PAPER • OPEN ACCESS

Assessment of 2D-based tests for the qualification of simulation software for dXCT

To cite this article: Fabrício Borges de Oliveira *et al* 2023 *Meas. Sci. Technol.* **34** 064005

View the article online for updates and enhancements.

You may also like

Assessment of 2D-based tests for the qualification of simulation software for dXCT

Fabrício Borges de Oliveira[1,*] ![ORCID]**, Tamara Reuter**[2] ![ORCID]**, David Plotzki**[3] ![ORCID]**,
Florian Wohlgemuth**[4]**, Markus Batscher**[1] ![ORCID]**, Carsten Bellon**[3] ![ORCID] **and Stefan Kasperl**[5,6] ![ORCID]

[1] Coordinate Metrology department, Physikalisch-Technische Bundesanstalt, Braunschweig, Germany
[2] Chair of Manufacturing Metrology (FMT), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
[3] Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany
[4] HEITEC PTS GmbH, Kuchen, Germany
[5] Fraunhofer Development Center for X-ray Technology (EZRT), Fürth, Germany
[6] University of Applied Sciences Georg Simon Ohm, Nuremberg, Germany

E-mail: fabricio.borges@ptb.de

Abstract

The interest in using computer simulations of dimensional x-ray computed tomography (dXCT) for various metrological purposes—such as measurement planning, performance prediction, performance optimisation and, finally, measurement uncertainty estimation—is increasing along with the ever-growing demand for more reliable measurements with dXCT. However, before a piece of simulation software can be used for tasks related to coordinate metrology, it has to be ensured that it is able to simulate physical laws, characteristic effects and basic CT system functionalities correctly and with sufficient accuracy. In short, the software must be qualified for dimensional metrology tasks. As one part of such a qualification process, a method is presented here for determining conformity intervals of 2D tests (projection-based tests) based on 3D tests (testing based on dimensional evaluations in a reconstructed volume) for the assessment of dXCT simulation software. The method consists of varying relevant parameter values in order to verify their influence on 3D measurement results. The results of the 3D tests with varied parameter values are then transferred to the quantities tested in the 2D tests and used as the basis for determining conformity intervals. Two approaches are applied for determining whether or not a variation of a parameter value is significant: (a) statistical and (b) heuristic. Two examples are presented, each based on simulated images, which show the application of the two different approaches for determining conformity intervals for the results of the 2D tests.

Keywords: x-ray computed tomography, dXCT, simulation

(Some figures may appear in colour only in the online journal)

---

* Author to whom any correspondence should be addressed.

1. Introduction

X-ray computed tomography (CT) systems have been used as coordinate measurement systems (CMSs) for almost two decades now [1]. Thanks to its holistic approach (acquisition of the inner and outer structures of an object with a single scan), CT enables measurement tasks in coordinate metrology that were previously not possible. Nonetheless, CT-based CMSs have yet to fully attain the same level of confidence and reliability enjoyed by conventional CMSs, i.e. tactile and optical systems. The inferior confidence still placed in CT for many applications can be attributed to the high complexity of the measurement process, which hampers the estimation of measurement uncertainty and, consequently, traceability to the SI unit of length, the metre.

Conventional methods applied in coordinate metrology for estimating measurement uncertainty often require either a comprehensive understanding of the measurement process (as described in the guide to the expression of uncertainty in measurement – GUM [2]) or a great deal of effort due to the need for repeated measurements of a calibrated standard (as described for tactile CMSs in ISO 15530-3 [3] and VDI/VDE[7]2617-8 [4], and specifically for CT in VDI/VDE 2630-2.1 [5]). Therefore, industry, research and calibration/measurement laboratories share a strong interest in acquiring time-, cost- and resource-efficient methods for estimating CT measurement uncertainty. Methods based on simulation seem to be a promising solution as they potentially yield more information with less effort and lower costs.

Several software packages are available for the simulation of CT measurements using radiographic modelling (e.g. [6–10]). These have already been used in research and industry for various applications as described, for example, in [11–14]. However, one important obstacle to using simulations for tasks related to coordinate metrology is the lack of standardised procedures (guidelines, standards, etc) for qualifying simulation software for the CT measurement process.

The general concept of simulation-based uncertainty estimation using the Monte Carlo method is described in GUM Supplement 1 [15]. Furthermore, ISO/TS 15530-4 specifies requirements for the application of uncertainty evaluating software to CMS measurements. It also describes testing methods for verifying uncertainty evaluating software and various test procedures for the evaluation of task-specific measurement uncertainty. However, both GUM Supplement 1 and ISO/TS 15530-4 do not yet consider the specifics of CT.

To overcome this limitation, the CTSimU project series was initiated with the aim of developing a standardised procedure to estimate measurement uncertainty using simulation. Due to the complexity of the matter, the approach was divided into three sequential research topics: (a) the CTSimU project dealt with the correct simulation of basic physical laws and functionalities (basic qualification) of CT simulation software [1], so that in step (b) methods to build and verify digital
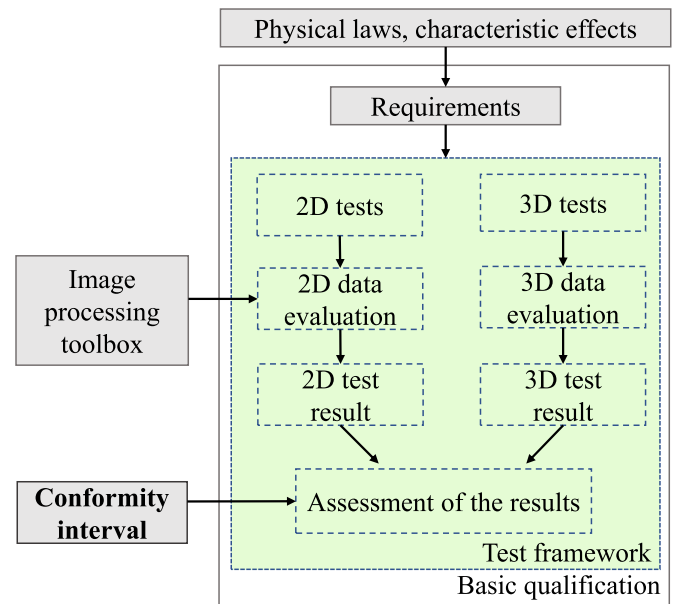

Figure 1. Basic qualification of CT simulation software.

twins (i.e. digital replicas of real CT systems) using qualified CT simulation software can be developed (CTSimU2), thus enabling (c) the development of standardised methods and procedures, using digital twins, to estimate the uncertainty of CT measurements. The outcomes of the CTSimU project were presented to and are being discussed by VDI/VDE technical committee 4.33 (until 2022: 3.33) for the development of a guideline for the basic qualification of CT simulation software for dimensional metrology using CT-based CMSs.

Before a CT simulation software tool can be used for tasks related to the estimation of measurement uncertainty, it has to be ensured that the relevant physical laws, characteristic effects and basic functionalities of the CT measurement process are correctly reproduced. In other words, the software needs to be validated for specific tasks related to dimensional measurements. This basic qualification process aims at qualifying simulation software based on a test framework (i.e. a set of tests) in which several physical laws, characteristic effects and software functionalities are tested separately. These physical laws and characteristic effects are translated into requirements that a simulation software package must fulfil to be considered 'basically qualified'. To test simulation software efficiently, the test framework is divided into 2D and 3D tests, see figure 1. The 2D tests are carried out based on a single or a few projections of specific test scenarios to evaluate one or more physical effects or functionalities. Data evaluation of the 2D tests is based on 2D image processing techniques developed within the CTSimU project [16–18]. The 3D test verifies the ability of the simulation software to perform complete CT scans. Data evaluation and testing are based on dimensional measurements performed on the reconstructed volume.

How does one decide if a simulation software program reproduces a physical law and/or a characteristic effect and/or a functionality with sufficient accuracy? To answer

---

[7] VDI = Association of German Engineers and VDE = Association for Electrical, Electronic and Information Technologies

this question, a method to determine conformity intervals for the 2D tests based on simulations of complete CT scans was developed and is presented in this contribution. The method consists of varying specific scan parameters (to be tested in the 2D test scenarios) in the simulation of a meaningful, complete CT scan to verify their influence on dimensional measurements. The conformity interval is, in this context, a range of values of a 2D metric in which the results produced by simulation software tools, while being tested, must lie in order to be considered 'basically qualified'.

Next, the approaches for CT simulation that were applied in the project are briefly presented, followed by a detailed explanation of the proposed method for determining conformity intervals. Finally, two examples of how to determine conformity intervals for 2D tests are described, one for the simulation of projection noise and another for the simulation of a geometric offset of the rotary table.

2. CT simulation

The two main approaches to the simulation of radiographic images are: (a) deterministic ray-casting simulations and (b) stochastic Monte Carlo-based particle physics simulations. Approach (a) essentially reproduces the Beer–Lambert law of radiation attenuation along rays in the virtual scene, and (b) follows the trajectories of individual photons and their interactions with matter by taking into account the phenomena of radiation extinction (i.e. absorption and scattering processes).

In both approaches, a virtual CT with its main components (i.e. x-ray source, rotary table/measurement object and x-ray detector) and their geometrical relations and radiation interactions must be implemented with adequate precision. For example, the CT geometry — meaning the relative positions and orientations of the x-ray source, rotary table/object and detector — and information on the object geometry have to be defined in the software environment. Simulation software programs additionally require information on the chemical composition and density of the materials that make up the components (e.g. the object, simulation environment) as well as the parameters for the generation of the x-rays (e.g. acceleration voltage, anode material) and the detection of the radiation (e.g. scintillator material, size of the detection elements). The object geometry can usually be imported into the simulation environment as surface meshes (e.g. STL files) or CAD geometry files.

3. Approach for determining conformity intervals for 2D tests

The main task of the test framework is to verify the ability of a software program to simulate physical effects and the CT geometry with sufficient accuracy. Insufficient simulation accuracy of the effects and geometry will have an impact on the reconstructed volume and potentially on the measurement result, causing measurement deviations. As a consequence, any software which is not able to reproduce effects and system geometry with sufficient accuracy can most likely not be
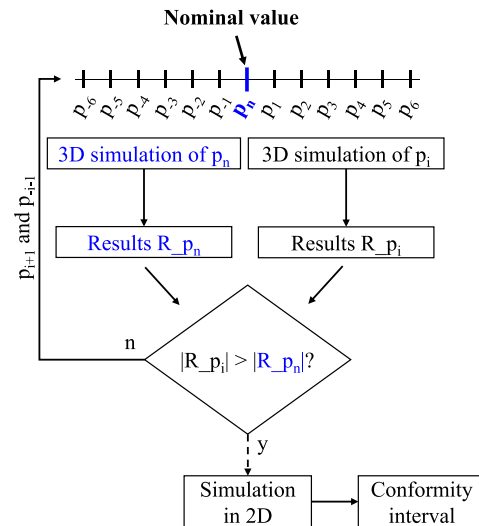

Figure 2. Concept for determining conformity intervals of 2D tests based on 3D simulation. Here, p_n is the nominal and p_i the variated parameter value. R_p_n and R_p_i are the measurement results of the nominal and variated parameter values, respectively.

used for metrological tasks. A simulation software tool should therefore be able to simulate physical quantities with sufficient accuracy such that no significant deviation from the expected measurement result is observed (e.g. manifested as a measurement error).

But how does one decide if a simulation software tool reproduces physical quantities with sufficient accuracy? To answer this question, this contribution proposes a method for determining conformity intervals for the assessment of 2D tests based on the dimensional results of simulations of complete CT scans (i.e. 3D simulations). The method consists of a stepwise variation of parameter values from their nominal value. The goal is to determine whether a significant 3D measurement deviation is caused by the varied parameter value. Parameter value variation in this context means that the value of a single parameter, such as noise, is varied in the positive and negative directions from a predefined nominal value (see figure 2).

3D simulations are carried out with the multi-geometry cuboid (section 3.2) and a specially designed ideal simulation scenario (section 3.3). Complete scans of the nominal (p_n in figure 2) and varied values of a parameter (p_i in figure 2) are simulated and dimensionally evaluated. The measurement results obtained with varied parameter values (R_p_i in figure 2) are then compared with the results achieved with nominal conditions (R_p_n in figure 2). If the variation value of a parameter significantly influences the measurement result (i.e. the result differs substantially from the result obtained with the nominal parameter value), the variation behaviour is used for the determination of the conformity interval. The tolerable amount of parameter variation (measurement result remains within limits) is used for defining the conformity interval. The effects are tested separately to achieve maximum sensitivity, thus ensuring that even a small simulation error will be caught by the test framework.

To determine whether the variation has a significant influence on the measurement result compared to the result obtained with the nominal parameter value, two approaches are applied: (a) a statistical method and (b) a heuristic threshold based on the literature and expert knowledge, see section 3.1.

Figure 2 schematically illustrates the concept for determining the conformity interval. Here, p is considered to be a parameter of a physical quantity to be tested based on a radiograph. The conformity interval for p is established by simulations of complete CT scans followed by dimensional measurements, where the parameter value p_i is varied stepwise (with sign) from its nominal value (p_n). If the measurement deviation generated by a varied parameter value p_i is significantly larger than the result obtained with the nominal value (i.e. $|R\_p_i| > |R\_p_n|$), then this variation value cannot be used for the determination of the conformity interval and the last tolerable parameter value is used for appointing the conformity interval.

Some quantities of the test framework, such as noise, require no further steps for the definition of the conformity intervals because the 2D test is carried out with the same parameters as the 3D test. As soon as the upper and lower limits in 3D are found, they can be directly used as the boundaries of the conformity interval. However, there are some test quantities where the parameters measured in 2D are different than the parameters measured in 3D, for instance geometrical parameters, which require an additional step for transferring the limits from the measurands in 3D to the measurands in 2D. In such cases, the last variated value (results of 3D) for which the result is similar (within limits) to the results obtained with nominal conditions is simulated in 2D to find the conformity interval for the 2D measurand of that test.

3.1. Definition of thresholds

How does one determine if the measurement deviation caused by the variation of a parameter is significant? Two methods are suggested and were applied for determining whether the 3D measurement deviation caused by the variation of a parameter value is significant. The approaches are based on (a) a statistical hypothesis test and (b) a heuristic threshold.

In the statistical method, both measurement results (varied R_p_i vs. nominal R_p_n) are compared statistically with one another based on a hypothesis test. The result of the statistical test indicates with a certain level of confidence whether both results (R_p_i vs R_p_n, cf figure 2) are considered different. If a value of a parameter variation p_i is found to cause a statistically significant difference from the result obtained using the nominal parameter value p_n, this variation value p_i will most likely (with 95% probability for a statistical significance of $\alpha/2 = 0.05$) cause a significant measurement deviation. This approach requires a series of repeated simulations. To improve the sensitivity of the test for different parameter variations, the test scenarios are simulated using (almost) ideal scenarios (see section 3.3), i.e. all influencing parameters, apart from those under test, were switched off (including noise). Consequently, repeated simulations can lead to identical results

if the parameter under investigation is not itself introducing an element of randomness to the simulation (as is the case for image noise). Such fully deterministic cases (like specific geometric misalignments), disallow the application of a statistical test since repeated measurements of a varied parameter would lead to identical results. Instead, a heuristic threshold must be applied here.

The definition of the heuristic threshold is based on the literature and expert experience ([19, 20]). The heuristic threshold method compares the dimensional measurement results of the simulations (difference of varied vs. nominal) with a predefined threshold that depends on a predefined fraction of the voxel size of the reconstructed volume (e.g. 10% × voxel size).

Not all measurands react sensitively to all variated parameter values, but as soon as a single measurand (of the measurands described in section 3.2) exceeds the threshold or shows statistical difference, the parameter variation p_i will be considered to have a significant influence on the measurement result. Consequently, the boundary of the conformity interval is determined from a 2D test of a simulation using the last parameter variation p_{i-1} and p_{-i+1} that was still considered 'insignificant'.

3.2. Reference standard used in the 3D simulations

The 3D simulations used for the determination of the conformity intervals of the 2D tests were carried out with the *multi-geometry cuboid* [21]. The multi-geometry cuboid has a scalable geometry (it was used here with dimensions of 22.2 mm × 21.8 mm × 15.0 mm) and features 37 inner and outer geometrical elements (i.e. cylinders, cones, planes, half-spheres/calottes, tori), see figure 3. Based on the geometrical elements, 54 measurands of different complexity (including distances, diameter, cylindricity, concentricity, etc) were measured, see table 1. To identify the geometrical elements in the reference standard, the labels of all 37 geometrical elements are presented in figure 3. The set of measurands intentionally included both simple and relatively complex measurement tasks, thus guaranteeing that a minimum number of measurands would sensitively react to the variation of the respective parameter values and detect potential undesired effects and hidden errors in the simulation software. Further applications of this standard can be seen in [22, 23]. For the current contribution, reconstruction (standard filtered back projection), local gradient-based surface determination and data evaluation were carried out automatically (using scripts) in the software VG Studio Max 3.5 by Volume Graphics GmbH, Heidelberg, Germany.

3.3. 3D basis scenario

A 3D test scenario was designed to simulate the variation of the parameter values with the multi-geometry cuboid. The scenario is designed to sensitively react to variations in different factors influencing the measurement results, but it still features a realistic scan scenario, see figure 4. To guarantee maximum sensitivity, effects were considered separately with simulations carried out under (almost) ideal conditions.
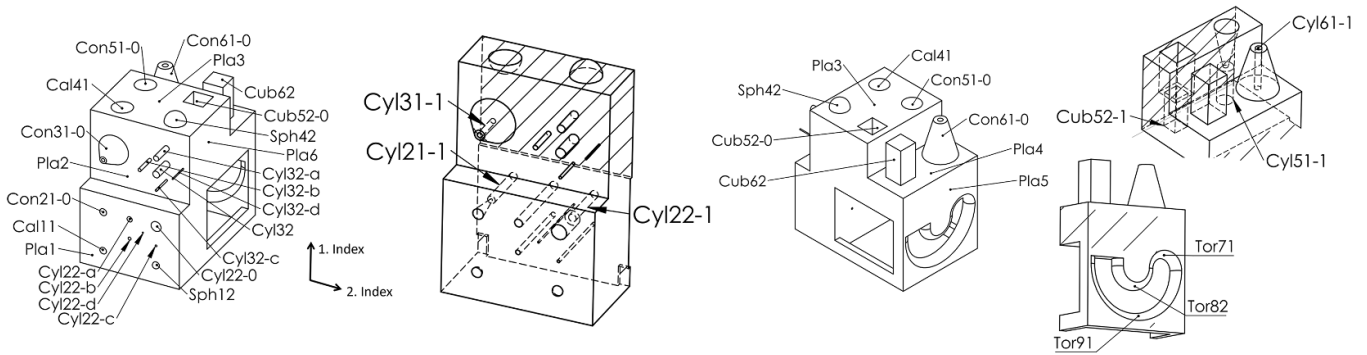
Figure 3. Multi-geometry cuboid with labels of all geometrical elements. Table 1 shows the explanation and allocation of the features to the specific measurement tasks.

Table 1. 54 measurands measured in the multi-geometry cuboid. Measurand numbers (in *italic*) match the presentation of results in section 4.

Measurand type	Designation	Geometrical elements used (bold face entries indicate reference datums)			
Coaxiality	Coa	*1)* Con21-0–**Cyl21-1**	*2)* Cyl31-1–**Con31-0**	*3)* Cyl61-1b–**Con61-0**	
Concentricity	Con	*4)* Cyl61-1 c–**Con61-0**			
Cylindricity	Cyl	*5)* Cyl22-0	*6)* Cyl22-c	*7)* Cyl22-d	*8)* Cyl32	*9)* Cyl32-a
		10) Cyl32-b	*11)* Cyl32-c	*12)* Cyl32-d	*13)* Cyl51-1
Diameter	Dia	*14)* Cal11	*15)* Cal41	*16)* Cyl22-0	*17)* Cyl22-a	*18)* Cyl22-b
		19) Cyl22-c	*20)* Cyl32	*21)* Cyl32-a	*22)* Cyl32-b	*23)* Cyl32-c
		24) Cyl51-1	*25)* Cyl61-1a	*26)* Cyl61-1b	*27)* Cyl61-1 c	*28)* Sph12
		29) Sph42			
Centre-to-centre distance	Dis	*30)* Cal11–Sph41	*31)* Cal11–Cal12	*32)* Cal11–Sph42	
Point-to-point distance	Dis	*33)* Poi51-1 c–Poi52-1 c	*34)* Poi52-0b–Poi62-d	*35)* Poi52-0b–Poi62-d	
Flatness	Fla	*36)* Cub52-1d	*37)* Cub62-b		
Torus major radius	MajRad	*38)* Tor71	*39)* Tor81	*40)* Tor91	
Torus minor radius	MinRad	*41)* Tor71	*42)* Tor81	*43)* Tor91	
Parallelism	Par	*44)* Cyl32–**Con31-0**	*45)* Cyl51-1–**Con61-0**	*46)* Con21-0–**Cyl61-1**	
Perpendicularity	Per	*47)* Cyl31-1–**Con51-0**			
Roundness	Rou	*48)* Cyl61-1a	*49)* Cyl61-1b	*50)* Cyl61-1 c	
Sphericity	SFo	*51)* Cal11	*52)* Sph41	*53)* Cal12	*54)* Sph42

This means, apart from the parameter under test, all other influencing factors were set to exactly match the reconstruction parameters (e.g. for the CT geometry parameters) and any image quality degrading effects were switched off (e.g. detector unsharpness, projection noise).

The simulations were performed with an ideal energy-integrating detector, a polychromatic x-ray spectrum (150 kV) and the multi-geometry cuboid made of aluminium. The environment was assumed to be air.

This scenario was designed to have a magnification factor of 16.6667, a beam opening angle of ~9.1°, and a detector pixel size of 680 μm, which results in a voxel size of [~40.8 μm]³. All 3D simulations were carried out with 1500 projections (in equidistant angular steps of a complete 360° rotation). All simulations were performed using the analytical simulation tool aRTist 2.10 from BAM, Berlin, Germany [6], which follows the approach (a) mentioned in section 2. The software also contains analytical models for the x-ray source and the detector. The also possible Monte-Carlo-based simulation of x-ray scattering processes within aRTist has not been used in this study.

4. Application examples of the method for determining conformity intervals

4.1. Projection noise

How can projection noise be tested in 2D? The simulation of noise can be tested based on a single free-beam projection image (i.e. radiographic image with no object between source and detector). A possible way of doing that is based on comparing the results of a nominal signal-to-noise ratio (SNR) with those of the measured SNRs. The SNR can be calculated as defined in equation (1). Data evaluation is carried out using the CTSimU Toolbox [16, 17].

$$\text{SNR} = \frac{\text{mean signal amplitude}}{\text{mean noise amplitude}} = \frac{\mu}{\sigma} \qquad (1)$$
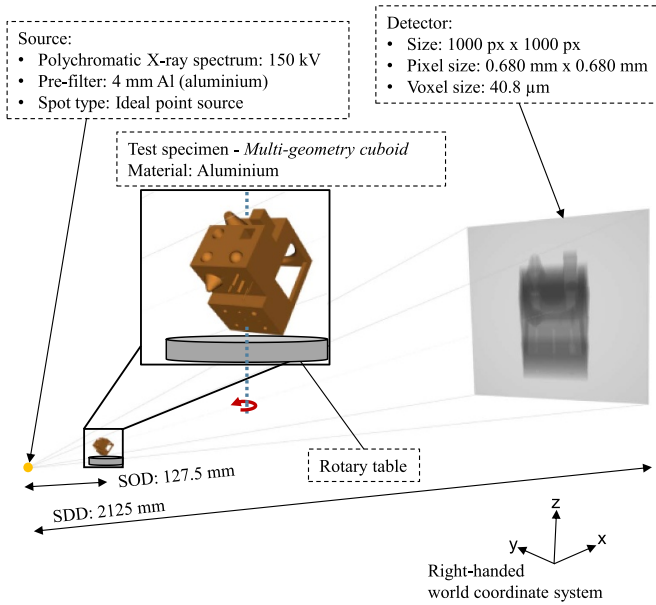
Figure 4. Schematic representation of the 3D test in an (almost) ideal scenario (source: adapted from [22]).

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}. \tag{2}$$

In equation (2), N is the number of pixels in a certain evaluated region of interest (ROI), i is a sequential pixel index within the ROI, and x_i is the grey value of pixel i. The mean grey value within the ROI is denoted as μ.

How does one define conformity intervals for the projection noise test? The definition of conformity intervals for the projection noise test is based on the proposed method presented in section 3.1 and on the proposed statistical test method. The generic test scenario presented in section 3.3 was used for this test.

The SNR values were varied stepwise and compared to the nominal value. For every new variation of the SNR value, five repeated 3D simulations were carried out and the results compared statistically with the results of the nominal SNR value, for which 20 repeated simulations were performed. When a result (of a single measurand) reveals a statistical difference, this means that this variated SNR value already influences the measurement results significantly and that the magnitude of its influence cannot be accepted as tolerable. In such cases, the last tolerable value is used for the conformity interval.

4.1.1. Statistical approach

Considering that different SNRs deliver results with unequal variances, a comparison based on the Welch's statistical hypothesis test (*t*-test for unequal variances) was conducted [24].

The Welch's test is based on the acceptance or rejection of two hypotheses: the null hypothesis (H_0) assumes that the nominal ($\bar{X}_N$) and variated ($\bar{X}_V$) averages of the sample are

equal, while the complementary hypothesis (H_1) assumes that they are different, see equations (3) and (4):

$$H_0 : \bar{X}_N - \bar{X}_V = 0 \tag{3}$$

$$H_1 : \bar{X}_N - \bar{X}_V \neq 0. \tag{4}$$

The sample averages were calculated from 20 repeated simulations of the nominal SNR (a number suited to obtain a more robust statistic given that the nominal SNR is the reference) and from five repeated simulations of the varied SNR values. The different sample sizes are considered in the calculation of the Welch's statistic and in the estimation of the degrees of freedom (equations (6) and (7)). The comparison is carried out for each measurand individually.

Considering a two-sided rejection region of the H_0, H_0 can be rejected and H_1 accepted if the statistic t_w is larger or smaller than the tabled *t*-student value for v degrees of freedom, $\alpha/2$ (0.05) significance level and 95% probability:

$$t_w > t_{v,\alpha/2} \; or \, t_w < -t_{v,\alpha/2}. \tag{5}$$

The Welch's statistic (t_w) is calculated as a function of both sample averages ($\bar{X}_N$ and $\bar{X}_V$), the standard deviation (s_N and s_V) and the number of repetitions (N_N and N_V) according to equation (6):

$$t_W = \frac{\bar{X}_N - \bar{X}_V}{\sqrt{\frac{s_N^2}{N_N} + \frac{s_V^2}{N_V}}}. \tag{6}$$

The number of degrees of freedom, v, is calculated based on equation (7),

$$v = \frac{\left(\frac{s_N^2}{N_N} + \frac{s_V^2}{N_V}\right)^2}{\frac{s_N^4}{N_N^2 v_N} + \frac{s_V^4}{N_V^2 v_V}} \tag{7}$$

where v_N and v_V are the degrees of freedom associated, respectively, with the standard deviations of the nominal sample and the samples with varied SNR values.

One advantage of the Welch's test over the standard t-test is that the Welch's test delivers a more robust statistic with respect to type 1 errors (i.e. false positives) than does the t-test, given that it estimates the statistical degrees of freedom for both samples.

4.1.2. Conformity intervals for the 'projection noise' test

How does one determine the conformity interval for the '2D projection noise' test for a nominal SNR of 100?

The goal is to find both a higher and a lower varied SNR value (around the nominal SNR of 100), p_i and p_{-i}, that significantly influence the measurement results. To do so, five repeated simulations of each varied SNR value (i.e. SNR …, 98, 99, 101, 102, …) were carried out and compared statistically with 20 repeated simulations of the nominal SNR of 100. When the normalised result $t_w/t_{\alpha/2,v}$ (derived from equation 5
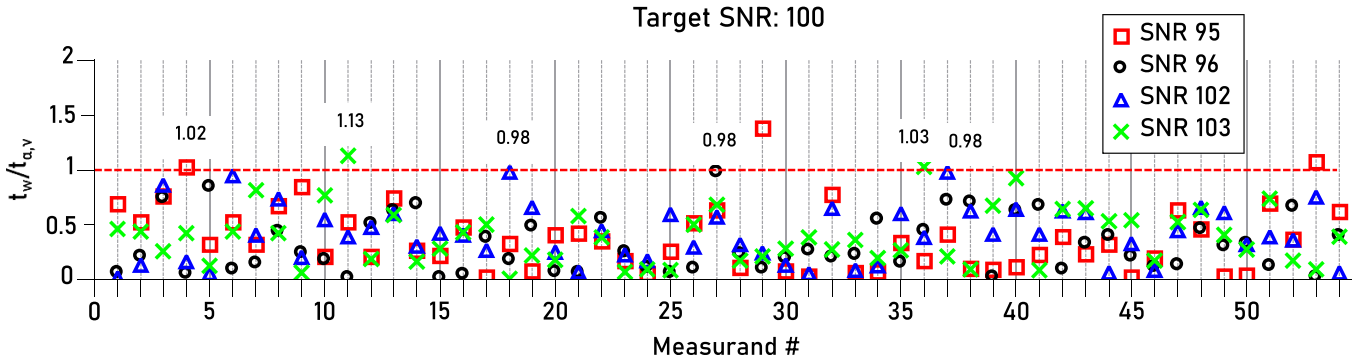
Figure 5. Results of the Welch's test for the nominal SNR of 100 vs. varied SNR values of 95, 96, 102 and 103.

was larger than 1, H_0 ($\bar{X}_N = \bar{X}_V$) could be rejected and H_1 ($\bar{X}_N \neq \bar{X}_V$) accepted, meaning that the samples could be considered statistically different. This statistical comparison was carried out individually for each of the 54 measurands. The proposed method suggests that as soon as a single measurand of a varied SNR value presents a significant reaction (statistical difference), this varied SNR value cannot be accepted as tolerable, and the next acceptable value should be used for determining the conformity interval.

The results of the statistical comparison based on the Welch's test for the 54 measurands are presented in figure 5.

For the sake of simplicity and graph legibility, figure 5 only presents the results of the statistical comparison of SNRs 95, 96 102 and 103. The graph shows that all results ($t_w/t_{\alpha,v}$) obtained with SNR 96 and 102 lie below 1, while the results of three measurands of SNR 95 (4. Con_Cyl61-1 c–**Con61-0**, 29. Dia_Sph42 and 53. SFo_Cal12) and two measurands of SNR 103 (11. Cyl_Cyl32-c and 36. Fla_Cub52-1d) were statistically different when compared to SNR 100. As SNRs 95 and 103 influence the measurement result significantly, they cannot be accepted as tolerable, so the conformity interval of SNR 100 is [96, 102] (for the given step size of one in the presented test series). This means that when simulating an SNR of 100, the simulation software is allowed to wrongly produce projection images with SNRs between 96 and 102, with no significant influence on the 3D measurement result (see figure 6). We note that this simplified test was carried out with Gaussian noise distributions. Other distributions are expected to have a different influence and could therefore result in different conformity intervals.

4.2. Position of the rotary table

How can the position of the rotary table be tested in 2D?

The conical hole sheet standard [21] can be used (see figure 7) to test, based on a single radiographic image, if a simulation tool correctly positions and orients the x-ray source, rotary table and detector relative to one another. The standard features 10 holes which have a distorted conical shape and whose surfaces are oriented towards the focal spot of the x-ray source. The holes are designed in this way so as to improve the precision of data evaluation (segmentation) by creating sharp projected hole edges on the detector.
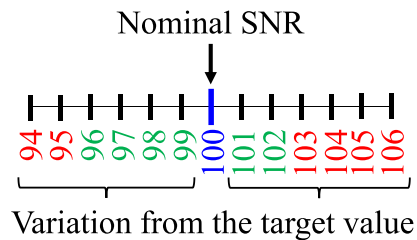
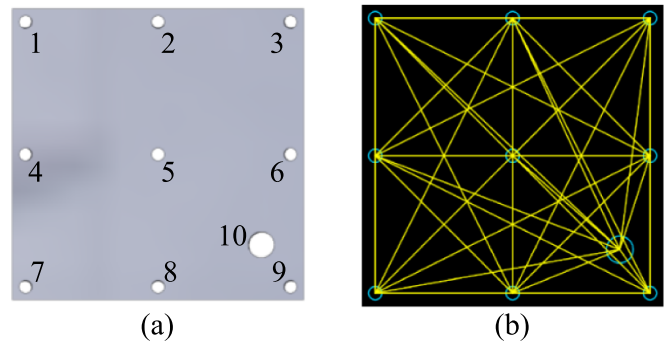**Figure 6.** Conformity interval for SNR 100.

Figure 7. (a) Conical hole sheet and (b) connecting vectors between all holes of the conical hole sheet used for the data evaluation.

To verify the correct positioning and orientation of the rotary table based on a simulated radiograph, the relative distances and vector orientations between all possible hole pairs of the hole sheet are used along with the absolute positions of all holes in the pixel coordinate system. Specifically, an incorrect positioning of the rotary table can be identified by determining the distance between holes as well as the absolute positions of the holes. The reference data evaluation of the geometrical parameters test is implemented in the CTSimU Toolbox [16, 17].

4.2.1. Heuristic approach   How does one define conformity intervals for the rotary table position test? The definition
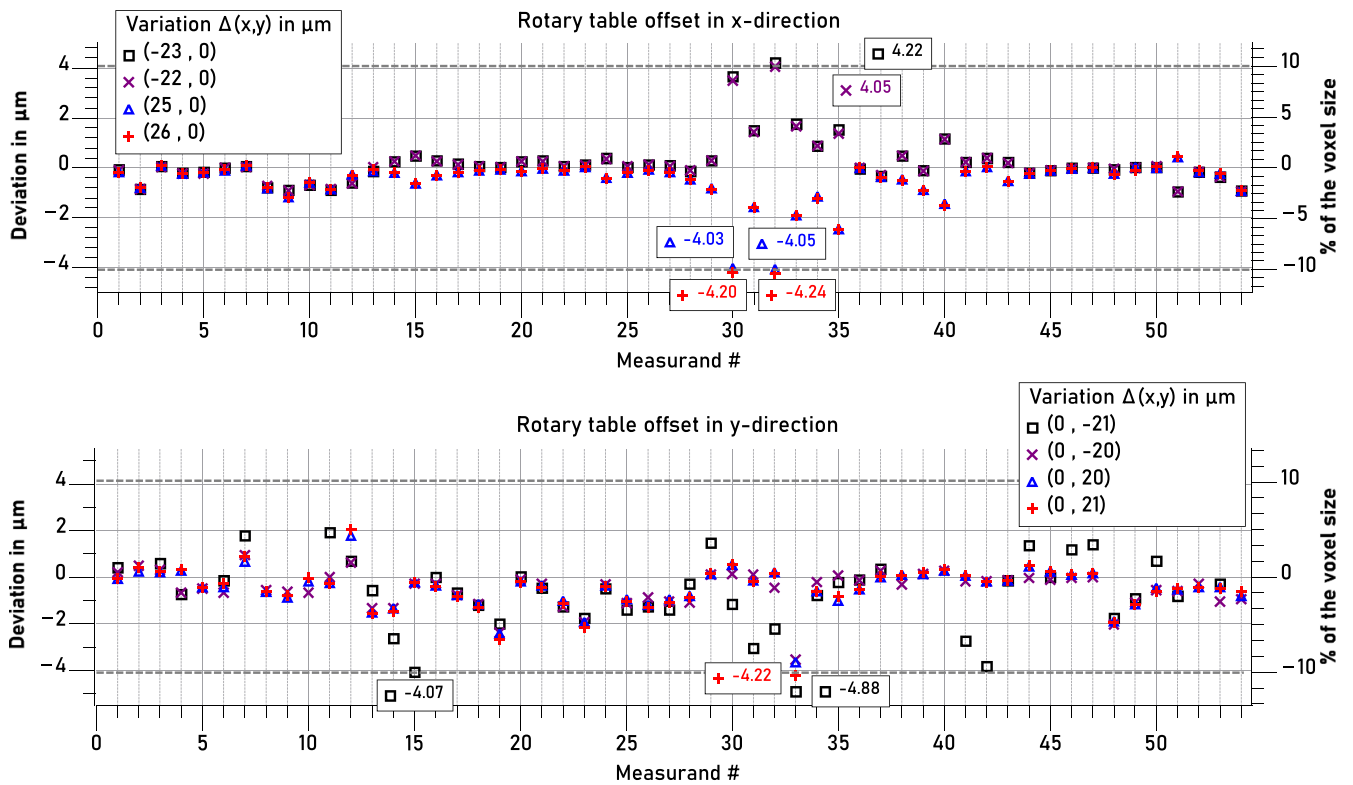
Figure 8. Results of the heuristic method (threshold = 10% × voxel size) for determining conformity intervals for the offset of the rotary table in the *x*- and *y*-directions.

of conformity intervals for this test is based on the approach presented in section 3.1 and on the heuristic threshold method. The generic test scenario with the multi-geometry cuboid presented in section 3 was used for this test. The position of the rotary table was varied stepwise (in 1 µm steps) and compared to the nominal position (source-object-distance – SOD = 127.5 mm) in the direction of the magnification (*x*-) axis and also perpendicular to the rotation axis (i.e. in the *y*-direction) separately, see figure 4 for the coordinate system. For every new position, a new 3D simulation was carried out and the results compared with the measurement results for the nominal position (difference between varied and nominal measurement result). If this difference (for a single measurand) exceeded the heuristic threshold of 10% × voxel size (i.e. 4.08 µm), this indicated that the associated variated position already exerted a significant influence on the measurement results and hence that this position could not be accepted as tolerable. The last tolerable value is therefore used for the conformity interval.

4.2.2. Conformity interval for the position of the rotary table
The results of the simulations with varied rotary table positions in the *x*- and *y*-directions (cf figure 4) are presented in figure 8. It shows the deviations of the 54 measurands from their nominal values for four different simulated distances of the rotary table in the *x*- (figure 8 top) and *y*-directions (figure 8 bottom) from its nominal position of 127.5 mm from the source.

For the sake of simplicity and graph legibility, only the results of the variations −23 µm, −22 µm, 25 µm and 26 µm in the *x*-direction and −21 µm, −20 µm, 20 µm and 21 µm in the *y*-direction are plotted. In the *x*-direction, the −22 µm and 25 µm variations resulted in deviations of less than 4.08 µm, as did the −20 µm and 20 µm variations in the *y*-direction. On the other hand, the results from −23 µm and 26 µm (measurands 32. Dis_Cal11–Sph42 and 30. Dis_Cal11–Sph41) in the *x*-direction and from −21 µm and 21 µm (measurand 33. Dis_Poi51-1 c–Poi52-1 c) in the *y*-direction exceeded the threshold of 10% × voxel size, indicating a significant influence on the measurement results. The last tolerable results (below 4.08 µm) were therefore used for the conformity interval: for the *x*-direction SOD = 127.478 mm (offset of −22 µm) and 127.525 mm (offset of 25 µm) and for the *y*-direction an offset of −20 µm and 20 µm.

Now that it is clear which 3D positional variations are tolerable, this 3D conformity interval must be transferred to the actual 2D test because the 2D test cannot measure the 3D geometry parameters directly. Instead, it uses the position and vector orientation of the projected holes on the detector plane to conclude whether or not the hole sheet is positioned correctly. The 2D test does not measure the hole sheet's position in 3D space, but uses very different, derived parameters: the mean distance between the holes to check for scale deviation; the mean vector rotation angle to check for rotational misalignments; and a mean translation vector to check for positional misalignments. Because all three of these parameters

Table 2. Results of the 2D data evaluation with the hole sheet standard using the data processing toolbox. The selected values used for the conformity interval are shown in bold. The selection of the values in bold was aided by the graphs presented in figures 9 and 10.

	Unit	Nominal conditions	$\Delta(-21, 0)$ µm	$\Delta(25,0)$ µm	$\Delta(0, -21)$ µm	$\Delta(0, 20)$ µm
Mean scale deviation		$2.40 \cdot 10^{-6}$	$\mathbf{4.18 \cdot 10^{-4}}$	$-4.94 \cdot 10^{-4}$	$2.08 \cdot 10^{-5}$	$-1.32 \cdot 10^{-5}$
Standard deviation of scale deviation		$1.33 \cdot 10^{-5}$	$1.40 \cdot 10^{-5}$	$5.30 \cdot 10^{-6}$	$1.30 \cdot 10^{-5}$	$1.20 \cdot 10^{-5}$
Mean rotation angle	°	$-2.88 \cdot 10^{-5}$	$-9.23 \cdot 10^{-5}$	$-2.49 \cdot 10^{-5}$	$-1.88 \cdot 10^{-4}$	$8.10 \cdot 10^{-6}$
Standard deviation of rotation angle	°	$9.06 \cdot 10^{-4}$	$8.74 \cdot 10^{-4}$	$4.06 \cdot 10^{-4}$	$6.28 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$
Mean translation vector	px	$1.01 \cdot 10^{-3}$ $-1.93 \cdot 10^{-3}$	$-5.46 \cdot 10^{-1}$ $-5.25 \cdot 10^{-1}$	$\mathbf{-4.63 \cdot 10^{-1}}$ $-4.81 \cdot 10^{-1}$	$6.19 \cdot 10^{-1}$ $\mathbf{-4.30 \cdot 10^{-1}}$	$-1.07 \cdot 10^{\pm 0}$ $-6.74 \cdot 10^{-2}$
Standard deviation of translation vector	px	$1.59 \cdot 10^{-2}$ $1.60 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$ $1.28 \cdot 10^{-2}$	$6.73 \cdot 10^{-3}$ $5.00 \cdot 10^{-3}$	$1.27 \cdot 10^{-2}$ $1.45 \cdot 10^{-2}$	$1.31 \cdot 10^{-2}$ $1.27 \cdot 10^{-2}$

are mean values for an ensemble of ten hole coordinates (and, more specifically, their possible combinations), they all come with a root mean square deviation (here called standard deviation). Those three standard deviations are important parameters for identifying if something went wrong in the 2D test. In total, this results in six distinct measurands for the 2D test.

To obtain a conformity interval for the 2D test, the geometry variations that were still deemed tolerable in the 3D test are used as deliberate misalignments of the hole sheet in simulations of the 2D test. Preliminary simulations have shown that the tested effects present a linear behaviour in 3D and 2D. The resulting projection images are run through the evaluation process, and the evaluation results from the Image Processing Toolbox [16, 17] then constitute the 2D conformity intervals for each of the six measurands of the 2D test. Each 2D measurand can react with a different sensitivity to the simulation of the still tolerable variations (found from the results of the 3D simulations). *The question now is how to select a conformity interval based on a single projection of the still tolerable variation for the 2D measurands?* The criterion used must take into account that the observed error should be large enough to be considered significant but at the same time small enough that a slightly imprecise reproduction of the effects is 'caught' in the 2D test. Thus, the smallest variation whose absolute mean result minus three times its standard deviation is still larger than the absolute mean of the nominal condition plus three times its standard deviation is taken as the conformity interval (see equation 8):

$$\left| \overline{\mathrm{Var}}_i - \overline{\mathrm{Nom}} \right| > 3 \times \sigma_{\mathrm{Nom}} + 3 \times \sigma\_\mathrm{Var}_i \qquad (8)$$

$\overline{\mathrm{Nom}}$ and σ_{Nom} are the respective mean and standard deviation acquired with nominal conditions. $\overline{\mathrm{Var}}_i$ and σ_Var_i are the respective mean and standard deviation acquired with the variated parameter values.

The results of the 2D measurands of the conical hole sheet are presented in table 2. In bold are the first results found to fulfil the criterion presented in equation (8) (i.e. showing a significant change when compared with the nominal 2D simulation [positioning error = 0]) and these are used as the conformity

Table 3. Conformity interval of the 2D test for the 'position of the rotary table'.

	Unit	Conformity interval
Mean scale deviation		$[-4.18 \cdot 10^{-4}, 4.18 \cdot 10^{-4}]$
Mean translation vector	px	$[-4.63 \cdot 10^{-1}, 4.63 \cdot 10^{-1}]$ $[-4.30 \cdot 10^{-1}, 4.30 \cdot 10^{-1}]$

interval for the 2D test of the rotary table position, see table 3. The graphical representation of the 2D test results is presented in figure 9. The graphic results permit the visual identification of the lowest variation mean that satisfies the criterion presented in equation (8); in other words, the smallest mean variated value that causes a significant change in the results in 2D. The criterion compares 'only' the means of the results. There are, however, parameter variations that cause a stronger change in the standard deviation than the mean of the results. For this reason, the variances of the variated parameter value and are also compared with the nominal values. To verify if the variances between the nominal and variated values are significantly different, the ratio between the standard deviations (σ_2/σ_1 for $\sigma_2 > \sigma_1$) of the nominal and variated parameter values is used, with σ_2 representing the greater of the standard deviations obtained from the nominal or variated values. The smallest standard deviation for which the ratio σ_2/σ_1 is greater than 10 is used for the conformity interval of the rotary table position test. A graphical representation of these example results is presented in figure 10.

The graphs in figure 9 make clear that for the mean of the 2D measurands (scale deviation, rotation angle and translation vector), the smallest mean that fulfils the condition expressed by equation (8) is yielded when the position of the rotary table is mistakenly offset by -21 µm and 25 µm in the *x*-direction and -21 µm in the *y*-direction. No significant difference was observed for the standard deviations from the tested variations. Now that the influence of the effects causing a significant deviation in 3D on the 2D measurands is known, the conformity interval is taken based on the smallest results to guarantee that wrongly simulated parameters (even small errors) will be detected.
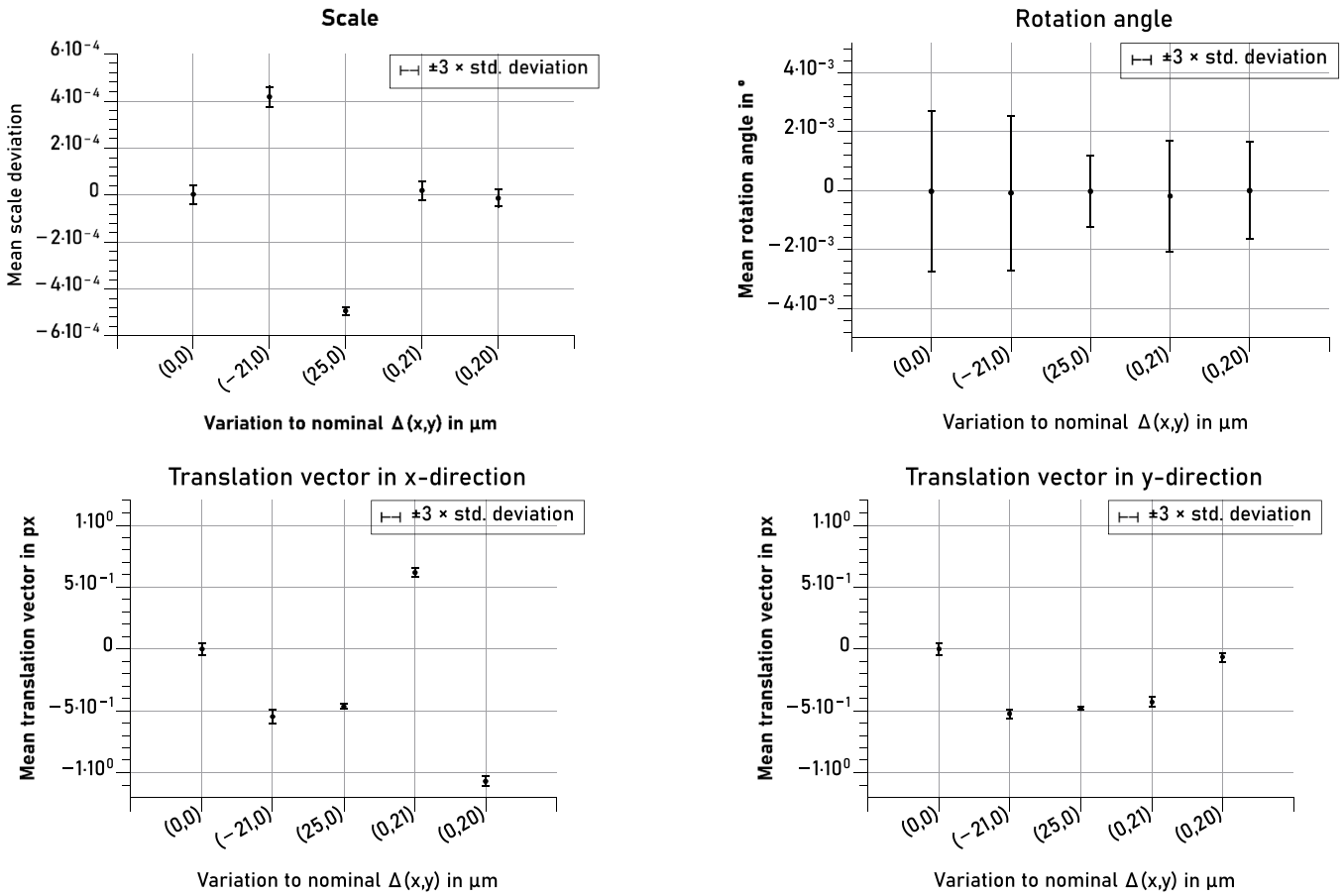
Figure 9. Graphical representation of the conformity interval determination based on the mean of the results.
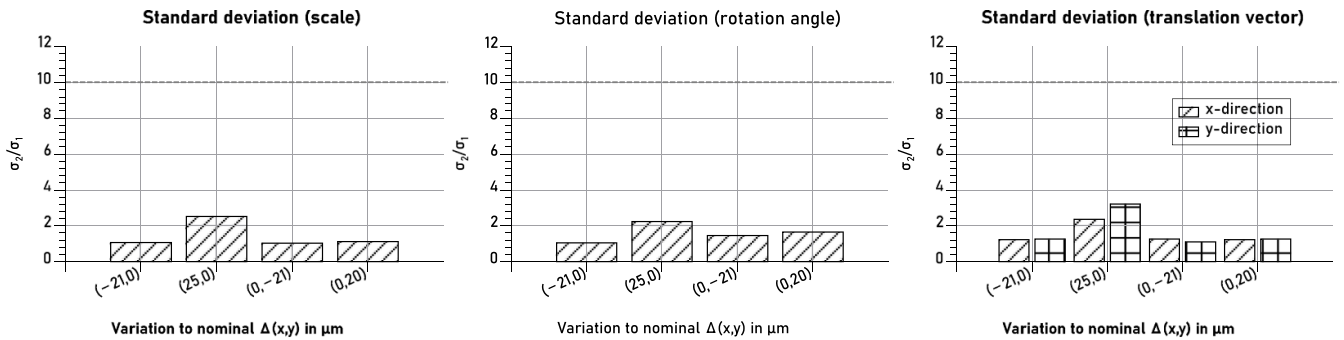

Figure 10. Graphical representation of the conformity interval determination based on the standard deviation of the results. σ_1 and σ_2 are the standard deviations of either the nominal or the variated parameter value, with $\sigma_2 > \sigma_1$ in order to achieve a ratio greater than one and verify if the variances between the nominal and variated values are significant.

5. Summary and discussion

The correct simulation of a measurement process is essential to ensure the ability to later estimate the measurement uncertainty using simulations of a digital twin.

A method for determining conformity intervals for the 2D testing of radiographic simulation software for physical laws, characteristic effects and basic functionalities was presented. The method is based on simulations of complete scans of the multi-geometry cuboid. Parameter values associated with

physical laws, characteristic effects or basic functionalities are varied, simulated and evaluated in 3D. Varied parameter values that are found to lead to a significant change of the measurement results serve as a basis for determining the conformity intervals. Two approaches were proposed in this contribution to determine whether a measurement change caused by a variation of a parameter value is considered significant: (a) a statistical approach and (b) a heuristic threshold. In (a) the results of the 3D simulations are compared statistically by means of a hypothesis test. In (b) a heuristic threshold depending on a

Figure 11. Official logo of the WIPANO funding Programm and the Federal Ministry of Economic Affairs and Climate Action

fixed fraction of the voxel size is applied. If the measurement deviation resulting from the varied parameter value was above this predefined threshold, the varied value was considered to influence the measurement result significantly. Based on this decision, the conformity intervals were determined.

Two application examples were presented for the determination of the conformity interval, one for the test of projection noise and one for the test of the rotary table's position. To verify the significance of the results, the statistical approach was applied to the projection noise example and the heuristic approach applied to the rotary table position.

As concerns the methods for testing the significance of the measurement result, the conclusion is that both approaches are promising and that they complement one another. One noteworthy advantage of the statistical approach is that the threshold depends on the variance of the effect being tested and on the measurand. However, the method is time-consuming as it requires a large number of simulations (per variation). Further, it is applicable only to non-deterministic effects since it considers the variance of the effect to create the threshold, and it involves relatively complex data evaluation. As for the heuristic approach, it is fast because it requires only one simulation (per variation), it involves simple and straightforward data evaluation, and it is generally applicable to distinct effects. On the other hand, the method works with a static threshold (depending only on the voxel size of the scan) that does not consider the effect variance, and this threshold was chosen empirically.

Besides the presented examples, the methods described in this contribution were applied to determine the conformity intervals of further effects, e.g. CT geometry, detector pixel sampling and unsharpness, and focal spot size. Beyond that, both methods will serve as the basis for the definition of conformity intervals in the CTSimU2 project for the evaluation of the CT system digital twins to be built during the project.

Credits

Markus Bartscher (MB), Florian Wohlgemuth (FW), Carsten Bellon (CB) and Stefan Kasperl (SK) contributed to funding acquisition and project administration. Fabrício Borges de Oliveira (FB), Tamara Reuter (TR), David Plotzki (DP), MB, FW, CB and SK contributed to the review and finalization of the full paper. FB, TR, DP and FW contributed to the conceptualization, methodology, investigation and formal analysis. FB, DP and TR wrote the original draft.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Fabrício Borges de Oliveira ⬤ https://orcid.org/0000-0002-3357-1286
Tamara Reuter ⬤ https://orcid.org/0000-0002-4300-1834
David Plotzki ⬤ https://orcid.org/0000-0002-3641-2635
Markus Batscher ⬤ https://orcid.org/0000-0002-2361-5863
Carsten Bellon ⬤ https://orcid.org/0000-0002-3861-8426
Stefan Kasperl ⬤ https://orcid.org/0000-0002-8118-7609

References

[1] Kruth J-P, Bartscher M, Carmignato S, Schmitt R, De Chiffre L and Weckenmann A 2011 Computed tomography for dimensional metrology *CIRP Ann.-Manuf. Technol.* **60** 821–42

[2] BIPM 1995 *JCGM 100:2008 Evaluation of measurement data—Guide to the expression of uncertainty in measurement* (available at: www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6) (Accessed 13 March 2023)

[3] ISO 15530-3:2011 2011 Geometrical product specifications (GPS)—coordinate measuring machines (CMM): technique for determining the uncertainty of measurement—part 3: use of calibrated workpieces or measurement standards

[4] VDI/VDE 2617 Blatt 8: 2018-10 2018 Genauigkeit von Koordinatenmessgeräten—Kenngrößen und deren Prüfung—Prüfprozesseignung von Messungen mit Koordinatenmessgeräten

[5] VDI/VDE 2630 Blatt 2.1: 2015-06 2015 Computertomografie in der dimensionellen Messtechnik — Bestimmung der Messunsicherheit und der Prüfprozesseignung von Koordinatenmessgeräten mit CT-sensoren

[6] Bundesanstalt für Materialforschung und—prüfung aRTist—analytical RT inspection simulation tool (available at: http://artist.bam.de/) (Accessed 16 December 2022)

[7] Hiller J, Fuchs T O J, Kasperl S and Reindl L M 2011 Einfluss der Bildqualität röntgentomographischer abbildungen auf koordinatenmessungen: grundlagen, messungen und simulationen *Tech. Mess. Tm* **78** 334–47

[8] Reiter M, Erler M, Kuhn C, Gusenbauer C and Kastner J 2016 SimCT: a simulation tool for x-ray imaging *iCT2016* (Wels) p 7

[9] van Aarle W, Palenstijn W J, De Beenhouwer J, Altantzis T, Bals S, Batenburg K J and Sijbers J 2015 The ASTRA Toolbox: a platform for advanced algorithm development in electron tomography *Ultramicroscopy* **157** 35–47

[10] EMPA Center for x-ray Analytics—3D-Image analysis and simulation (available at: www.empa.ch/web/s499/rt1-3d-image-analysis-and-simulation) (Accessed 29 August 2022)

[11] Hiller J 2011 Abschätzung von unsicherheiten beim dimensionellen Messen mit industrieller Röntgen-Computertomographie durch simulation *PhD Thesis* Universität Freiburg

[12] Bircher B A, Meli F, Küng A and Thalmann R 2019 CT geometry determination using individual radiographs of calibrated multi-sphere standards *iCT 2019* (Padova) p 7 (available at: www.ndt.net/article/ctc2019/papers/iCT2019_Full_paper_43.pdf) (Accessed 31 August 2022)

[13] Fuchs P, Garbe C S and Kröger T 2019 Self-supervised learning for pore detection in CT-scans of cast aluminum parts *DIR 2019* (Fürth) p 10 (available at: www.ndt.net/article/dir2019/papers/We.1.A.1.pdf)

[14] Matern D, International Y and Bogen E 2019 Analysis of cone beam artefact influences with respect to calibration of metrology qualified x-ray computed tomography systems *iCT 2019* (Padova) p 10 (available at: www.ndt.net/article/ctc2019/papers/iCT2019_Full_paper_13.pdf) (Accessed 13 March 2023)

[15] BIPM JCGM 101:2008 *Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement"—Propagation of distributions using a Monte Carlo method* (available at: www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c) (Accessed 13 March 2023)

[16] Plotzki D *et al* 2022 The CTSimU software toolbox for CT-related image processing and quality assessment *iCT 2022* (Wels) p 7 (available at: www.ndt.net/article/ctc2022/papers/ICT2022_paper_id242.pdf) (Accessed 13 March 2023)

[17] Plotzki D 2022 CTSimU software toolbox. Bundesanstalt für materialforschung und—prüfung (available at: https://github.com/BAMresearch/ctsimu-toolbox) (Accessed 31 August 2022)

[18] Lehrstuhl für Fertigungsmesstechnik der Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) *Project WIPANO-CTSimU* (available at: www.ctsimu.forschung.fau.de/wipano-ctsimu/ctsimu/wipano-ctsimu1/) (Accessed 13 March 2023)

[19] Reinhart C 2008 Industrial computer tomography—a universal inspection tool *17th World Conf. on Nondestructive Testing* (Shanghai) p 10

[20] Carmignato S 2012 Accuracy of industrial computed tomography measurements: experimental results from an international comparison *CIRP Ann.-Manuf. Technol.* **61** 491–4

[21] Borges de Oliveira F, Reuter T, Plotzki D, Bartscher M, Hausotte T and Bundesanstalt P-T 2022 Novel reference standards for the qualification of radiography-based computed tomography simulation software *iCT 2022* (Wels) p 12 (available at: www.ndt.net/article/ctc2022/papers/ICT2022_paper_id243.pdf) (Accessed August 31 2022)

[22] Reuter T, Borges de Oliveira F, Plotzki D and Hausotte T 2022 Influence of detector misalignments on different geometrical and dimensional measurands using a dedicated test specimen *iCT 2022* (Wels) p 7 (available at: www.ndt.net/article/ctc2022/papers/ICT2022_paper_id216.pdf) (Accessed August 31 2022)

[23] Reuter T, Plotzki D, Oliveira F B D and Hausotte T 2021 Simulative untersuchung des einflusses von bildrauschen auf dimensionelle messungen mit industriellen computertomografen *presented at the DGZfP-Jahrestagung 2021* p 11 (available at: https://jahrestagung.dgzfp.de/portals/jt2021/bb176/inhalt/p8.pdf) (Accessed 31 August 2022)

[24] Welch B L 1947 The generalization of `student's' problem when several different population variances are involved *Biometrika* **34** 28–35