uncertainties for the connecting rod (top) and corresponding E_Ns relative to the reference values (bottom).

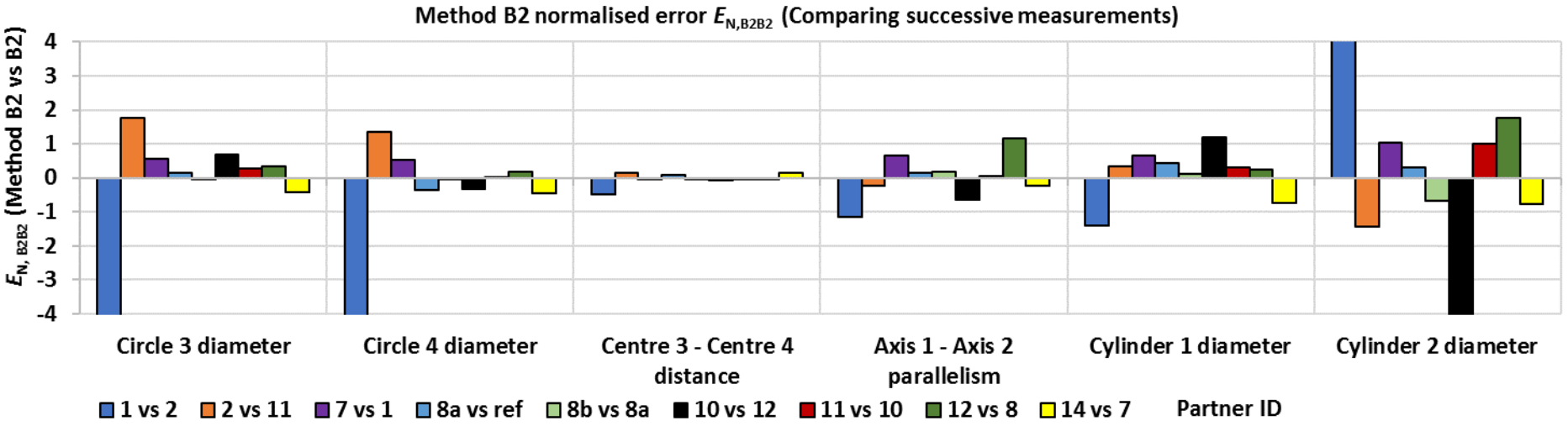


Figure 6-7: Connecting rod normalised errors contrasting consecutive B2-evaluated measurements.

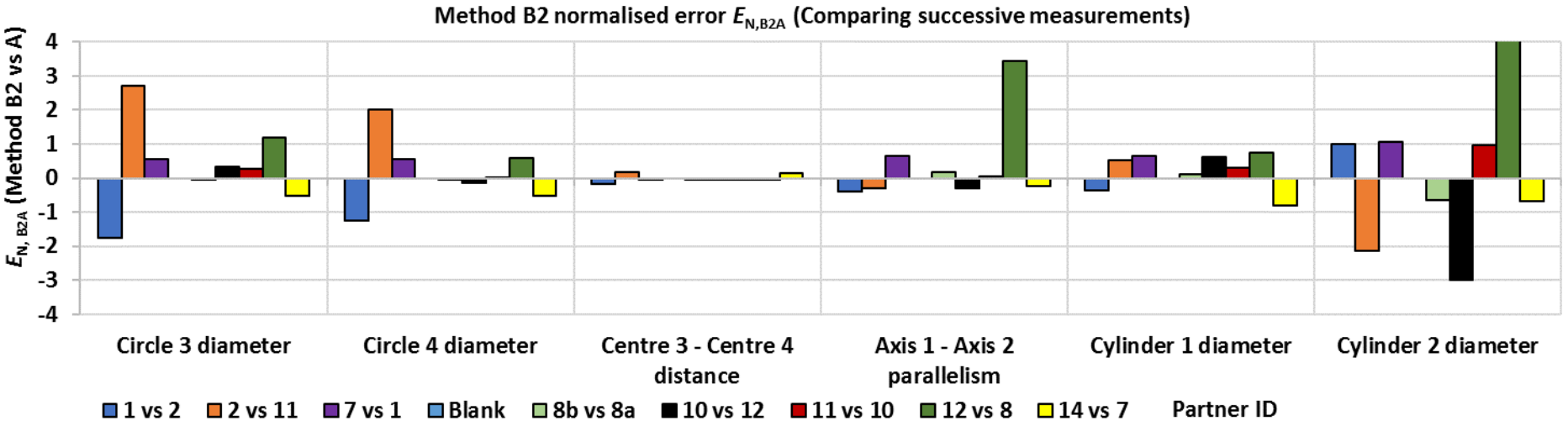


Figure 6-8: Connecting rod normalised errors contrasting methods B2 and A.

6.1.4 Method B2 uncertainties

The uncertainties derived from CMM MPEs and length measurement errors (E_L from EN ISO 10360-2) uncertainties are shown in **Figure 6-6**. For some CMMs, E_L data was not available. These datasets have been marked as “MPE only.” In contrast to method A, method B2 variation is due to slight differences between measurand models. The exact models (including intermediate numerical values) are specific to the connecting rod and independent of any measured values (which method B does not use) or the CMM performance (here, $E_{L,MPE}$ and EN ISO10360-2 E_L test lengths). Hence the variation of the B2 uncertainty follows the same general pattern in each dataset, although it scales with the overall uncertainty, which does depend on the CMM characteristics.

In terms of magnitude, uncertainties are similar to reference (VCMM) uncertainty estimates but tend to be slightly lower. Set 8a stands out, having been recorded on the same CMM. Both VCMM and method B2 are modelling the measurement machine. However, the VCMM is a much more detailed representation of the equipment and measurement. B2 is a more direct approach and unable to account for as many influence factors and should thus be expected to yield lower uncertainty estimates, given the same measurement conditions. By the same token, method A assesses some uncertainty contributions empirically, which are not necessarily captured by MPE values or the data from the length measurement test E_L.

6.1.5 Method B2 conformity testing

As seen before, the conformity testing does work very well. (**Figure 6-7**) Method B conformity testing is based on the same measurement data as method A, and thus has the same problems. Regardless of which results are compared – B2 and reference, B2 internally or B2 and A – there are always several failures. In every case, the normalised errors from the centre-centre distance are again very low, as seen before with method A. A possible explanation would be that B2 estimates are also too high for this measurand. Taken together, the implication of the results so far would be that all methods tested overestimate centre-centre uncertainties, at least in this particular case. This would include the VCMM, as U_r-values are similar to any other estimates. However, given the problems with the connecting rod and its measurement, this is an unlikely conclusion, especially in light of the results from other workpieces.

6.2 Multi-feature check

Nine measurements of the LQ-MFC were available, and three of the HQ-MFC. The measurands cover a wide range of dimensional and geometrical characteristics.

The measurement volume of the CMM for LQ-set 6 was barely large enough to accommodate the workpiece in all orientations. The resulting difficulties have affected the measurement results as well. Some results are not shown in the following charts as they distort the scale and obscure the remaining data points. Sets 4a-f were recorded on three different CMMs and different sampling strategies. Sets 4b, d, f acquired surface points by scanning, the others used individual points and thus much more sparse data for the same measurement. Pairs 4ab, 4cd and 4ef each used the same CMM for acquisition. The reference uncertainties were determined with the VCMM method.

6.2.1 Method A uncertainties

The majority of the uncertainties are consistent with each other, both within individual datasets and across measurements. (**Figure 6-9**) The measurand or feature type is therefore not a factor for this type of uncertainty estimate. Since the MFC covers a wide range of typical features, the method will likely work for most, if not all prismatic measurands.

Measurement group 4 of the LQ confirms that the sampling strategy has no obvious effect on the result or method A uncertainty. Beside the reference standard measurements, the uncertainties are calculated from derived features, i.e., measurands based on one or more fitted elements. Hence, the amount of data available – detailed scanned profiles or sparse point clouds – only indirectly affects the uncertainty.

A few measurands in 4a and 4b were omitted. In both cases, the results from the third orientation were significantly higher than from any other repeat measurement. Since both datasets reported the exact same values, this is most likely a transcription error rather than a measurement outlier. It was not possible to recover the missing data.

The high uncertainties in LQ set 6 are a result of the difficulties with the measurement itself. Method A is based on measurement data. Thus, if the system is not suited to a particular measurement task and performs a poor measurement, both the measurement value and uncertainty will reflect this. Most of the features were evidently safe to measure since the results and uncertainties are comparable with other measurements. This issue only appears when the CMM is operating at edge of the normal measurement volume. It's not clear whether the uncertainty estimates are reasonable estimates for poor measurements, or whether the mathematical model for method A breaks down, but coincidentally still delivers plausible numbers in this instance. If it is the former, the results could in principle still be used, stating the high associated uncertainty. In either case, method A does appear to highlight problematic measurements. This would be very useful particularly for stand-alone measurements, where, without reference values, problems might otherwise go unnoticed. The omitted results and uncertainties have been collected in **Table 6-1**.

The error in 4a/4b has relatively little impact on the measurement results but does affect the uncertainty, which increases almost ten-fold compared to other uncertainties from the same datasets. The uncertainty breakdown correctly identifies u_{geo} as the main contributor to the uncertainty. Since only one orientation is affected, the CMM geometry would at first appear to be the cause.

In measurement 6, both values and uncertainties indicate problems. Individual measurement results vary significantly across repeat measurement and re-orientation measurements. Here, both u_{geo} and u_{rep} are the main uncertainty sources, as expected, exceeding any other contributions by several orders of magnitude.

Table 6-1: Outliers or erroneous values not included in the evaluation

ID	Characteristic	Method A result / mm or °		Reference result / mm or °	
		Value	Uncertainty	Value	Uncertainty
4a	Plane perpendicularity	0.015	0.017	0.011	0.004
4a	Plane flatness	0.012	0.013	0.007	0.002
4a	Plane parallelism	0.016	0.015	0.012	0.003
4b	Plane perpendicularity	0.017	0.015	0.011	0.004
4b	Plane flatness	0.013	0.013	0.007	0.002
4b	Plane parallelism	0.018	0.014	0.012	0.003
6	Line straightness	1.362	2.746	0.002	0.001
6	Cone diameter	49.982	0.013	49.854	0.004
6	Cone angle	120.592	0.972	120.009	0.004

6.2.2 Method A conformity testing against the reference

The normalised errors of both artefacts are shown in **Figure 6-11**. The majority of the LQ results conforms to the reference. E_N is usually evaluated as an absolute value. This was not done in this report in case the sign might add useful information. Here, there is a tendency for negative E_N-values, indicating a prevalence of test measurements which underestimate the “true” or reference values. There are also some instances of very small E_N-values, a potential indicator for overestimated uncertainties.

About a quarter of the HQ-MFC results are failures. The difference between the two MFCs stems from the different reference uncertainties. The HQ-MFC uncertainty is very low even compared to the LQ-MFC reference uncertainty. The lower the uncertainties are, the more restrictive the normalised error criterion becomes. As **Figure 6-10** shows, the measurement deviations of both artefacts are very similar and spread evenly around 0. Both reference and method A uncertainties are estimated to be higher for the LQ-MFC, making it easier for measurement to pass the test. The opposite applies to the HQ-MFC. This difference may reflect the age of the LQ-MFC, as hoped during the initial artefact selection. Surfaces worn down by repeated measurement over a long period of time would change the actual values – which the reference measurement would reflect and cancel out – and affect the repeatability of measurements. This would directly impact the empirical method A estimate. In this context, it is important to note that the EUCoM measurement strategy followed the manufacturer’s suggested measurement strategy. It seems probable that many previous users of the workpiece followed the same strategy, leading to more cumulative wear in the areas of interest on the MFC.

6.2.3 Method A conformity testing against other method A results

Tests of method A results against each other were mostly successful. The LQ-MFC results are shown in **Figure 6-12** by way of example. A few very large E_N-values involve the low quality measurands from set 6 or the anomalies from 4a and 4b. Few other violations remain and are most likely the result of normal variation. Large data collections such as these will almost

inevitably contain some fringe values and outliers. There are no patterns visible indicating other issues with individual measurements, measurement systems or the uncertainty evaluation.

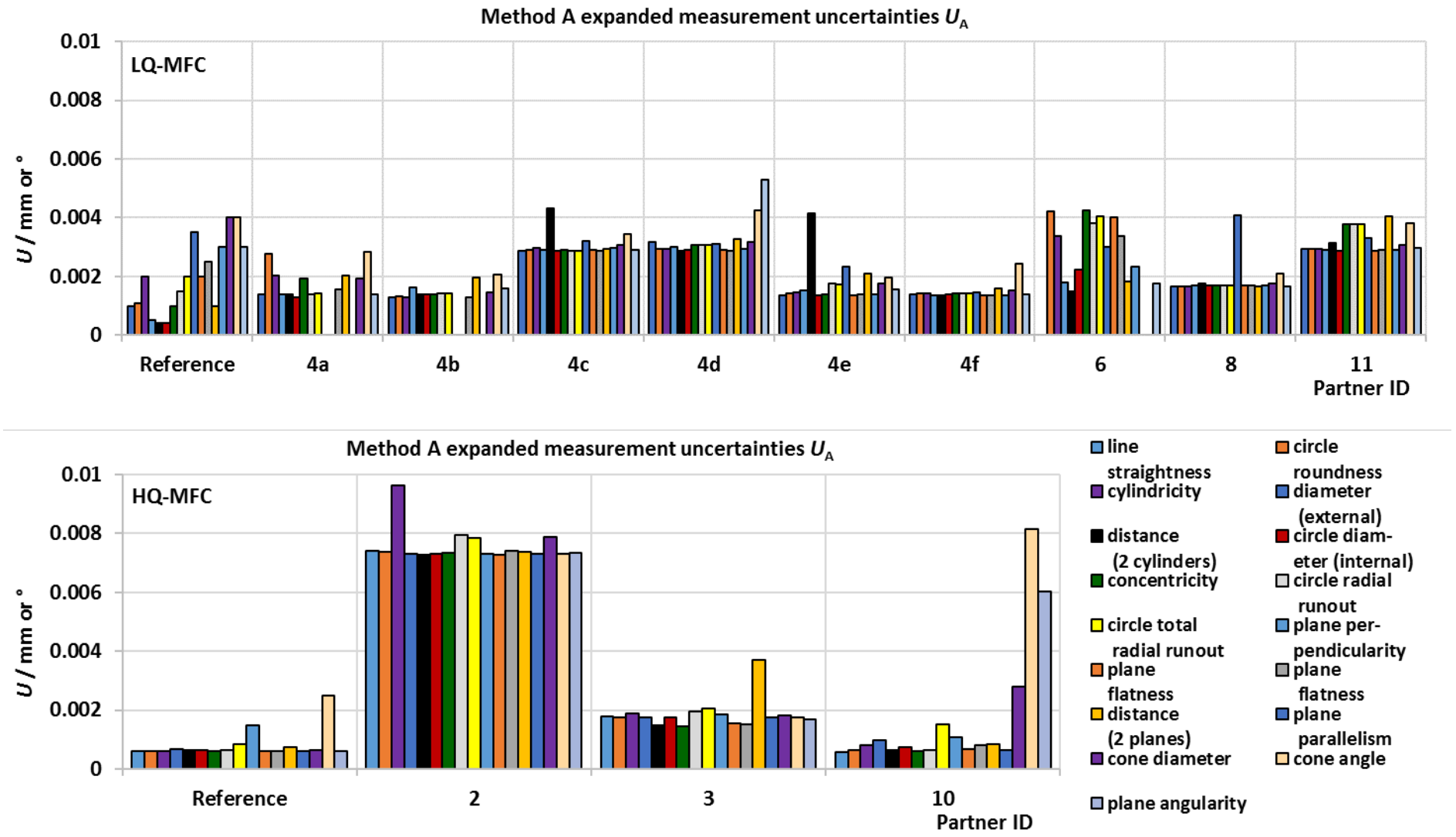


Figure 6-9: Multi-feature check method A measurement uncertainties. Not shown: LQ set 6 line straightness and cone angle.

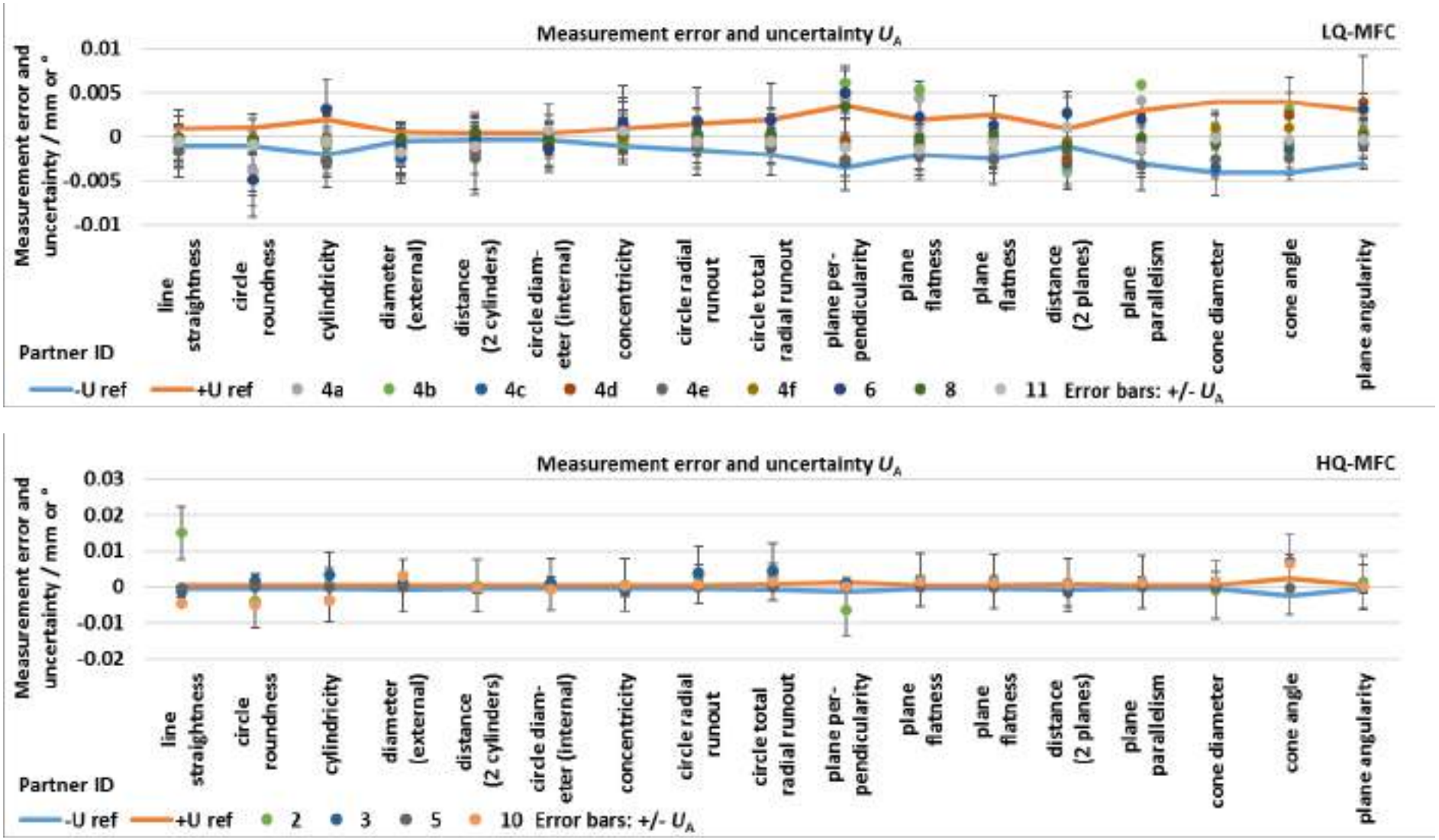


Figure 6-10: Deviations from the reference measurement. Not shown: LQ set 6 line straightness, cone angle and cone diameter.

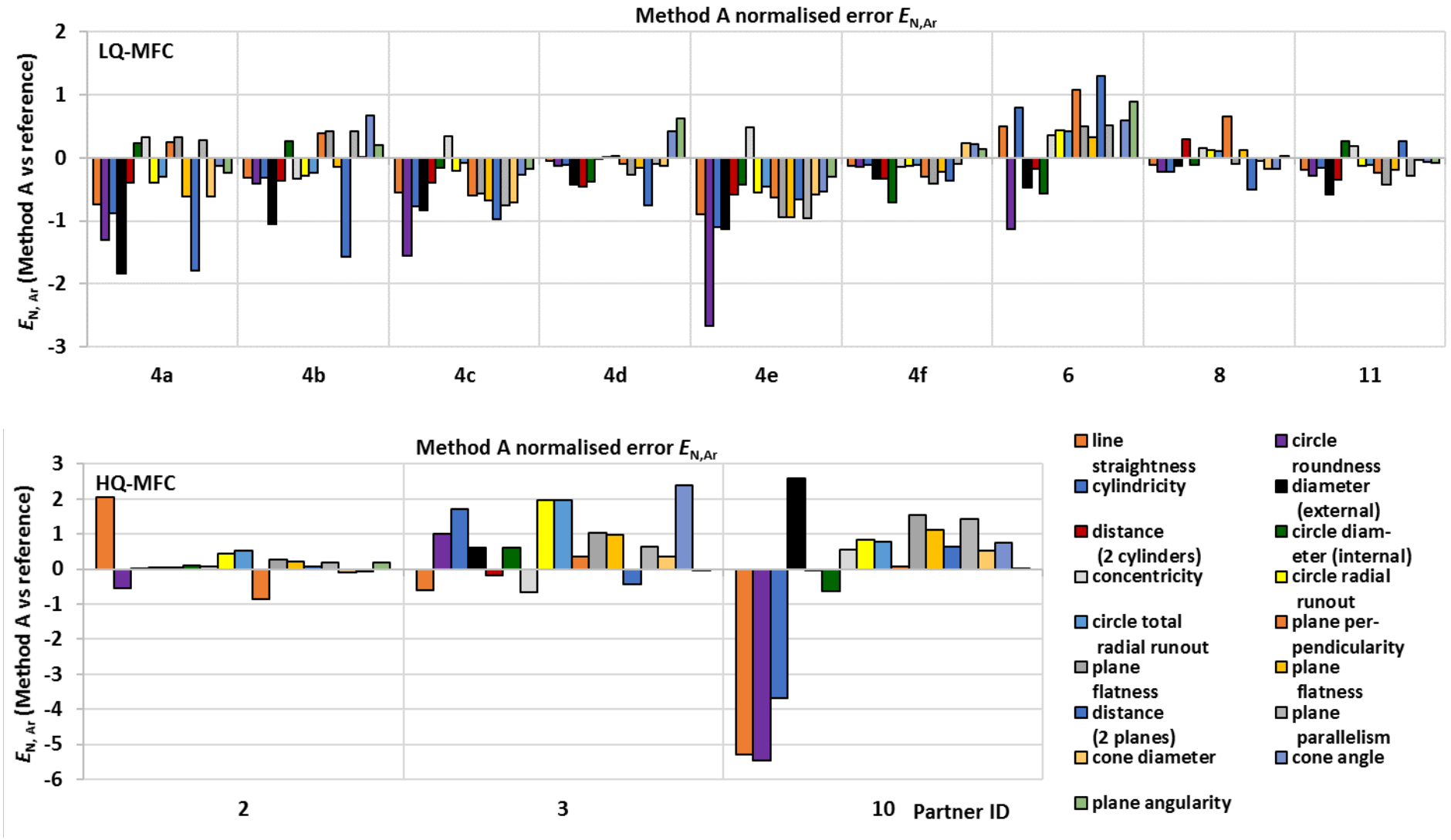


Figure 6-11: Normalised errors comparing method-A evaluated uncertainties to reference measurements.

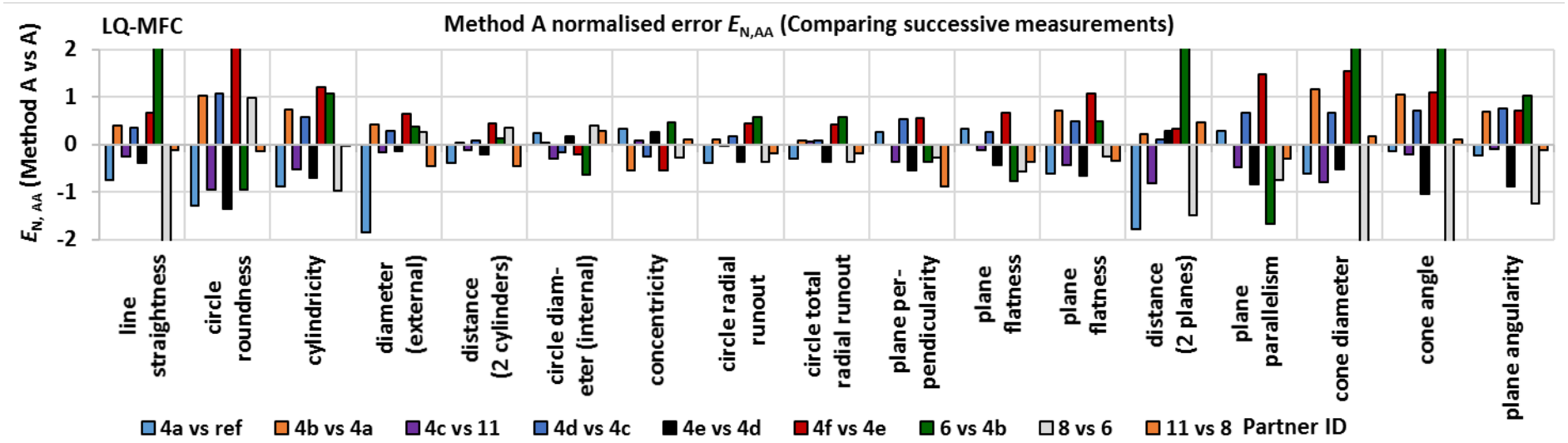


Figure 6-12: LQ-MFC normalised errors calculated for paired measurements using only method A uncertainties. (As discussed in the text, some values are outliers or measurement issues exceeding the scale. The E_N pass criterion is $|E_N|<1$, so no crucial information is lost.)

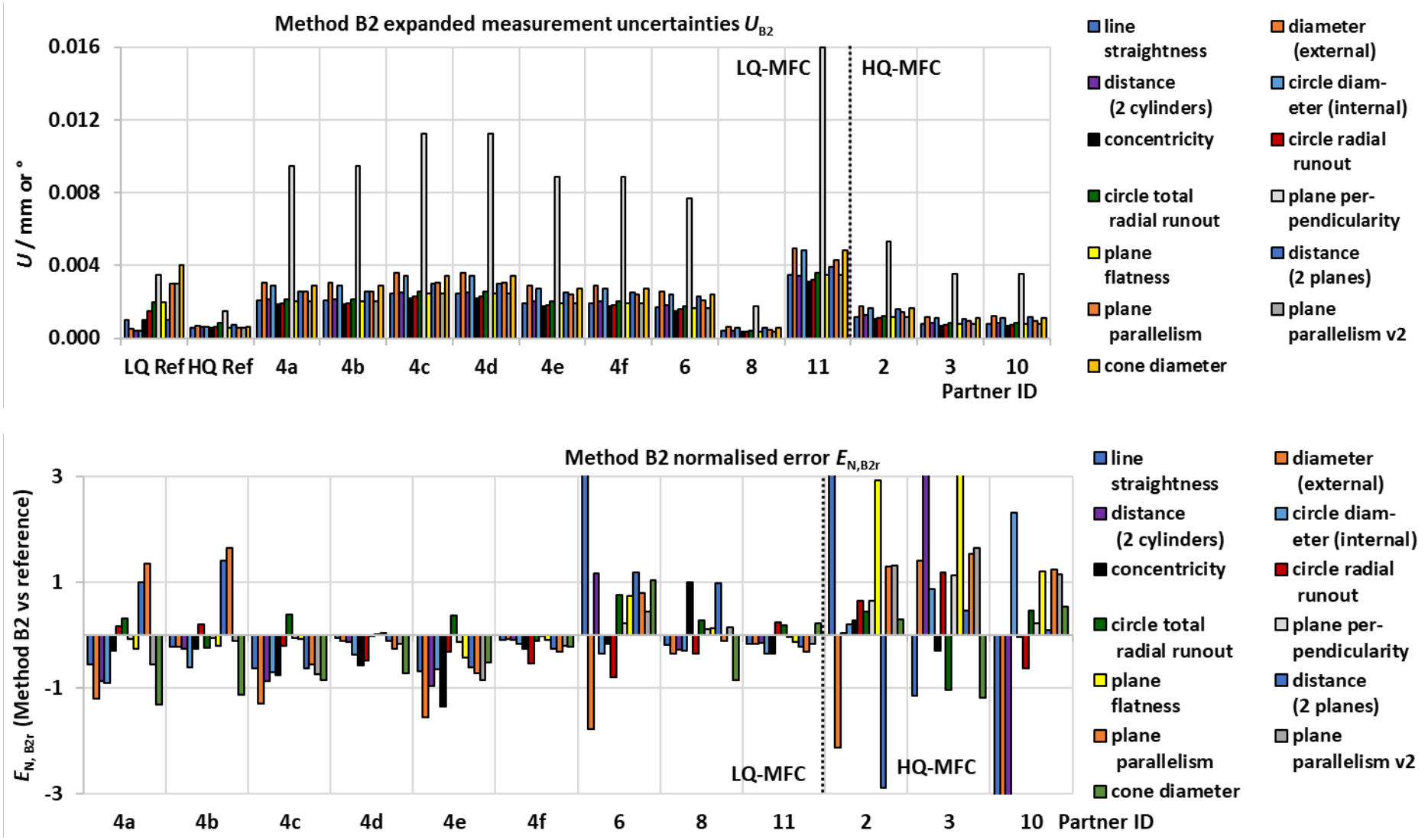


Figure 6-13: Method B2 uncertainties (top) and normalised errors relative to the reference (bottom) for both multi-feature checks.

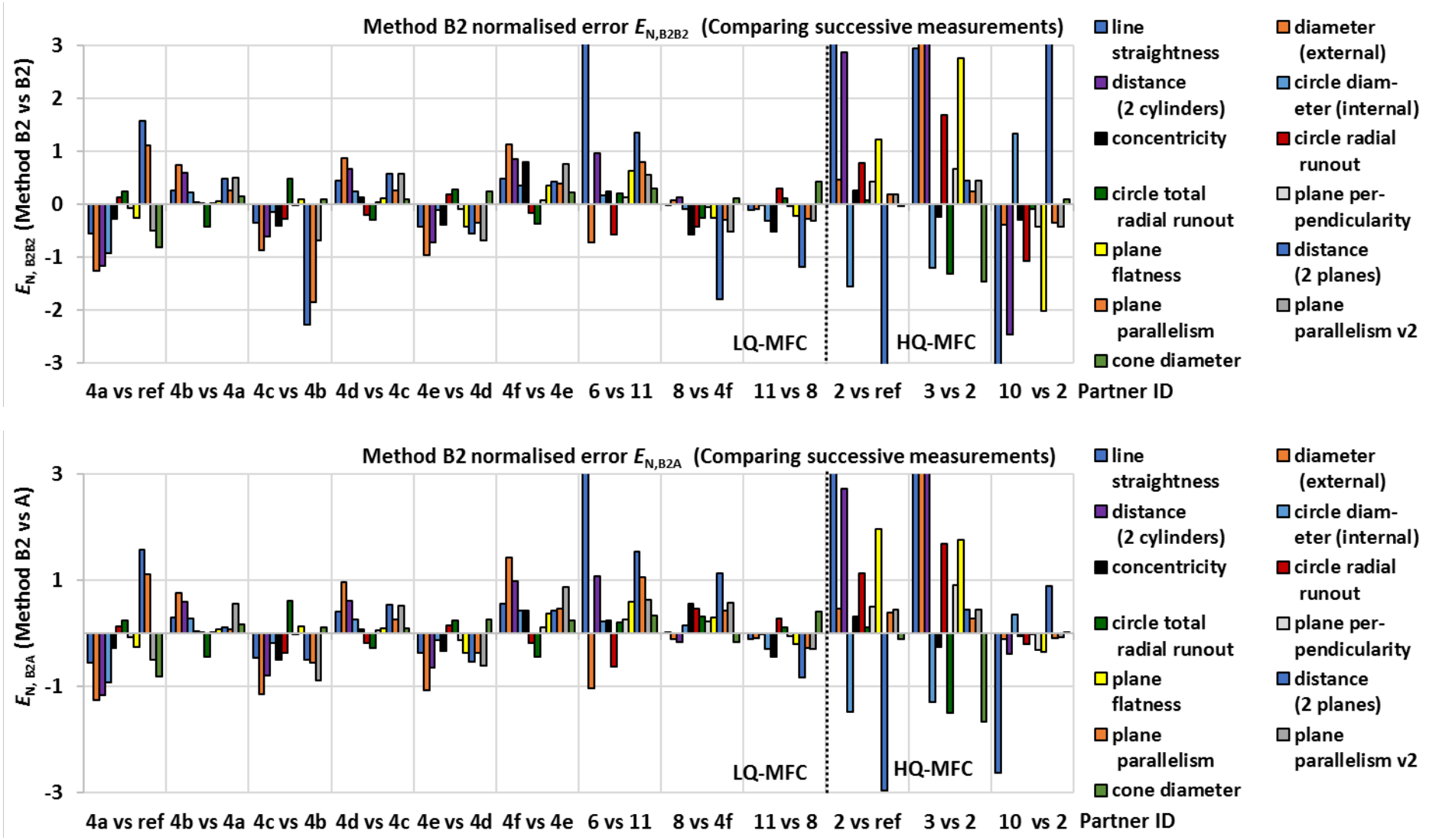


Figure 6-14: Normalised errors when comparing B2 results to each other (top) or comparing methods A and B2.

6.2.4 Method B2 uncertainties

The multi-feature checks were also evaluated according to the B2 sensitivity analysis method. Not all characteristics were evaluated – each feature type requires a specific model. Roundness, cylindricity, cone angle and angularity were not evaluated. The second flatness measurement (plane F) was omitted since the modelled uncertainty is the same as for the first (plane E). There are two models for parallelism with different definitions of the deviation, hence this characteristic is evaluated twice.

The uncertainties are shown in **Figure 6-13**. Within the datasets, the uncertainty varies little due to feature type, except for the LQ-MFC reference uncertainties and for plane perpendicularity. Since the datum (axis of C, a short cylinder) is relatively small compared to the plane E (forward face of the MFC), this characteristic has a higher uncertainty. The average uncertainty is approximately the same for all datasets from one workpiece. Set 8 is an exception, with very low uncertainties, which are, on average, even below the reference uncertainties. The average HQ-MFC uncertainty is slightly lower than LQ-MFC uncertainties.

6.2.5 Method B2 conformity testing

As with the connecting rod, normalised errors were calculated using measurement data from the round robin, obtained from the method A measurements. The normalised errors relative to the reference, $E_{N,B2r}$, are shown in **Figure 6-14**. Most LQ-MFC results pass, with the exception of set 6, which has issues with some features due to the small measurement volume of the CMM. The HQ-MFC has many more failures. As with method A, this is likely a consequence of the low reference and B2 uncertainties, which render the E_N-criterion much more restrictive. Similar results are obtained when comparing B2 results against each other: most LQ results pass, while roughly half of the HQ results fail. The results from the A-B cross-comparison ($E_{N,B2A}$) are again similar to the previous evaluation; LQ-MFC results mostly pass, HQ-MFC results show significant numbers of failures.

6.3 Hyperbolic paraboloid (freeform)

Nine measurements of the hyperbolic paraboloid are available. The reference uncertainty was a global VCMM estimate and is the same for all points.

6.3.1 Method A uncertainties

The method A uncertainty estimates for the individual surface points are shown in **Figure 6-15**. Overall, uncertainties are similar in magnitude and higher than the reference. Most measurements are very consistent across all points, but some show a certain degree of variability, e.g., set 6, ranging from 1.9 µm to 3.2 µm. The variability correlated to higher u_{geo} contributions, i.e., CMM geometry uncertainties obtained from the re-orientation of the workpiece, which becomes the dominant contributor. Other contributions are of similar magnitude and increase proportionally in sets with elevated uncertainties (e.g., set 4). Set with high overall uncertainty and variability feature both the dominant u_{geo} peak and increased contributions from other factors. **Figure 6-18** shows a selection of points from several of the datasets to illustrate these changes in the uncertainty contributions.

Such variations were not observed with the prismatic workpiece measurements. Therefore, the reason for the variability is unlikely to be the measurement setup, since at least some prismatic measurements – typically run on the same kinds of setups – should have shown something similar.

The root cause could stem from data processing and data sparseness. The data used should essentially be raw data, and we would therefore expect to see a certain degree of variation from the hardware (e.g., positioning errors), the environment (e.g., temperature-related) and random noise. The data is also very limited, ideally one point per measurement. Prismatic measurands are derived from ideal geometries that are fitted to point clouds. To some extent, the fitting process filters out random variations of individual points.

Because of the sparse data, any changes to the data should also affect the result more significantly. Depending on the hard- and software, the data could be pre-filtered by different degrees, resulting in cleaner, more stable results. Where no or fewer filters are in place, the variability is more visible. Some filters are intrinsic to a system or system design. Others may be part of “black box” commercial software where the user is either unaware or unable to disable them.

However, the variation observed in the datasets is small relative to the average uncertainty of the data. Averaging the uncertainties of the same feature type into a single value would simplify real-world application of the method A estimate. When this is done, care must be taken to ensure the combined uncertainties actually fit together. For point measurements, this should be uncontroversial. However, anisotropic features could be dangerous to combine. For example, lengths measured in different directions might have different associated uncertainties due to differences in the CMM axes. An empirical uncertainty estimate should detect these differences, which must not be conflated with the random variation discussed here.

The last thing to note is that the uncertainties are uniformly higher than the VCMM reference estimate. This was also the case with prismatic measurements and suggests that method A is a more pessimistic uncertainty estimator than the VCMM.

6.3.2 Method A conformity testing

The point-wise normalised errors are shown in **Figure 6-17**. With very few exceptions, measurements conform with the reference measurement. **Figure 6-16** shows that, on average, point deviations are small relative to the uncertainty of the reference measurement. Since the measurements vary little, method A is thus consistent and comparable with the VCMM. However, the normalised errors tend to be fairly small, which may indicate that method A is actually overestimating the measurement uncertainties. When the uncertainties used to calculate E_N are too large, the value necessarily becomes very small. Most E_N-values are below 0.5, and the mean normalised error for all data including failures is only 0.34.

Similarly, when comparing method A measurements directly to each other, the normalised errors are again quite low. Most values are below 0.5, and the average for $|E_{N,AA}|$-values is 0.4.

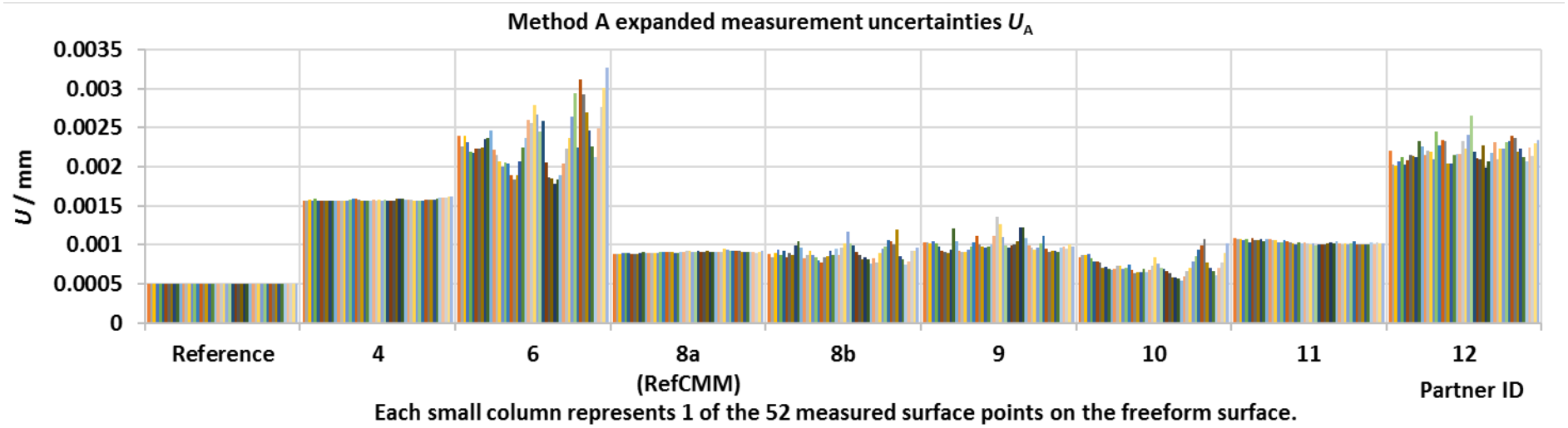


Figure 6-15: Hyperbolic paraboloid method A uncertainties. The reference uncertainty was determined by VCMM and applied equally to all points.

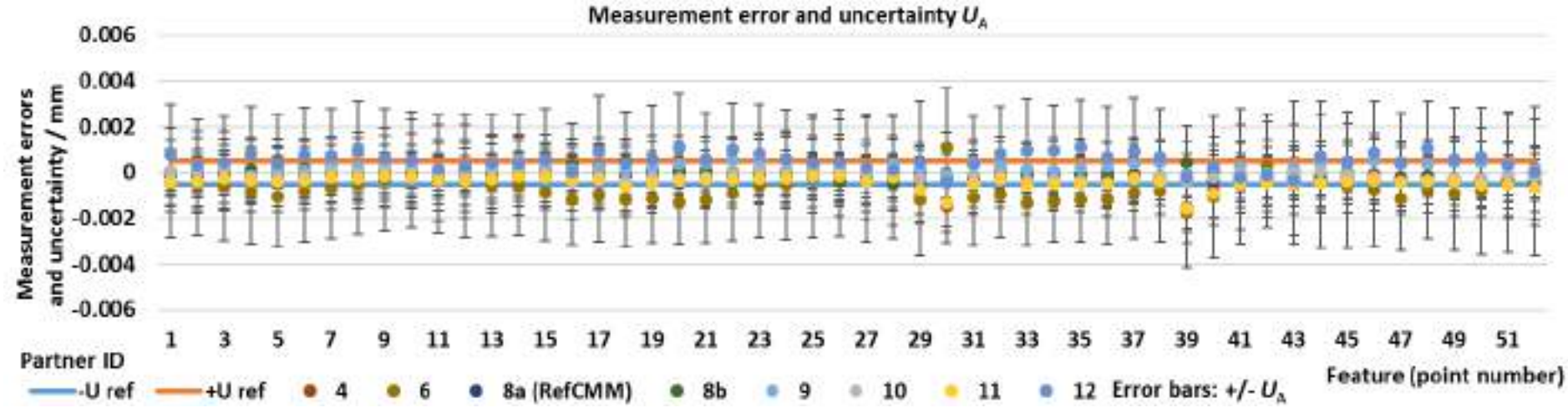


Figure 6-16: Deviations from the reference measurement.

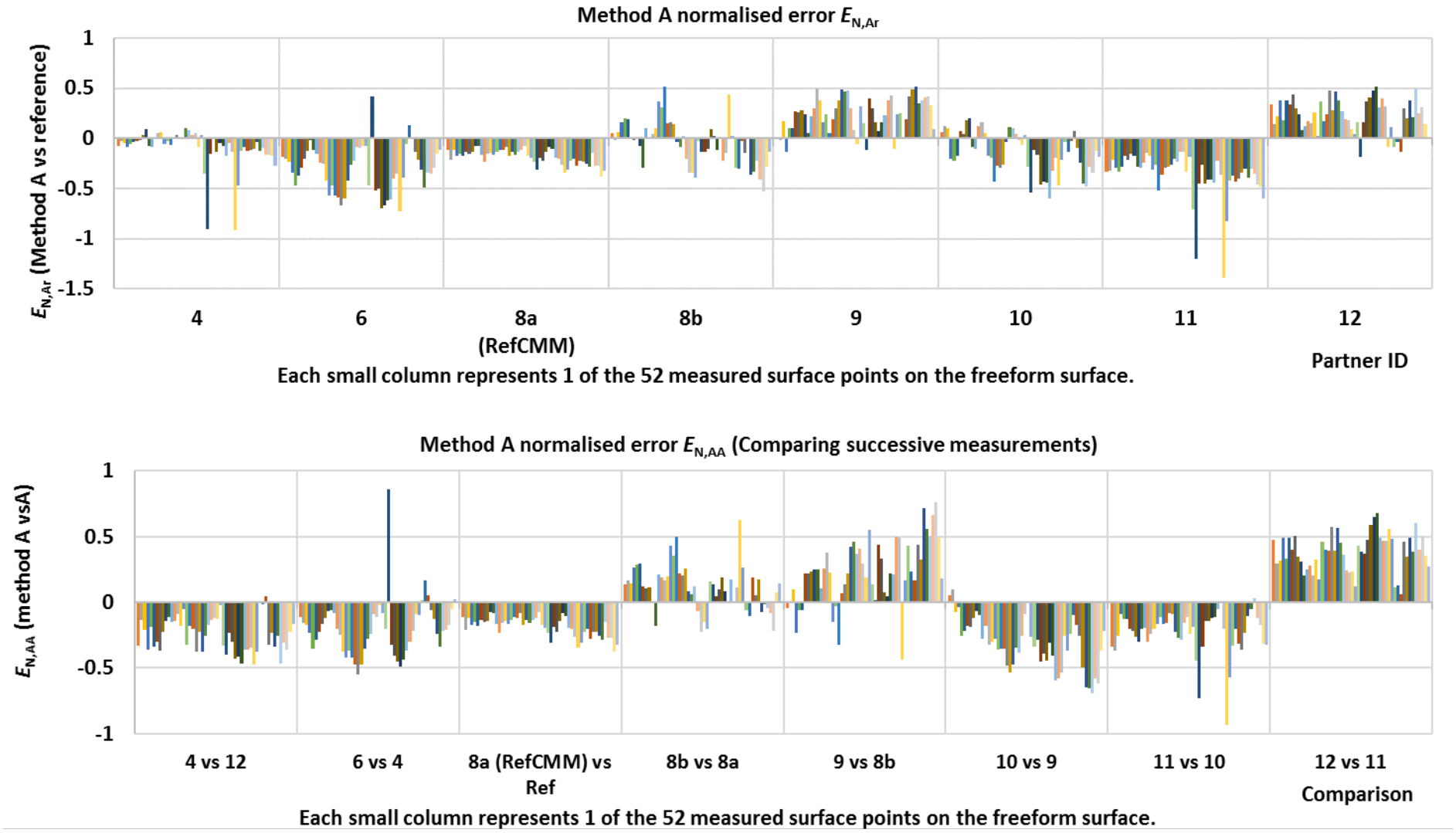


Figure 6-17: Hyperbolic paraboloid method A normalised errors calculated against the reference (top) and previous measurements. (bottom)

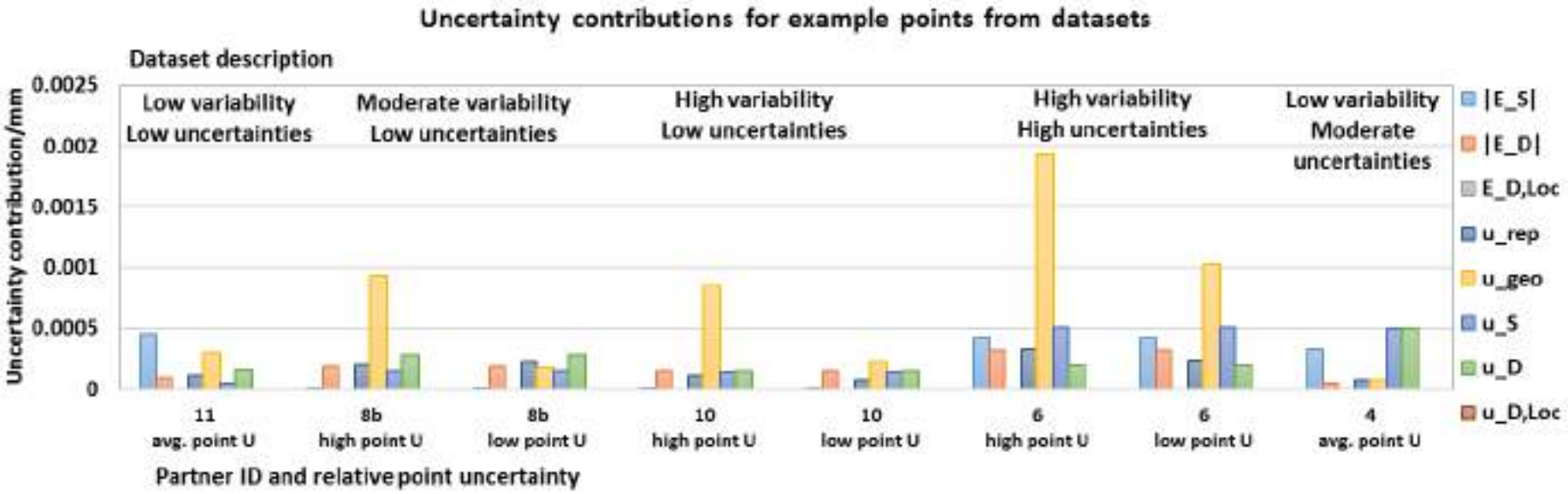


Figure 6-18: Uncertainty breakdown of representative points taken from different datasets. The variation of U_A within datasets correlates with relatively high u_{geo} contributions, which often lead to higher combined uncertainties as well. Only when variability is low are higher uncertainties associated with other contributions (set 4). Average, high, and low *U* refer to the relative combined uncertainty of the selected points.

6.3.3 Method B1 uncertainties

The B1 uncertainties are shown in **Figure 6-19**. Not all datasets could be included as the necessary information (e.g. MPEs) was not always available. Due to time constraints, the evaluation could not be fully implemented; the uncertainties presented do not yet consider any correlations between the contributing factors.

Since the model is a priori and applied to the same feature type (point deviations), the uncertainty values are uniform in each dataset. Variations are on the order of 1 nm. The datasets range from 0.3 µm to 2.7 µm, compared to a reference uncertainty of 0.5 µm. Most B1 uncertainties are higher than the reference (from VCMM), except for 8a, (0.3 µm), which is also the CMM used to acquire the reference data.

6.3.4 Method B1 conformity testing

The conformity tests were quite successful. Comparing B1 to the reference ($E_{N,B1r}$), most points passed easily with a failure rate of 1.6 % overall. Intercomparison of B1 results ($E_{N,B1B1}$) and intercomparison to method A ($E_{N,B1A}$) only include two failures each (0.5 %). (**Figure 6-22**)

6.3.5 Method B1 direct validation

Another statistical approach to testing method B1 works by comparing results obtained with a priori and a posteriori model parameters. The procedure is described in detail in the method development report (deliverable D2, ch. 10). [2] This is a short summary of the procedure and results.

Normally, the model parameters are estimates **a priori**, that is, based on previous knowledge and experience, since no actual data on the measurement task is available. In this case, there is a lot of data from method A, both repeat measurements and measurements in different orientations. This information was used to estimate **a posteriori** model parameters. The a priori and posteriori statistical parameters were then compared to check their consistency. This is quantified by $\widehat{\sigma_0}$, which represents the combined effects of all model parameters (indicated by the subscript; the ^ indicates that is an a posteriori estimate). For good agreement, the value should be equal to 1. Lower values indicate an overestimate of the uncertainty by the a priori model. The approach is limited by the degrees of freedom of the available data, i.e., the number of orientations – four, due to the method A measurement strategy – and the number of times a feature type was measured. The latter depends on the artefact. The connecting rod and hyperbolic paraboloid were both investigated. On the rod, each type is only present once or twice (e.g., distance, diameter), hence there are few degrees of freedom and $\widehat{\sigma_0}$ is distributed across a broad range. The paraboloid measures fifty-two point deviations – one feature type – and thus has more degrees of freedom for the analysis.

The results vary greatly between datasets, and it is difficult to arrive at a definite conclusion, particularly for the connecting rod with few degrees of freedom. However, several a posteriori results are quite close to the initial estimates. (The numbers were omitted from this report as associating them with specific measurements or partner IDs would compromise the anonymity of the round robin. See [2], chapter 10.)

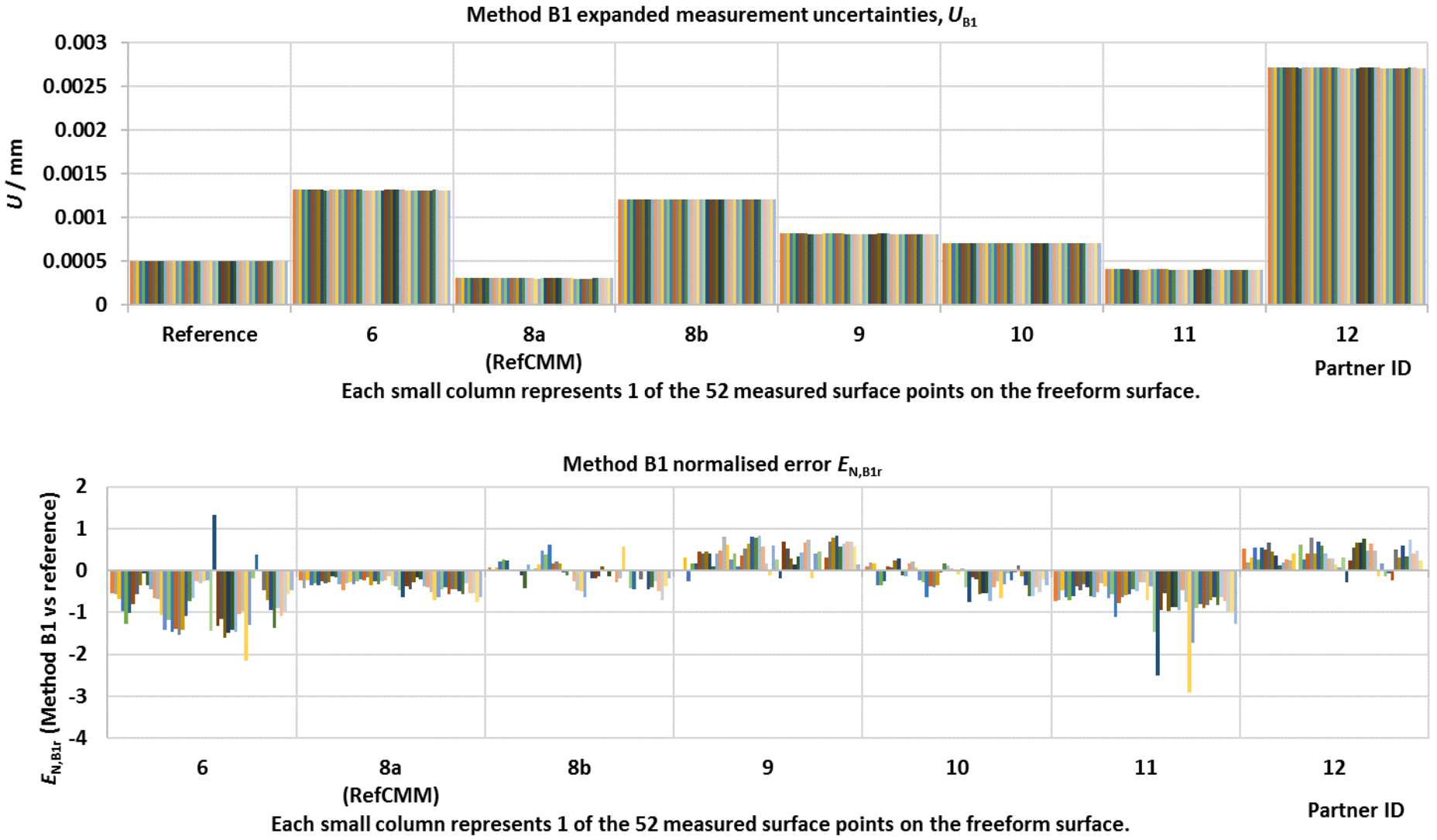


Figure 6-19: Method B1 uncertainties (top) and E_N-values relative to the reference (bottom) for the hyperbolic paraboloid.

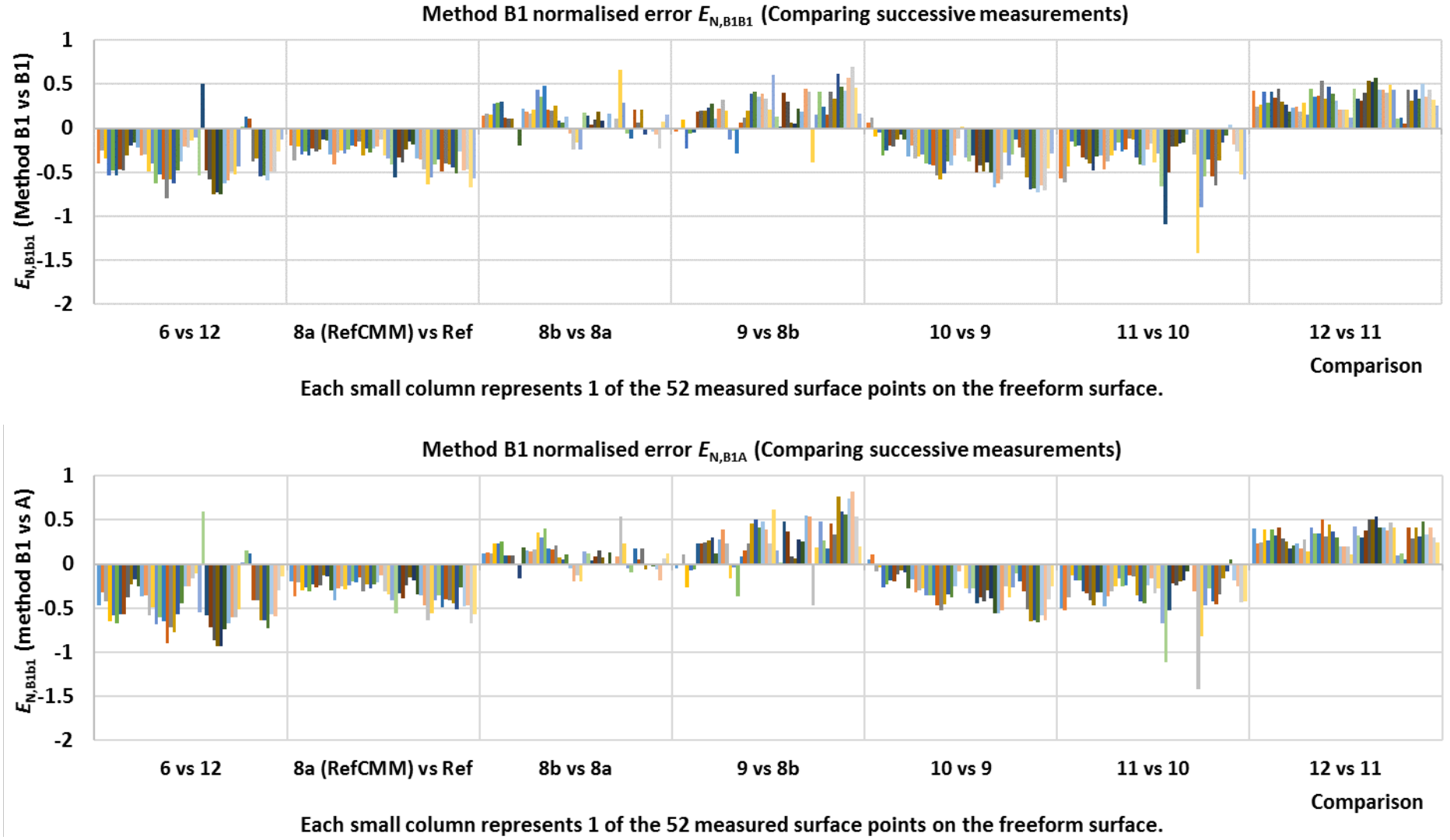


Figure 6-20: Comparison of hyperbolic paraboloid results to EUCoM methods. Top: B1 vs B1. Bottom: B1 vs A.

6.4 Involute gears (freeform)

Nine measurements of the smooth, and eight measurements of the sinusoidal involute gear are available.

The involute flanks were measured with approximately 1000 points, interpolated, and then evaluated for 995 sample points. For each involute there are two sets of results: an evaluation with a datum, where the involute is registered using other features of the workpiece, and without a datum. Here, instead of registering the workpiece, the measured profile is fitted against the nominal profile by least squares before computing the point deviations. The deviation was determined for each individual nominal profile point. The nominal profiles represent the ideal parts, not the actual manufactured workpieces. The nominal points were calculated from the equations for an involute and, for the sinusoidal profile, the (ideal) sine components.

An evaluation similar to typical profile evaluations, which looks at the minimum and maximum deviations to estimate an overall profile deviation, was also tried. However, the results tended to be very chaotic. In some examples, the first few points of the profile showed very high and varied distributions. This could be due to artefact design (the profile starts close to a corner-point with another feature, where the surface is difficult to reach) or interpolation issues (in case of low density data). The evaluation based on max/min also tends to pick outliers from the repeat measurement and combine them. While these problems could be addressed by filtering, this would affect the data in toto and influence the validation as well. Hence the profile evaluation was dropped to avoid potential issues arising from data processing.

Due to software incompatibilities, the measurement instructions had to be re-written by partners 5 and 7. This included recalculated nominal profiles and is one potential reason for the high profile deviations in parts of their measurements.

6.4.1 Smooth profile (with datum)

6.4.1.1 Method A uncertainties

The complete profile results are shown in **Figure 6-22**. As noted, some datasets start with very high profile deviations. Other than this, profile scans tend to be in good agreement. The method A uncertainty tends to be fairly consistent across the profile. Most datasets lie in the range of 1-2 µm, similar or higher than the reference uncertainty (~1 µm).

Some datasets appear noisier than others, which can be attributed to the different systems and measurement environments. Set 14 stands out with occasional peaks and, elevated plateaus or valleys. The apparent smoothness of set 7 is a side-effect of a relatively high degree of interpolation. While most measurements used close to 1000 sample points, there are only 100 in the set 7 measurements. These were still interpolated, sampled evaluated in the same way as any other, with the initial portion of this profile affected by interpolation issues due to ten-fold lower data density.

6.4.1.2 Method A conformity testing

Almost all of the E_N-values based on the reference uncertainty pass (**Figure 6-23**). The first section of set 7 fails the test on account of the large deviations evaluated for this part of the profile. Set 5a is much more variable than any of the other sets.

Most datasets also pass the test when comparing two method A results. The results outside the acceptance range for E_N (-1 to 1) are related to other issues. Sets 5a vs 7 and 7 vs 8 are again affected by the high deviations measured at the start of the profile in set 7.

These graphs also illustrate the variability of the uncertainty estimates and measurements. Since $E_{N,AA}$ folds measurements and uncertainties together, the variation is amplified and more visible, giving the curves a noisy or sometimes wave-like appearance. We can explain this by

considering the feature type and the individual uncertainty contributions, an example of which is shown in **Figure 6-21**. The contributions determined from the sphere and length standards are the same for all points. Repeatability (u_{rep}) and geometry (u_{geo}) are both derived from profile point deviations. The relative magnitude of u_{rep} is small. It lies in a narrow band and hence does not add much variation to the overall uncertainty. u_{geo} is a lot more variable and has much more impact on the combined uncertainty.

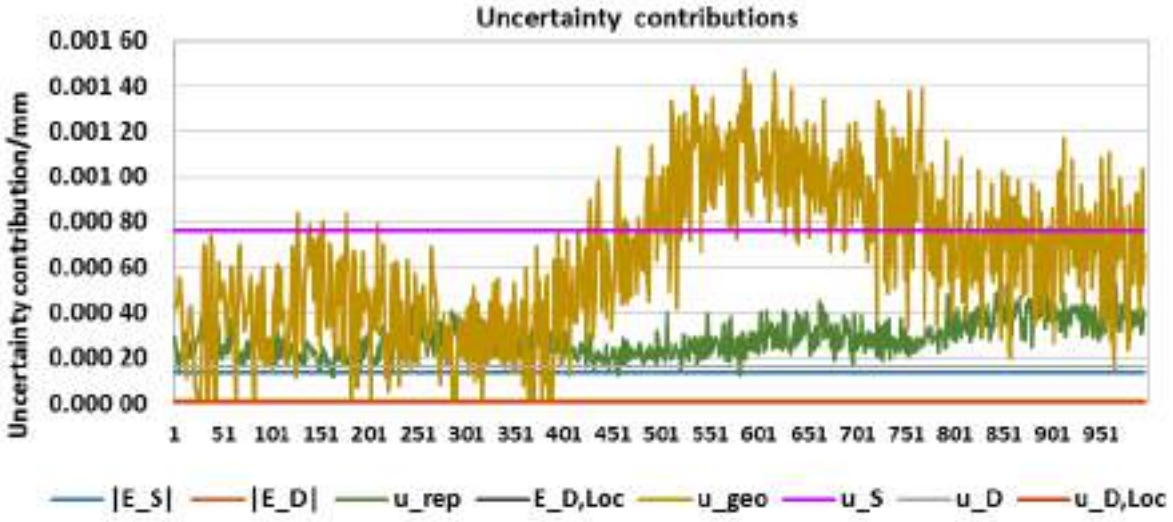


Figure 6-21: Uncertainty breakdown for the smooth involute from dataset 11a.

6.4.2 Smooth profile (No datum / best fit)

The results for the fitted profiles are shown in **Figure 6-24**. The results are very similar to the evaluation as a profile with datum, i.e., the registered parts. Because of the least squares fit, the apparent profile deviations are much lower than those of the registered parts. The registration establishes a fixed position in space that is not directly related to the nominal profile position. Nominal and actual registration features may differ, due to differences between the part design a manufactured part. This would result in systematic errors, such as offsets or alignment errors. The best fit minimises the overall error and hence the measured deviations. The majority of fitted datasets are in close agreement, however, so the outcome of the best fit is consistent.

Profile 7 shows the largest disagreement. As discussed, this profile is heavily interpolated. In addition, the first segment was omitted as the large deviations would have affected the least-squares fit. The omission and data interpolation also affect the best fit to some extent and thus still result in the noticeable differences to the other profiles.

The measurement uncertainties follow the same pattern as the datum evaluation. This includes the noisy appearance and prominent features, such as the higher peaks in set 11a, which appear in the same places along the profile. The uncertainties are in the same range (1-2 µm) as before and mostly close to, but slightly higher than the reference (0.8 µm).

The normalised errors against the reference, $E_{N,Ar}$, (**Figure 6-25**) are reduced relative to the registered profile evaluation. This can be attributed to the best fit reducing the measured deviations (smaller numerator when calculating E_N) rather than any change in the uncertainties (unchanged denominator).

The normalised error, $E_{N,AA}$, obtained when comparing method A results is the same as with the registered profiles. The best fit reduces point deviations systematically for all datasets. Therefore, when comparing two datasets, the difference between the measured values is

approximately unchanged, and the normalised error stays the same. Note that the datum and no datum evaluations paired the same datasets for the comparison.

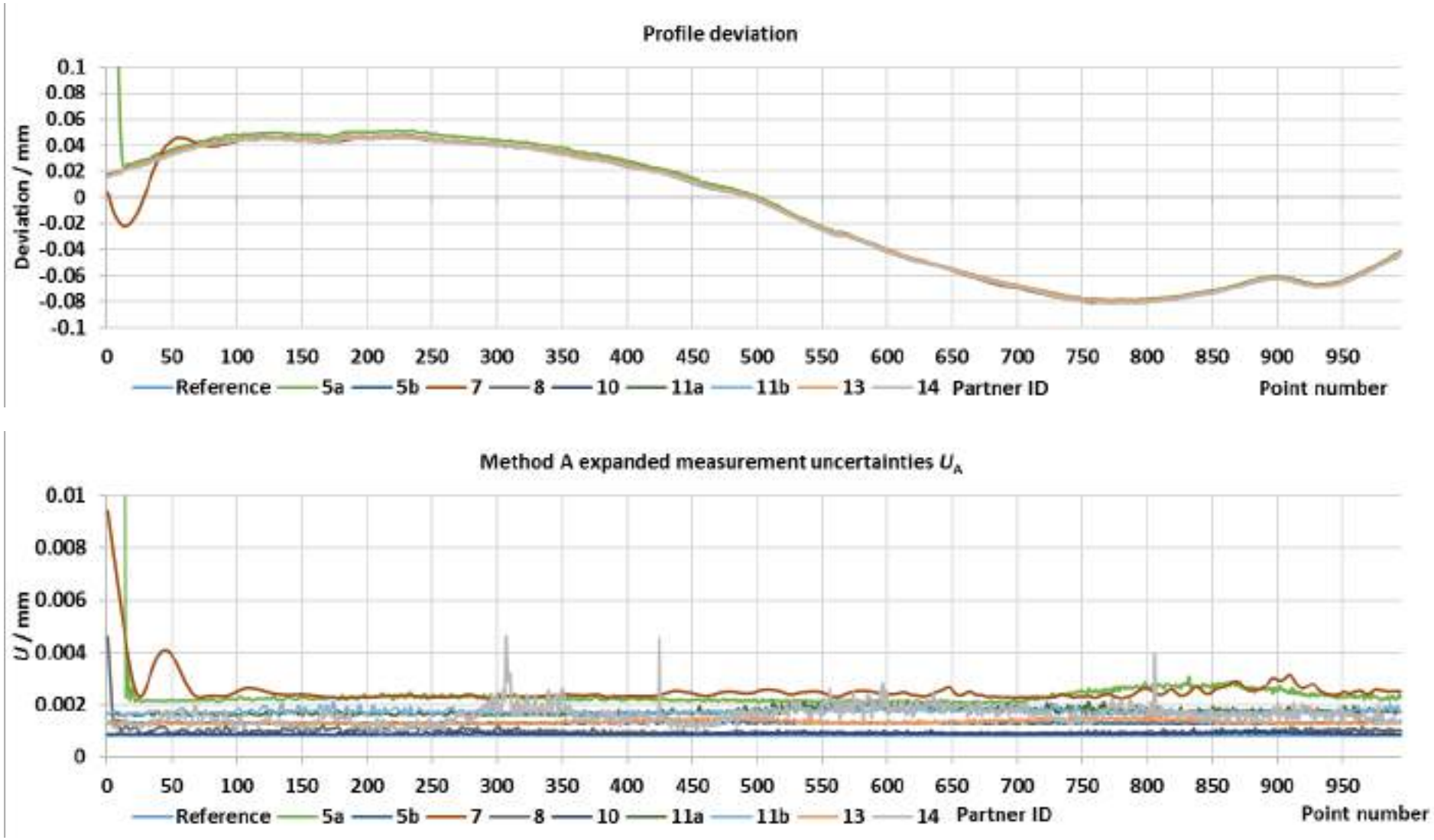


Figure 6-22: Deviations (top) and Method A uncertainty (bottom) for all points of the full smooth involute profiles (with datum).

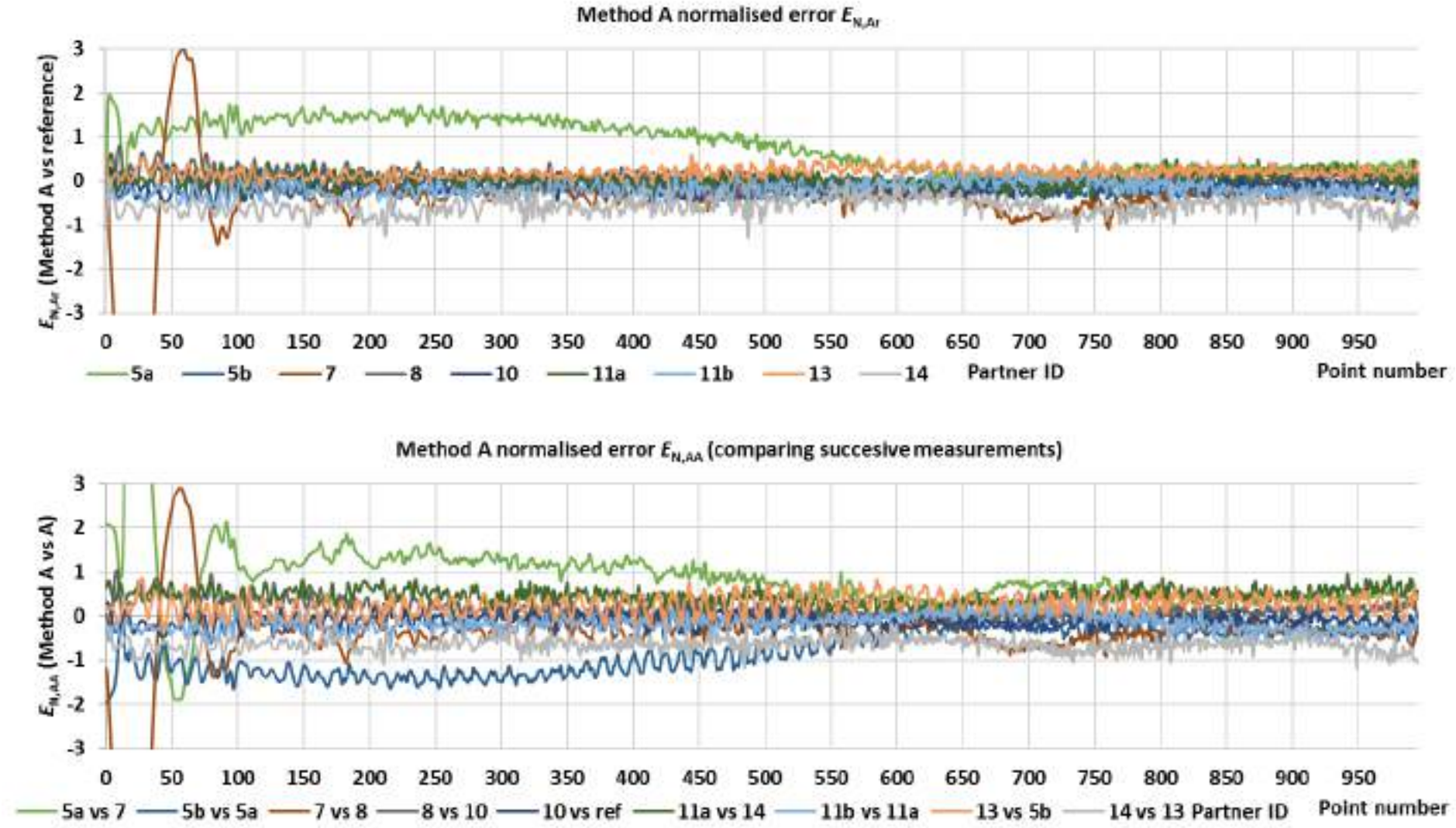


Figure 6-23: Method A normalised errors relative to the reference (top) and other method A evaluations, for each profile point (with datum).

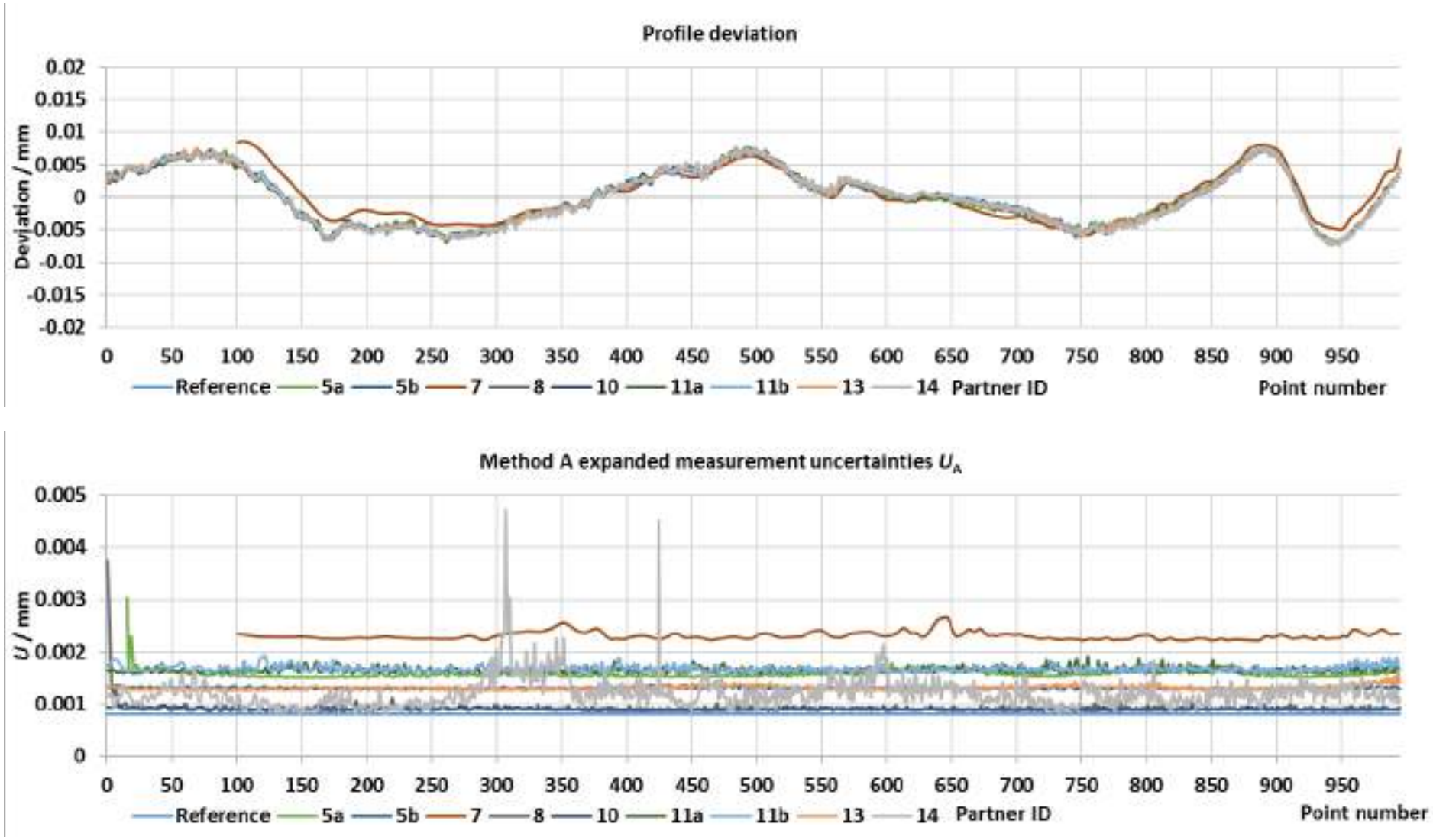


Figure 6-24: Deviations (top) and Method A uncertainty (bottom) for all points of the full smooth involute profiles (best fit).

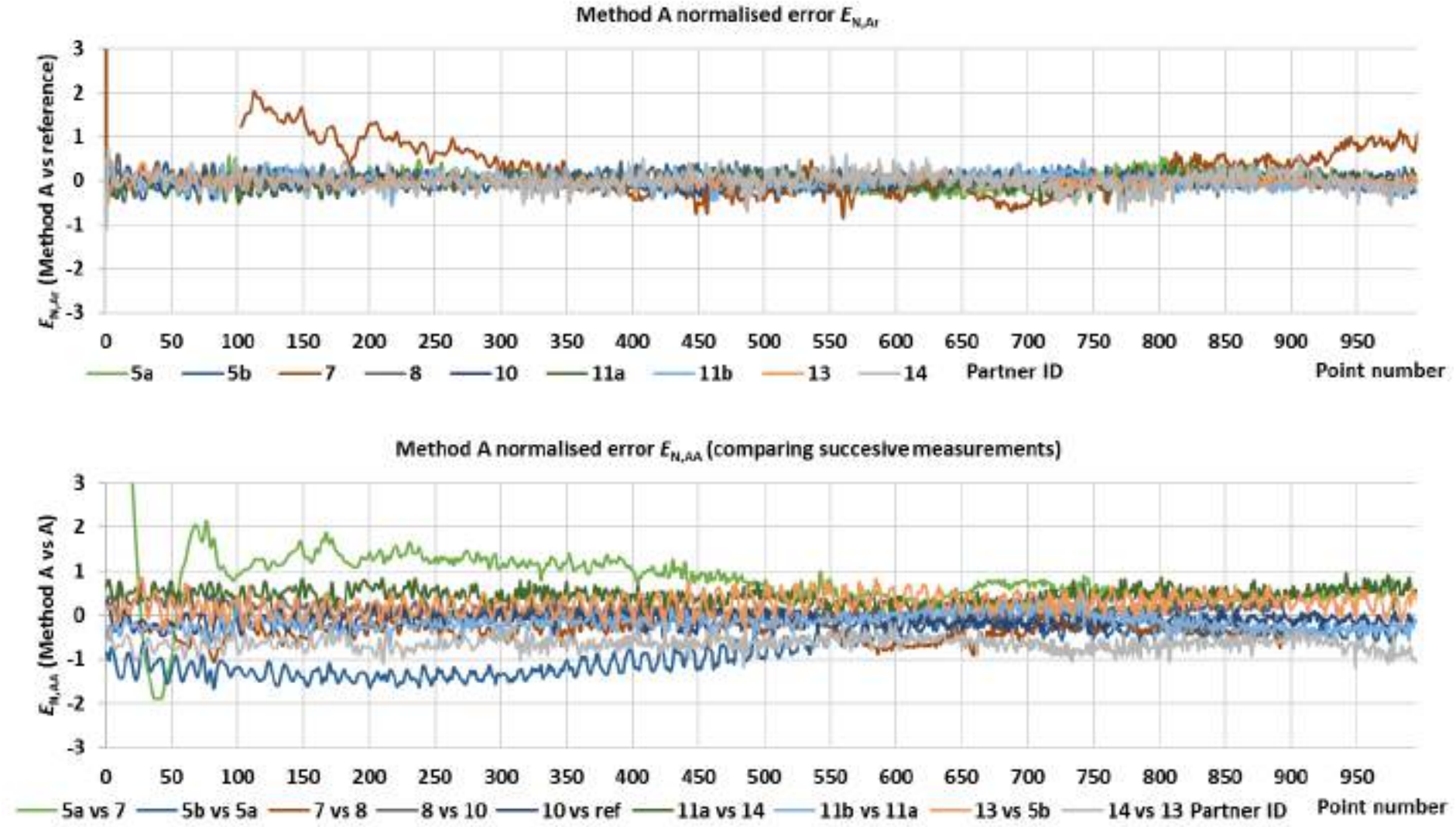


Figure 6-25: Method A normalised errors relative to the reference (top) and other method A evaluations, calculated for each profile point (best fit)

6.4.3 Sinusoidal profile (with and without datum)

The results for the sinusoidal involute track closely with the smooth involute. The profile deviations are shown in **Figure 6-27**. The individual measurements are in close agreement, as highlighted by the inset graph on the top chart. As with the smooth part, the registration introduces a slight, but reproducible mismatch of registrations between the real part and the nominal part on which the reference values are based. The misalignment also places the real and nominal sine patterns out of phase. Thus, when calculating the deviations, subtracting the wave patterns creates another, similar wave pattern, which can be seen in the deviation plots. The pattern persists in the “no datum” profiles. The best fit aligns the full profiles, rather than the fine-grained sine with an amplitude of ~20 µm, which in any case are imperfect measurements of the nominal profile they are fitted to (c.f. **Figure 3-8**, **Table 3-2**). Some of the profiles also show the same high deviations as seen with the smooth involute.

The method A uncertainties still range from 1 µm - 2 µm both with and without datum (**Figure 6-28**). The normalised errors against the reference (**Figure 6-29**) or other method A results are (**Figure 6-30**) mostly acceptable. The variability of the uncertainties and normalised errors is again due to the measurement-induced variation of the u_{geo} and u_{rep} contributions. In this example (set 14, **Figure 6-26**) the magnitude of u_{geo} and u_{rep} are roughly equal for both the registered and fitted profiles.

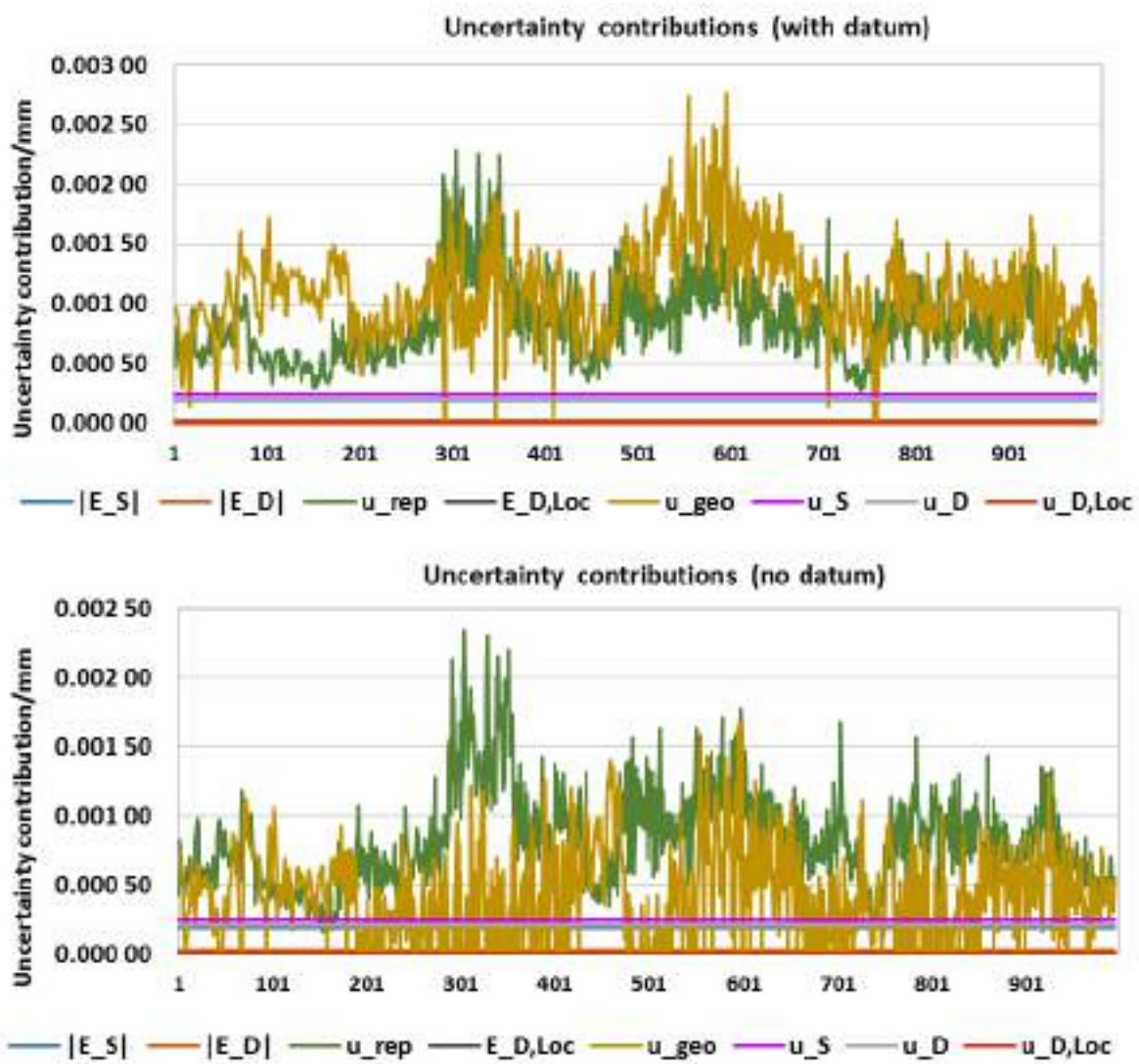


Figure 6-26: Uncertainty breakdowns from set 14 for both wave profile evaluations.

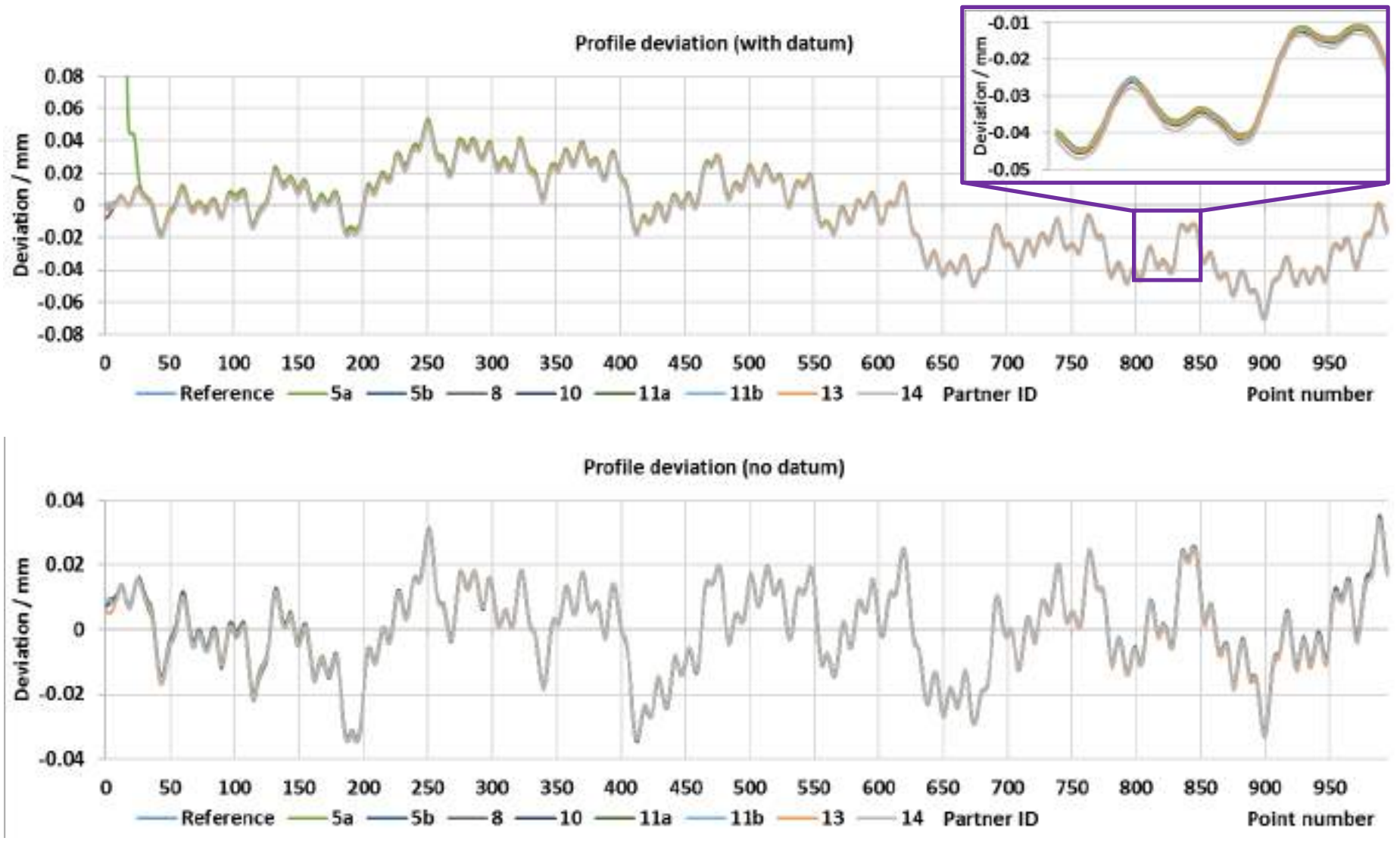


Figure 6-27: Sinusoidal involute profile deviations with (top) and without datum (bottom).

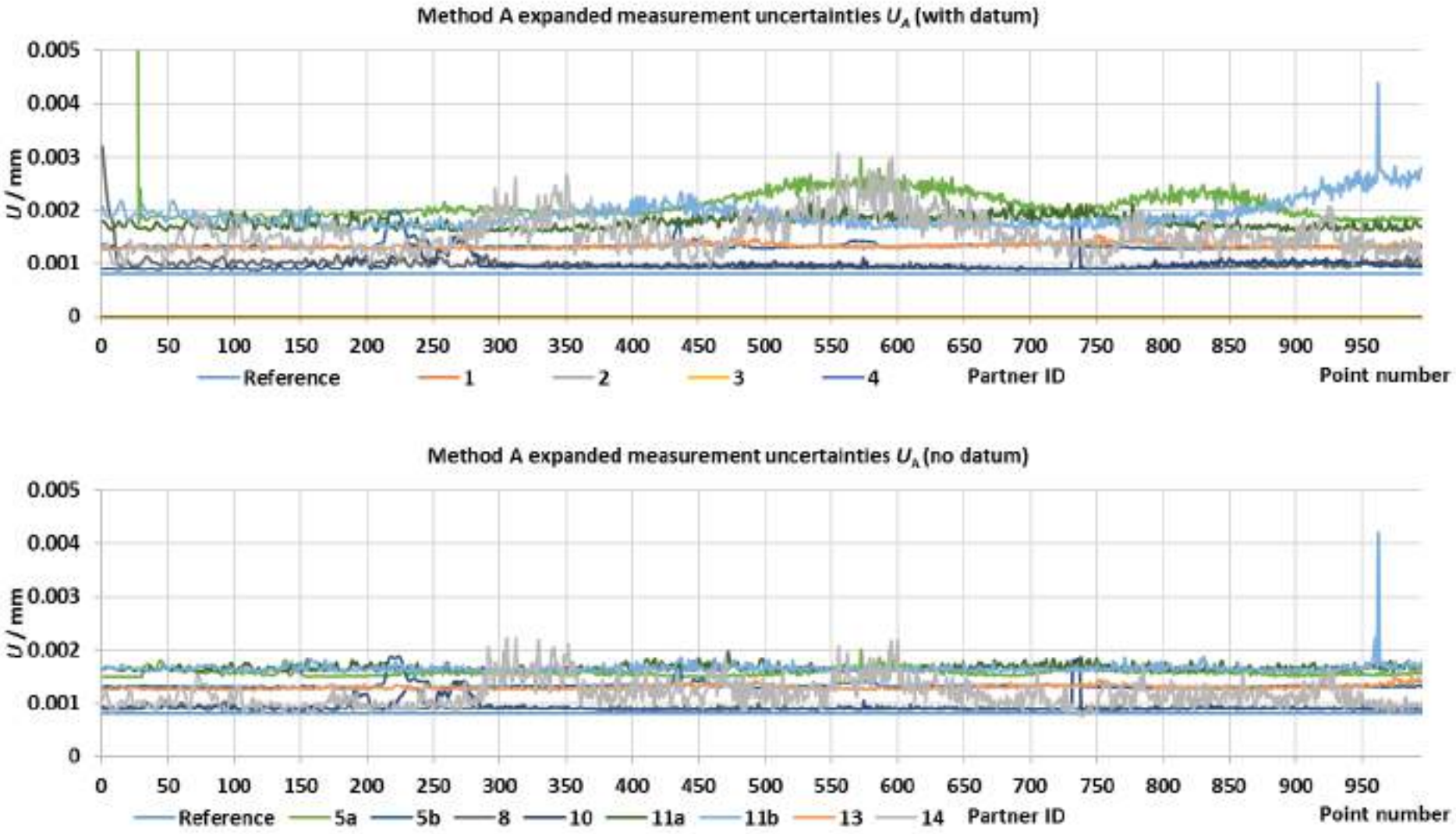


Figure 6-28: Method A uncertainties for the sinusoidal involute.

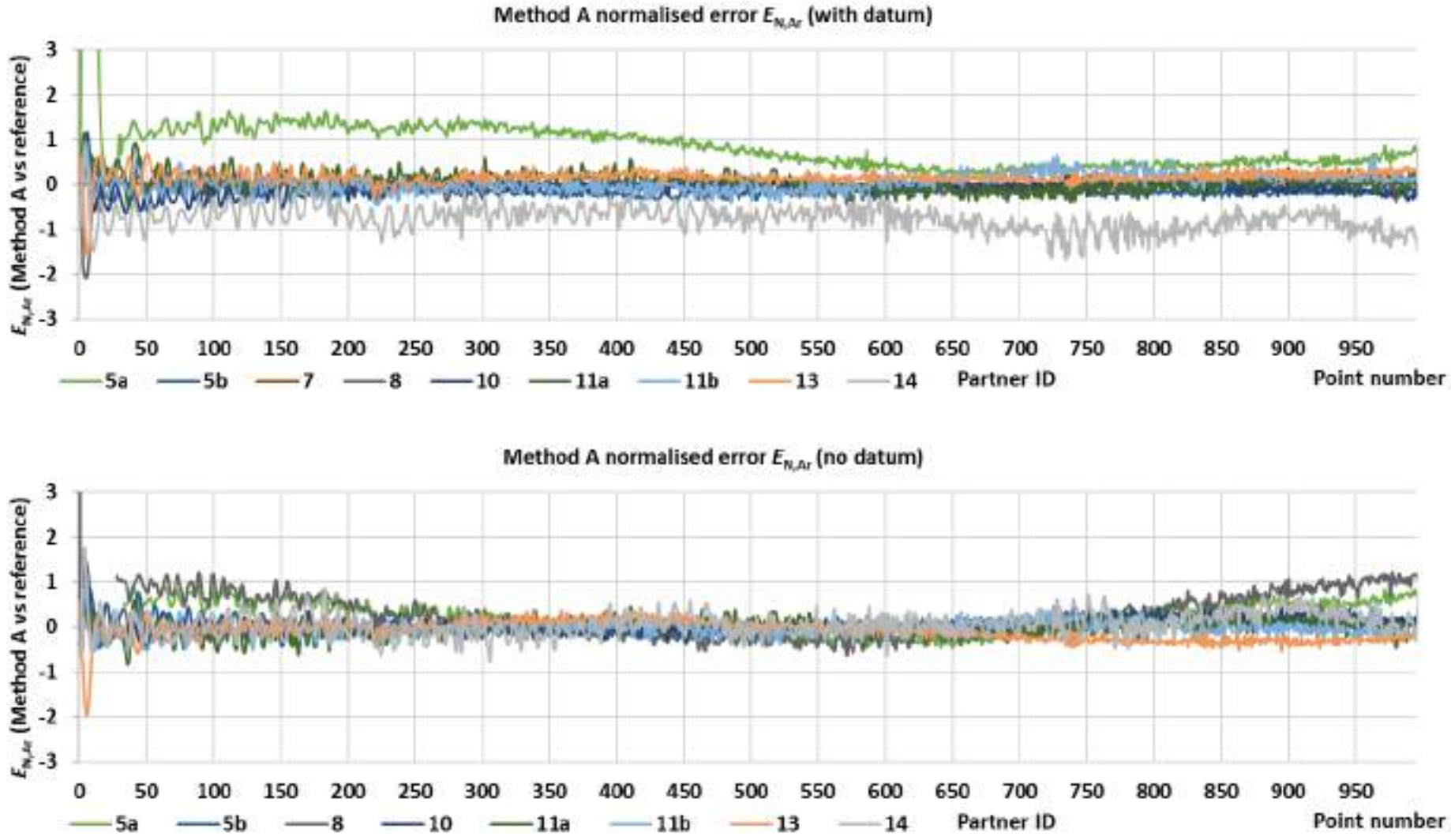


Figure 6-29: Normalised errors comparing method A results for the sinusoidal involute with the reference data.

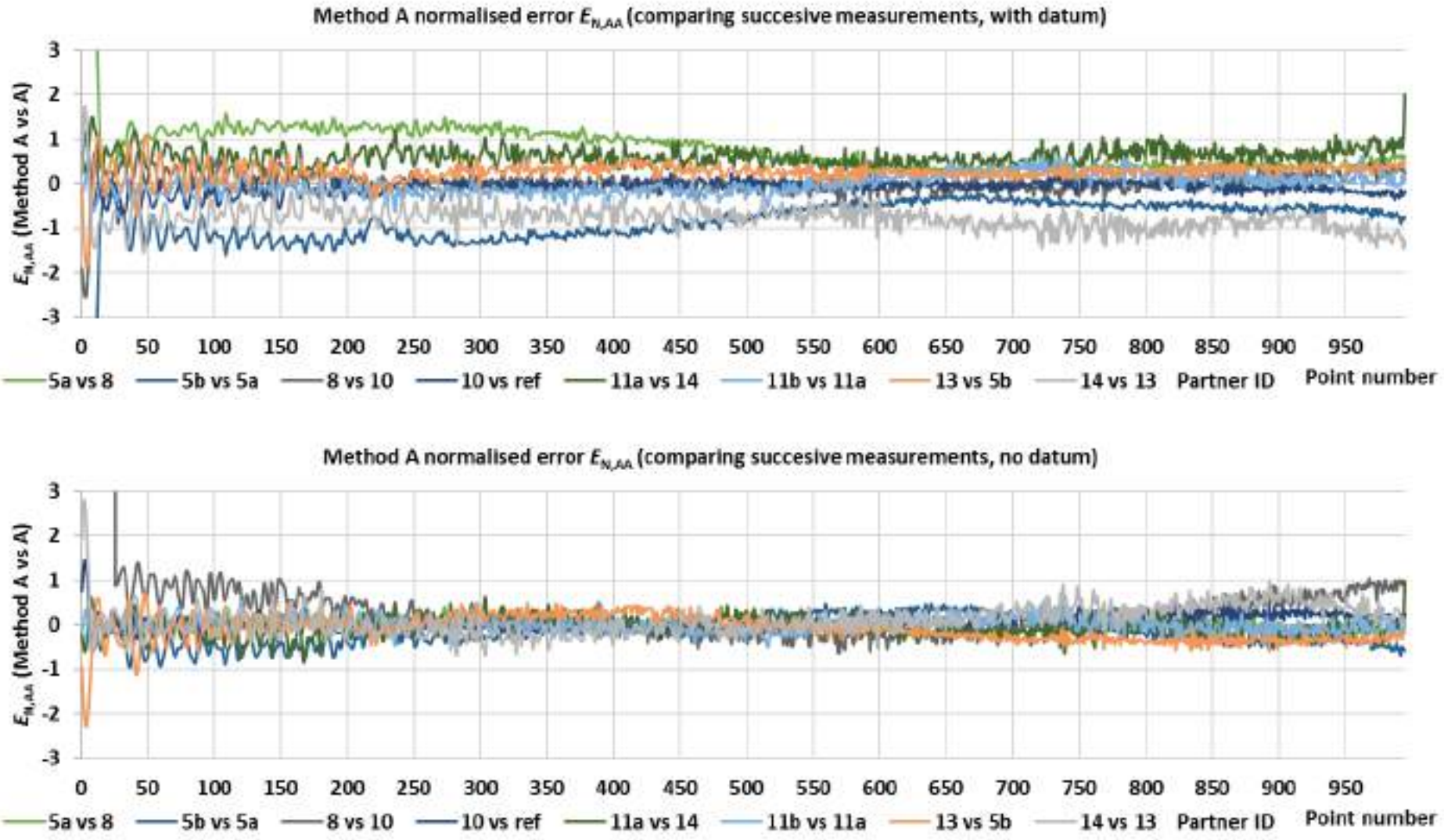


Figure 6-30: Normalised errors comparing method A results for the sinusoidal involute against each other.

7 Summary for prismatic geometries

A total of twenty-one independent measurements of the connecting rod and two multi-feature checks were evaluated. A tabulated summary is shown in **Table 8-1** and **Table 8-2**. The connecting rod results are somewhat mixed due to unexpected problems with the artefact itself, discussed in section 3.3.1. The connecting rod was evaluated with methods A, B1 and B2. The two multi-feature checks were evaluated with methods A and B2.

7.1 Method A: a posteriori

Method A uncertainties followed the same trends for all workpieces (including freeforms). Within a dataset, uncertainties of different features tended to be similar.

Connecting rod normalised errors were too high in many cases. This is probably due to the problems with the measurands, but issues with the uncertainty estimates cannot be ruled out.

The multi-feature checks had no such problems. On the low-quality multi-feature check (LQ-MFC), most measurands resulted in acceptable normalised errors, both relative to the reference (12 failures, 8 %) and other method A results (27 failures, 18 %). The higher $E_{N,AA}$ failure rate of the LQ-MFC could also mean that U_A is actually too small. E_N would then be more difficult to pass when two U_A uncertainties are compared, instead of U_A and U_r. Further analysis is required. The HQ-MFC performed worse, with respective failures rates of 26 % ($E_{N,Ar}$) and 14 % ($E_{N,AA}$). This may be due to the magnitude of the uncertainties, which were lower for the HQ-MFC, making it more difficult to satisfy the E_N-criterion.

One measurement of the LQ-MFC also demonstrated that method A can be sensitive to adverse measurement conditions. In this case, the measurement volume was too small to allow proper measurement of some features. The results as well as the associated uncertainties came out as clear outliers. Other measurements tested a full and sparse sampling strategy on the same CMMs. This resulted in no significant changes, suggesting that a simplified sampling strategy does not necessarily impact method A uncertainties, even though the estimates are based on measurement data.

7.2 Method B: a priori

7.2.1 Method B2 sensitivity analysis

The connecting rod and multi-feature check results were also evaluated according to method B2. Method B uncertainty estimates are not derived in any way from the actual measurement data, which was obtained by the measurements for method A, but they can be applied to the results. In absolute terms, the B2 uncertainties are on par with method A uncertainties and generally higher than the reference VCMM estimate. As with method A, the problems with the connecting rod make an evaluation using normalised errors more difficult.

The comparison of the reference method and method B2 results from the LQ-MFC in most cases fulfil the E_N-criterion. The HQ-MFC had a higher failure rate attributed to very low reference and method B2 uncertainties, which, when combined, make it much more difficult to satisfy the E_N-criterion. The uncertainty depends on the feature type. One feature – perpendicularity to a short datum axis – consistently yielded uncertainties three or four times higher than any others in a given dataset. The other uncertainties were much more closely distributed – ca. ±0.5 µm range for uncertainties ~3 µm.

7.3 Inter-method comparisons

Results from the EUCoM methods were compared to each other in the same way reference and EUCoM results were compared. E_N scores were calculated using the EUCoM estimates and measurement data obtained from the round robin.

7.3.1 Method A and B2

Method A and B2 normalised errors were calculated for the connecting rod and multi-feature checks. For all three, the results tendentially followed those obtained from the previous comparison to the reference values. The connecting rod results are mostly negative on account of the workpiece issues already discussed. The MFC results are mostly positive as most features pass conformity testing. As noted in section 7.2.1 of this report, the HQ-MFC has very low associated uncertainties, both reference and method A/B2. Hence the failure rate is accordingly higher, as the normalised error increases when uncertainties decrease. Overall, methods A and B2 seem in most cases to be consistent with each other as well as with the methods used to obtain the reference uncertainties.

8 Summary for freeform

A total of twenty-six freeform measurements were collected and evaluated. **Table 8-3** and **Table 8-4** provide an overview of the results. Rather than evaluating different feature types, each freeform measurement is effectively a set of individually evaluated points from a profile or surface. Thus, each set consists of a large population of evaluated features, which offers some insight into the variation or stability of the uncertainty evaluation. Further, the features do not rely on fitted or derived geometries, which would act to stabilise features by combining more data into a single feature and thus, for example, filter or average out outliers and mitigate the impact of random noise.

8.1 Method A: a posteriori

The freeform method A results are very promising. Uncertainties are usually close to the reference uncertainty estimates and thus plausible. Absolute normalised errors are often clearly smaller than 0.5 when comparing results to reference or other method A-evaluated results, indicating an overestimation of the uncertainty. Exceptions are few enough to attribute to normal variation in a large dataset rather than to the method or the artefacts.

From the large number of evaluated points, it is also possible to elucidate the general behaviour of method A. The key factors are u_{geo} (CMM geometry) and u_{rep} (repeatability). In cases where the combined uncertainty varies significantly within a given dataset, u_{geo} is usually the dominant (and varying) contributor. Otherwise, elevated combined uncertainties are usually due to a proportional increase of all uncertainty contributions.

As the freeforms only consider point deviations, these results alone do not provide any information on the behaviour of other features when evaluated by Method A, or any other of the methods tested here.

8.2 Method B1: a priori

The hyperbolic paraboloid was evaluated with a preliminary implementation of method B1. This does not yet consider correlations between different uncertainty contributions but should suffice for a first impression of the method's performance. B1 uncertainties were of similar magnitude as other uncertainties and tended to be higher than the reference values.

Normalised errors in the various cross-comparisons – B1 vs reference, B1 vs B1, B1 vs A – performed very well with very few failures, probably caused by the fact that B1 uncertainties are generally higher than reference or other uncertainties. In most sets, the average score was reasonable. Only one set, acquired on the same CMM as the reference data, scored E_N-values low enough to indicate a possible overestimate of uncertainties.

Since only point deviations are determined in the hyperbolic paraboloid measurements, it is not possible to comment on how the uncertainty varies for different feature types, or on the transferability to other, untested feature types.

The connecting rod had also been investigated. However, between time constraints, the issues with the part itself and the low number of degrees of freedom of the data with respect to method B1 (four orientations, practically no repetition of the different feature types), there was no detailed evaluation.

Method B2 could not be implemented for either freeform artefact but was applied to prismatic workpieces.

Table 8-1: Summary of the method A results for the connecting rod and multi-feature checks.

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Connecting rod	Reference	0.00061	0.00006	-	-	-	-	-	-	-
	1	0.00065	0.00013	2	4	66.7	4	2	33.3	6
	2	0.00432	0.00068	3	3	50.0	3	3	50.0	6
	7	0.00260	0.00033	2	4	66.7	6	0	0.0	6
	8a (Ref CMM)	0.00093	0.00008	6	0	0.0	6	0	0.0	6
	8b	0.00088	0.00005	5	1	16.7	5	1	16.7	6
	10	0.00090	0.00002	2	4	66.7	5	1	16.7	6
	11	0.00285	0.00005	4	2	33.3	5	1	16.7	6
	12	0.00140	0.00004	4	2	33.3	4	2	33.3	6
	14	0.00143	0.00003	2	4	66.7	6	0	0.0	6
	All data	0.00177	0.00119	30	24	44.4	44	10	18.5	54

(Table 8-1 continued)

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
LQ-multi-feature check	Reference	0.00194	0.00118	-	-	-	-	-	-	-
	4a	0.00177	0.00049	11	3	21.4	11	3	21.4	14
	4b	0.00149	0.00023	12	2	14.3	11	3	21.4	14
	4c	0.00305	0.00035	16	1	5.9	17	0	0.0	17
	4d	0.00322	0.00061	17	0	0.0	16	1	5.9	17
	4e	0.00177	0.00066	14	3	17.6	15	2	11.8	17
	4f	0.00147	0.00025	17	0	0.0	11	6	35.3	17
	6	0.00296	0.00099	14	3	17.6	10	7	41.2	17
	8	0.00186	0.00057	17	0	0.0	12	5	29.4	17
	11	0.00323	0.00041	17	0	0.0	17	0	0.0	17
	All data	0.00233	0.00092	135	12	8.2	120	27	18.4	147
HQ-MFC	Reference	0.00080	0.00047	-	-	-	-	-	-	-
	2	0.00757	0.00056	16	1	5.9	16	1	5.9	17
	3	0.00186	0.00049	12	5	29.4	12	5	29.4	17
	10	0.00166	0.00208	10	7	41.2	16	1	5.9	17
	All data	0.00369	0.00302	38	13	25.5	44	7	13.7	51

Table 8-2: Summary of the method B2 results for the connecting rod and multi-feature checks.

	ID	U_{B2} / mm		$E_{N,B2r}$			$E_{N,B2B2}$			$E_{N,B2A}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Connecting rod	Reference	0.00061	0.00006	-	-	-	-	-	-	-	-	-	-
	1	0.00088	0.00005	2	4	66.7	1	5	83.3	4	2	33.3	6
	2	0.00052	0.00003	1	5	83.3	3	3	50.0	3	3	50.0	6
	7	0.00272	0.00025	2	4	66.7	5	1	16.7	5	1	16.7	6
	8a	0.00030	0.00002	6	0	0.0	6	0	0.0	6	0	0.0	6
	8b	0.00278	0.00027	6	0	0.0	6	0	0.0	6	0	0.0	6
	10	0.00064	0.00004	2	4	66.7	4	2	33.3	5	1	16.7	6
	11	0.00409	0.00038	6	0	0.0	6	0	0.0	6	0	0.0	6
	12	0.00034	0.00002	2	4	66.7	4	2	33.3	3	3	50.0	6
	14	0.00092	0.00006	2	4	66.7	6	0	0.0	6	0	0.0	6
	All data	0.00147	0.00130	29	25	46.3	41	13	24.1	44	10	18.5	54

(Table 8-2 continued)

	ID	U_A / mm		$E_{N,B1r}$			$E_{N,B2B2}$			$E_{N,B2A}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
LQ-multi-feature check	Reference	0.00199	0.00118	-	-	-	-	-	-	-	-	-	-
	4a	0.00290	0.00193	9	4	30.8	8	5	38.5	8	5	38.5	13
	4b	0.00290	0.00193	10	3	23.1	13	0	0.0	13	0	0.0	13
	4c	0.00343	0.00230	12	1	7.7	11	2	15.4	12	1	7.7	13
	4d	0.00343	0.00230	13	0	0.0	13	0	0.0	13	0	0.0	13
	4e	0.00274	0.00181	11	2	15.4	13	0	0.0	12	1	7.7	13
	4f	0.00274	0.00181	13	0	0.0	12	1	7.7	12	1	7.7	13
	6	0.00239	0.00157	8	5	38.5	11	2	15.4	8	5	38.5	13
	8	0.00056	0.00036	12	1	7.7	12	1	7.7	12	1	7.7	13
	11	0.00481	0.00328	13	0	0.0	12	1	7.7	13	0	0.0	13
	All data	0.00288	0.00231	101	16	13.7	105	12	10.3	103	14	12.0	117
HQ-MFC	HQ Ref	0.00079	0.00046	-	-	-	-	-	-	-	-	-	-
	2	0.00166	0.00108	7	6	46.2	7	6	46.2	7	6	46.2	13
	3	0.00111	0.00072	3	10	76.9	5	8	61.5	5	8	61.5	13
	10	0.00112	0.00072	6	7	53.8	7	6	46.2	12	1	7.7	13
	All data	0.00130	0.00090	16	23	59.0	19	20	51.3	24	15	38.5	39

Table 8-3: Summary of the method A results for the involute gears and hyperbolic paraboloid.

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Involute gear (smooth, with datum)	Reference	0.00080	0.00000	-	-	-	-	-	-	-
	5a	0.00352	0.01031	555	440	44.2	597	398	40.0	995
	5b	0.00140	0.00199	995	0	0.0	579	416	41.8	995
	7	0.00254	0.00062	915	80	8.0	915	80	8.0	995
	8	0.00099	0.00017	995	0	0.0	993	2	0.2	995
	10	0.00100	0.00200	995	0	0.0	995	0	0.0	995
	11a	0.00176	0.00012	995	0	0.0	995	0	0.0	995
	11b	0.00174	0.00009	995	0	0.0	995	0	0.0	995
	13	0.00143	0.00199	995	0	0.0	995	0	0.0	995
	14	0.00160	0.00035	977	18	1.8	983	12	1.2	995
	All data	0.00177	0.00371	8417	538	6.0	8047	908	10.1	8955

(Table 8-3 continued)

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Involute gear (smooth, without datum)	Reference	0.00080	0.00000	-	-	-	-	-	-	-
	5a	0.00158	0.00008	980	15	1.5	597	398	40.0	995
	5b	0.00130	0.00002	995	0	0.0	579	416	41.8	995
	7	0.00229	0.00008	804	191	19.2	915	80	8.0	995
	8	0.00093	0.00012	995	0	0.0	993	2	0.2	995
	10	0.00090	0.00019	995	0	0.0	995	0	0.0	995
	11a	0.00165	0.00004	995	0	0.0	995	0	0.0	995
	11b	0.00166	0.00005	995	0	0.0	995	0	0.0	995
	13	0.00038	0.00025	995	0	0.0	989	6	0.6	995
	14	0.00121	0.00030	995	0	0.0	799	196	19.7	995
	All data	0.00132	0.00054	8749	206	2.3	7857	1098	12.3	8955

(Table 8-3 continued)

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Involute gear (sinusoidal, with datum)	Reference	0.00080	0.00000	-	-	-	-	-	-	-
	5a	0.00660	0.02817	595	400	40.2	617	378	38.0	995
	5b	0.00133	0.00008	995	0	0.0	619	376	37.8	995
	8	0.00098	0.00014	989	6	0.6	989	6	0.6	995
	10	0.00106	0.00186	993	2	0.2	993	2	0.2	995
	11a	0.00176	0.00011	993	2	0.2	972	23	2.3	995
	11b	0.00191	0.00025	995	0	0.0	995	0	0.0	995
	13	0.00053	0.00188	978	17	1.7	978	17	1.7	995
	14	0.00156	0.00035	816	179	18.0	520	475	47.7	995
	All data	0.00197	0.01017	7354	606	7.6	6683	1277	16.0	7960

(Table 8-3 continued)

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Involute gear (sinusoidal, without datum)	Reference	0.00080	0.00000	-	-	-	-	-	-	-
	5a	0.00158	0.00006	964	31	3.1	965	30	3.0	995
	5b	0.00132	0.00007	995	0	0.0	965	30	3.0	995
	8	0.00092	0.00002	896	99	9.9	935	60	6.0	995
	10	0.00097	0.00103	991	4	0.4	991	4	0.4	995
	11a	0.00166	0.00005	995	0	0.0	995	0	0.0	995
	11b	0.00165	0.00010	995	0	0.0	995	0	0.0	995
	13	0.00041	0.00105	986	9	0.9	983	12	1.2	995
	14	0.00118	0.00025	992	3	0.3	956	39	3.9	995
	All data	0.00121	0.00067	7814	146	1.8	7785	175	2.2	7960

(Table 8-3 continued)

	ID	U_A / mm		$E_{N,Ar}$			$E_{N,AA}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Hyperbolic paraboloid	Reference	0.00050	-	-	-	-	-	-	-	-
	4	0.00158	0.00001	52	0	0.0	52	0	0.0	52
	6	0.00233	0.00035	52	0	0.0	52	0	0.0	52
	8a (Ref CMM)	0.00091	0.00001	52	0	0.0	52	0	0.0	52
	8b	0.00091	0.00009	52	0	0.0	52	0	0.0	52
	9	0.00101	0.00010	52	0	0.0	52	0	0.0	52
	10	0.00074	0.00011	52	0	0.0	52	0	0.0	52
	11	0.00103	0.00002	50	2	3.8	52	0	0.0	52
	12	0.00221	0.00013	52	0	0.0	52	0	0.0	52
	All data	0.00134	0.00060	414	2	0.5	416	0	0.0	416

Table 8-4: Summary of the method B1 results from the hyperbolic paraboloid

	ID	U_{B1} / mm		$E_{N,B1r}$			$E_{N,B1B1}$			$E_{N,B1A}$			Number of features
		Mean	Std. dev.	Pass	Fail	Fail / %	Pass	Fail	Fail / %	Pass	Fail	Fail / %	
Hyperbolic paraboloid	Reference	0.00050	0.00000	-	-	-	-	-	-	-	-	-	-
	6	0.00132	0.00000	51	1	1.9	52	0	0.0	52	0	0.0	52
	8a (Ref CMM)	0.00031	0.00000	52	0	0.0	52	0	0.0	52	0	0.0	52
	8b	0.00120	0.00000	52	0	0.0	52	0	0.0	52	0	0.0	52
	9	0.00081	0.00000	52	0	0.0	52	0	0.0	52	0	0.0	52
	10	0.00071	0.00000	52	0	0.0	52	0	0.0	52	0	0.0	52
	11	0.00041	0.00000	47	5	9.6	50	2	3.8	50	2	3.8	52
	12	0.00271	0.00000	52	0	0.0	52	0	0.0	52	0	0.0	52
	All data	0.00107	0.00076	358	6	1.6	362	2	0.5	362	2	0.5	364

9 Conclusion

Following the development or elaboration of the new uncertainty estimation approaches, Method A [1] and Method B (B1 and B2) [2], a measurement campaign was conducted to provide a practical test and validation of each of the three methods. A set of prismatic and freeform artefacts were selected, and measurement strategies devised (see chapter 3.3). Consortium partners measured the artefacts using their respective tactile measurement setups to provide a diverse library of measurements obtained from different quality hardware and under different conditions. The uncertainties for each dataset were evaluated according to the new methods and the results compared (chapter 6).

Comparisons used reference measurements and uncertainties (obtained by a method other than A or B) or used intra-method or inter-method comparisons to test methods for internal consistency, and against each other. Comparisons were done by calculating the normalised errors (E_N) according to EN ISO/IEC 17043. [6] The normalised error compares two independent measurement values while also taking the associated uncertainties into account (see chapter 5).

Method A – an empirical approach to measurement uncertainty utilising a large cache of measurement data – in most cases performed well against reference data and in internal consistency checks. Uncertainty estimates were generally consistent within each dataset, independent of the feature type. Across datasets, results were also mostly consistent. Uncertainties were tendentially higher than reference values determined by other methods, but not implausibly high. The freeform measurements reveal the measurement-dependent variability of U_A, for which u_{geo} – CMM geometry – is the primary cause. The nature of the workpiece – freeform or prismatic – did not seem to have any significant impact in this investigation.

Anomalies are discussed in the corresponding chapter. Most importantly, there are some measurements indicating that U_A may be too high, resulting in very low E_N scores. Conversely, the trend for $E_{N,AA}$-scores (method A vs method A comparisons) to have higher failure rates than $E_{N,Ar}$ (A vs reference), could indicate that U_A is an underestimate of the measurement uncertainty. A follow-up is strongly recommended.

Overall, method A is a very direct approach towards measurement uncertainty. Instead of modelling, large amounts of data are collected to effectively “measure” the uncertainty. Unfortunately, measurements in four positions are necessary for method A, which comprises a significant effort.

Method B1, employing a model based on expert knowledge to describe the CMM a priori, was tested on the hyperbolic paraboloid freeform artefact. The uncertainty was generally higher than reference or method A uncertainties. Comparative E_N-scores were accordingly low, although not so low as to suggest that determined U_{B1}-values are universally too high. B1 results were tested against reference, method A and other B1 results. However, due to time constraints and the difficulties in developing the method, a more detailed analysis could not be completed. No conclusions can be drawn on the performance with different feature types. Further work should be done to apply this method to the remaining artefacts of the EUCoM project.

Method B2 – sensitivity analysis based on MPE values and length measurement errors E_L according to EN ISO 10360-2 – was tested on the connecting rod and on the multi-feature checks. An implementation for a freeform workpiece could not be completed in time.

As before, uncertainties were generally higher than reference values. Performance was generally similar to method A. The feature type being assessed had little impact on the uncertainty, with the exception of perpendicularity, although this was attributed to the physical

artefact (short datum feature leading to high uncertainty), not the method itself. In fact, this highlighted the responsivity of the method B2 to ill-conditioning of the measurement. As with method B1, an analysis of the other workpieces should be conducted in the interest of fully validating the method. Some validation results are found in [2]. At this time, no comment can be made on the impact of freeform-type surfaces, or on how determined uncertainties might vary for a given feature type.

Both B-type methods appear to be working, in the sense that results are mainly plausible and consistent across datasets. A full validation, however, could not be completed. Being “a priori”, the necessary information for either approach should generally be available from documentation or previous measurements on a system.

Method B2, in particular, needs only MPE statements and (optionally) E_L data, which should be available immediately, even for a newly installed CMM, being part of the specification and acceptance testing. In principle, this method can therefore be applied “out of the box” with minimal preparation, even in the absence of any experience with the system as required by method B1.

The downside is the complexity of the models. For method B1, this requires rather many parameters. Method B2, is comparatively simple, which may carry the risk of underestimating uncertainties. In the study, it was also more difficult to apply the B-type methods to the different artefacts. By contrast, method A, once the basic template was complete, was relatively easy to apply to other workpieces.

10 References

[1] A. Balsamo and O. Sato, “Deliverable D1: Procedure for a posteriori (type A) evaluation of measurement uncertainty based on multiple measurement strategies,” EMPIR 17NRM03 EUCoM, 2021.

[2] A. Forbes, W. Plowucha and A. Balsamo, “Deliverable D2: Guidelines on uncertainty of coordinate measurements a priori using type B evaluation,” EMPIR 17NRM03 EUCoM, 2021.

[3] EN ISO 10360-2:2009, “Geometrical product specifications (GPS) — Acceptance and reverification tests for coordinate measuring machines (CMM) — Part 2: CMMs used for measuring linear dimensions,” 2009.

[4] I. Linkeová, P. Skalíik and V. Zelený, “Calibrated CAD Model of Freeform Standard,” in *XXI IMEKO World Congress "Measurement in Research and Industry"*, Prague, Czech Republic, 2015.

[5] I. Linkeová, P. Skalník and V. Zelený, “Calibration of freeform standard,” in *euspen's 15th International Conference & Exhibition*, Leuven, Belgium, 2015.

[6] EN ISO/IEC 17043:2010, “Conformity assessment — General requirements for proficiency testing,” 2010.