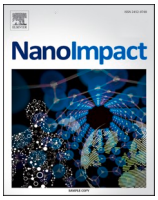
NanoImpact 27 (2022) 100410

Contents lists available at ScienceDirect

NanoImpact

journal homepage: www.elsevier.com/locate/nanoimpact



Reproducibility of methods required to identify and characterize nanoforms of substances

Richard K. Cross [a,*], Nathan Bossa [b], Björn Stolpe [c], Frédéric Loosli [d], Nicklas Mønster Sahlgren [e], Per Axel Clausen [e], Camilla Delpivo [b], Michael Persson [c], Andrea Valsesia [f], Jessica Ponti [f], Dora Mehn [f], Didem Ag Seleci [g], Philipp Müller [g], Frank von der Kammer [d], Hubert Rauscher [f], Dave Spurgeon [a], Claus Svendsen [a], Wendel Wohlleben [g]

[a] UK Centre for Ecology and Hydrology, Pollution, Wallingford, Oxfordshire, United Kingdom
[b] LEITAT Technological Center, Carrer de la Innovació 2, 08225 Terrassa, Barcelona, Spain
[c] Nouryon, Bohus, Sweden
[d] Environmental Geosciences, Centre for Microbiology and Environmental Systems Science, University of Vienna, Wien, Austria
[e] The National Research Centre for the Working Environment, Copenhagen, Denmark
[f] European Commission, Joint Research Centre (JRC), Ispra, Italy
[g] BASF SE, Department of Material Physics and Department of Experimental Toxicology & Ecology, Ludwigshafen, Germany

ARTICLE INFO

Editor: Dr. Bernd Nowack

Keywords:
Nanoform
Similarity
Reproducibility
Grouping

ABSTRACT

Nanoforms (NFs) of a substance may be distinguished from one another through differences in their physicochemical properties. When registering nanoforms of a substance for assessment under the EU REACH framework, five basic descriptors are required for their identification: composition, surface chemistry, size, specific surface area and shape. To make the risk assessment of similar NFs efficient, a number of grouping frameworks have been proposed, which often require assessment of similarity on individual physicochemical properties as part of the group justification. Similarity assessment requires an understanding of the achievable accuracy of the available methods. It must be demonstrated that measured differences between NFs are greater than the achievable accuracy of the method, to have confidence that the measured differences are indeed real. To estimate the achievable accuracy of a method, we assess the reproducibility of six analytical techniques routinely used to measure these five basic descriptors of nanoforms: inductively coupled plasma mass spectrometry (ICP-MS), Thermogravimetric analysis (TGA), Electrophoretic light scattering (ELS), Brunauer–Emmett–Teller (BET) specific surface area and transmission and scanning electron microscopy (TEM and SEM). Assessment was performed on representative test materials to evaluate the reproducibility of methods on single NFs of substances. The achievable accuracy was defined as the relative standard deviation of reproducibility (RSD_R) for each method.

Well established methods such as ICP-MS quantification of metal impurities, BET measurements of specific surface area, TEM and SEM for size and shape and ELS for surface potential and isoelectric point, all performed well, with low RSD_R, generally between 5 and 20%, with maximal fold differences usually <1.5 fold between laboratories. Applications of technologies such as TGA for measuring water content and putative organic impurities, additives or surface treatments (through loss on ignition), which have a lower technology readiness level, demonstrated poorer reproducibility, but still within 5-fold differences. The expected achievable accuracy of ICP-MS may be estimated for untested analytes using established relationships between concentration and reproducibility, but this is not yet the case for TGA measurements of loss on ignition or water content. The results here demonstrate an approach to estimate the achievable accuracy of a method that should be employed when interpreting differences between NFs on individual physicochemical properties.

* Corresponding author.
E-mail address: riccro@ceh.ac.uk (R.K. Cross).

https://doi.org/10.1016/j.impact.2022.100410
Received 30 September 2021; Received in revised form 24 June 2022; Accepted 25 June 2022
Available online 3 July 2022
2452-0748/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The registration of nanoforms (NFs) of substances according to REACH (European Parliament and Council, 2006; European Commission, 2018) requires their identification by composition, surface functionalisation or treatments, particle size distribution, specific surface area, and shape (European Commission, 2018). We designate these properties in the following as NF basic descriptors. If one can justify that the assessment of hazard, exposure and risk can be performed jointly for certain NFs, one can register these NFs as "sets of similar NFs" that must have "clearly defined boundaries" in each of the NF basic descriptors (ECHA, 2019; Janer et al., 2021b). If the hazard assessment for a certain endpoint can be performed jointly, one can apply the GRACIOUS framework or other relevant approaches to justify a grouping of several NFs (Stone et al., 2020) – and again one needs to describe the limits of the group by the NF basic descriptors.

Each single NF is polydisperse in each of the NF basic descriptors and has a batch-to-batch variability (Mülhopt et al., 2018). The raw data, e.g. the size distribution, is typically reduced to scalar descriptors, e.g. the median size, which is then compared between NFs (Jeliazkova et al., 2022) Any variability in the NF basic descriptors that goes beyond batch-to-batch variability creates another nanoform and therefore it is essential to understand the significance of measured variabilities. From the metrological point of view, the NF concept therefore requires an exact knowledge of the significance of deviations and of the reproducibility of measurements between independent laboratories to make conclusions on the similarity of different NFs. Moreover, the need to register ranges of the descriptors due to batch-to-batch variation raises the repeatability issue, and the comparison of NFs from different manufacturers must consider if differences are significant, before NFs can be rejected from specific sets or groups (Janer et al., 2021a).

Similarity assessment on individual properties requires understanding of the measurable range of the property, the width of acceptable similarity and the required accuracy of the assay (Jeliazkova et al., 2022). The width of acceptable similarity about the mean value of a property when comparing two or more NFs is a balance between achievable accuracy of the analytical method, and knowledge of the sensitivity of a toxicokinetic, environmental fate or toxicity endpoint that is modulated by the property of the nanoform (Loosli et al., 2021). In other words, for two NFs to be considered sufficiently dissimilar on the basis of a single property, it must be demonstrated that the difference measured between the two NFs is:

a) of sufficient magnitude to be greater than the reproducibility or achievable accuracy of the method, and
b) of sufficient magnitude to result in a meaningful difference in the endpoint under consideration.

The present contribution focuses on the metrological aspect a) only, whereas assessing issue b) is the purpose of a grouping framework and not in scope here.

To estimate the achievable accuracy of a given method, the reproducibility of measurement of a property on a single nanoform must be evaluated under conditions that mimic routine practice within the field. To this end, we assess the reproducibility of six analytical techniques that are routinely used to describe these five basic descriptors of a nanoform (Table 1). Representative test materials are employed to assess the reproducibility of measurements on single nanoforms. It should be noted that the reproducibility of a method will be material specific. Therefore, the achievable accuracy estimated herein for different techniques is specific only to the representative test materials assessed and the specific protocols followed. These representative test materials represent a range of commonly used nanomaterial classes evaluated in the literature and can be viewed as indicative of the achievable accuracy of these materials.

Here, a small inter-laboratory comparison was performed.

Table 1
Scheme of reproducibility test across different methods. [1] The original NanoDefine-SOP for size and shape determination (based on TEM) was found to be inappropriate for the fibre material. Instead, another SOP based on SEM was applied in all laboratories for the fibre material. [2] Surface potential is not required as part of NF registration but is a recommended property in grouping considerations.

Analytical method	Basic descriptors of the nanoform	Descriptors/ effect level endpoints	Nanoforms tested	Number of participating laboratories
ICP-MS	Composition	Impurities: elemental composition (atomic percentage)	SiO_2-NM200 Silica_silane	3
TGA-MS	Composition (often attributed to surface treatments)	Percentage of organic constituents	CeO_2-NM212 SiO_2-NM200 Silica_silane $BaSO_4$-NM220	3
TEM + SEM	Size median, Size distribution and Shape category	Diameter (Feret minimum), length and aspect ratio: D10, D50, D90 AR10, AR50, AR90	CeO_2-NM212 (TEM) CNT Mitsui7 (SEM[1])	3
BET	Specific surface area	Specific surface area (by mass)	CeO_2-NM212	4
ELS	Surface potential [2]	Iso-electric point (IEP)	CeO_2-NM212	4

Participating laboratories re-analysed a benchmark material such as CeO_2-NM212 and another NF that is suited to challenge the measurement method and to differ significantly from the CeO_2 in the property measured. The reproducibility check was performed by triplicate measurements of each listed material by each listed participant, using the Standard Operating Procedure (SOP) of the leading lab. Importantly, the objective of this study is to examine the reproducibility of a single analytical technique, not to optimise specific protocols for each analyte tested and so a single SOP was adhered to across each participating laboratory. The following Table 1 summarizes for each property the test materials, methods, and participating laboratory. For traceability, the full repository codes for each representative test material assessed are as follows: CeO2-NM212-JRCNM02102a; ZnO-NM110-JRCNM62101a; SiO2-NM200-JRCNM02000a; BaSO4-NM220-JRCNM50001a; MWCNTs-Mitsui7-JRCNM40011a. For practical reasons, we use the generic abbreviated forms throughout the test to refer to these representative test materials, for example, we refer to CeO2-NM212-JRCNM02102a simply as CeO_2-NM212.

2. Methods

2.1. Analytical methods and protocols for measurement of select physicochemical properties

Below follows a brief overview of the methods followed for each of the 5 analytical techniques employed in the reproducibility testing as provided by the leading laboratory for each analytical method. The detailed standard operating procedures followed for each method are available on the Zenodo platform (accessed at https://zenodo.org/communities/horizon2020-gracious/). The selection of methods used to characterize the 5 basic descriptors of nanoforms was informed by the refinement of descriptor array for nanoforms with most relevance for grouping nanoforms in the GRACIOUS Framework (Loosli et al., 2021). A summary of the results for each method is presented in section

4 of the Supplementary Information.

2.1.1. ICP-MS impurities analysis

The silica-based nanoforms SiO_2-NM200 (JRC, Italy) and (silane-modified colloidal silica from Nouryon, the Netherlands) were analysed for 6 impurities: Cr, Mg, Mn, Ni, Pb and Sr. These were selected to exemplify trace level impurities that may need to be monitored for their contribution to toxicity of the nanoform, particularly elements such as Cr and Ni for example. Selection of these two silica nanoforms for impurities analysis was informed by existing knowledge on semi-quantitative ICP-MS for SiO_2-NM200 (Rasmussen et al., 2013). Approximately 0.2 g of NM200 or 1 g of Silica_silane was weighed with a precision of ±0,0001 g and mixed with 4 ml hydrofluoric acid (HF) and 4 ml HNO_3 in a polytetrafluoreten (PTFE) bowl. The mixture was heated at 165 °C until dryness, 4 ml of HNO_3 was added and the mixture was again heated until dryness. The residue was dissolved in ultra-pure water (Millipore synergy >18 MΩ cm) with 1% HNO_3 and was analysed by ICP-MS (Agilent 7700) equipped with a collision cell with He gas to enable the analysis of elements suffering from spectral interferences.

Since no silica-based reference material with certified concentrations of the elements of interest was available, the trueness of the ICP-MS results was assessed by comparison with two other analysis methods, inductively coupled plasma optical emission spectroscopy (ICP-OES) and X-ray fluorescence (XRF), both at the Nouryon laboratory. Methods are reported in the Supplementary Information.

2.1.2. TGA organic and water content

The thermogravimetric analysis (TGA) of the NFs was performed to quantify a possible organic coating and water content using a STA 449 *F3 Jupiter* TGA coupled to a QMS 403 D *Aëolos* MS (both from Netzsch, Germany). Three representative test materials, CeO_2-NM212, SiO_2-NM200 and $BaSO_4$-NM220 (JRC, Italy) as well as Silica_silane (Azko-Nobel, Netherlands) were measured. Full details of the procedure are publicly available (Sahlgren et al., 2019). The TGA was carried out in an oxygen atmosphere (air) at a flow of 40 ml/min to ensure complete oxidation of the organic compounds. The temperature program for all samples was heating from 30 °C to 800 °C at a rate of 2.5 °C/min. The MS source and transfer line temperature was 300 °C. The sample holders (crucibles) used for the TGA measurements were made of alumina (Al_2O_3) and had a volume of 3.4 ml. Sample masses were 2–60 mg and samples were not conditioned to equilibrate with known air humidity. Data was corrected for buoyancy.

For some materials with high surface area such as nanoform structures, there can be more water in the internal structure. This may lead to delayed evaporation as the water takes longer to diffuse through the material, e.g. SiO_2-NM200 (SI Fig. 1). However, this is not considered critical for the screening method used in this work to investigate the coating on the nanomaterial. The analysis of the TGA thermograms are based on the NANoREG protocol 2.04 (NANoREG, 2018), where water-loss assumed up to 100–110 °C, organic-loss from 110 to 500 °C and from 500 to 800 °C losses are assumed to be loss on ignition. In cases where we observe water losses after 110 °C it would be possible for future experiments to introduce 5–10 min of waiting time to ensure complete evaporation of any water from the sample.

This method was developed to be used as a screening method to select materials for other investigative instrumentation, for more detailed analysis.

2.1.3. Electrophoretic light scattering (Zeta-potential and isoelectric point)

The Zeta-potential (ZP) and isoelectric point (IEP) of the representative test material CeO_2-NM212 (JRC, Italy) were determined by electrophoretic light scattering (ELS) using Malvern Nano ZS. The measured electrophoretic mobility is converted into the zeta-potential using the Smoluchowski approximation. Full details on the protocol can be found in the SOP on the Zenodo platform (Loosli, 2020).

Briefly, a 100 mg/l stable-NM dispersion (meaning that the-NM hydrodynamic diameter remains constant as a function of time) was prepared from a well vortex 10 g/l stock dispersion by dispersing the particles in a low monovalent electrolyte (1 mM $NaNO_3$) at pH 3 (acidic condition) and sonicated to favour-NM dispersion. Possible modification of the CeO_2 through sonication at this low pH was not monitored in this study, however, this does not contradict the aims of the assessment, which are to produce dispersion conditions for the nanoform in which IEP determination can be reproduced across laboratories. The pH was adjusted with NaOH (1, 0.1 or 0.01 M depending on the pH value to reach) and the ZP determined for all pH units in the pH domain 3 to 11. The system was left to equilibrate for 20 min before each pH adjustment and measurement. The ZP of the-NM dispersions were measured at 25 °C. After a temperature equilibrium time of 2 min the ZP was measured 3 times under an applied voltage of 150 V with 20 runs and a delay between measurements of 1 min. Quality control certified standards (Malvern Zeta potential transfer standard) were run prior to analysis. The trueness of the protocol was evaluated determining if there was significant difference between results provided from the measurements and the certified value of a certified standard. The absolute bias, Δ_m, which corresponds to the difference between the mean measured value, C_m, and the certified value, C_{CRM}, and its expanded uncertainty, U_Δ were calculated. They were no significant differences between the measurement value and the certified value with ($\Delta_m \approx 0.6$ which is $\leq U_\Delta$).

2.1.4. BET specific surface area

The determination of the specific surface area (SSA) of the representative test material CeO_2-NM212 (JRC, Italy) by the Brunauer-Emmett-Teller (BET) method was carried out in an Quantachrome NOVA 2000e, an Autosorb 6 Quantachrome Instrument, a Micromeritics-Gemini VII 3.04 and a Micromeritics Tristar II Plus by UNIVIE, LEITAT, JRC and Nouryon, respectively following the SOP based on ISO 9277.

The CeO_2-NM212 was outgassed under vacuum at 300 °C for 700 min prior the surface area measurement. The SSA was calculated using the multi-points (6 points at C/C0 = 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3) method on the N_2 adsorption isotherm.

2.1.5. Electron microscopy (TEM and SEM) for size and shape

Electron microscopic analysis was used to determine the NF particle ferret minimum diameter, length and aspect ratio. CeO_2-NM212 (JRC, Italy) analysis was done by TEM using a TEM Tecnai Spirit 120 kV, Jeol JEM-2100 and Thermo/Talos F200i, for LEITAT, JRC and BASF, respectively. The carbon nanotube CNT Mitsui7 (Mitsui & Co. Ltd., Japan, JRC Nanomaterials Repository, European Commission, JRC, Italy) analysis was performed by SEM using a FE-SEM, Merlin (Zeiss), FE-SEM, Nova600 (Thermofisher) and Zeiss Gemini for LEITAT, JRC and BASF, respectively. Triplicate TEM grids were prepared. From the CeO_2 dispersion (200 mg/l in ethanol ultra-sonicated bath for 5 min), 6 µl was let dry on a 200 mesh Cu grid coated with carbon. From a CNT suspension (250 mg/l in ethanol ultra-sonicated bath for 15 min), 10 µl were dropped onto a Si wafer on a SEM stub (1 cm^2) holder then and allowed to air dry.

The bright field signal was used for image formation. Particle analysis was done both manually and using automated approaches representing two commonly used protocols. Image analysis was performed using manual and automatic methods in the software ImageJ. Automated approaches are increasingly able to account for agglomeration artefacts, but there are still some instances where verification of media diameter through manual analysis can be prudent (Holzwarth and Ponti, 2020). Automated analysis did not set the number of particles for analysis a priori. A minimum of 100 particles were measured in the manual analysis, obtaining an intermediate precision similar to that achieved for CeO_2-NM212 in the evaluation of TEM as an approach to measure particulate nanomaterials through interlaboratory comparison

between 19 laboratories (Verleysen et al., 2019). For the reproducibility assessment, the mean ferret diameter, length and aspect ratio calculated either manually or automatically for triplicate individual samples of CeO_2-NM212 and CNT Mitsui 7, was evaluated across participants.

Two automatic methods were used for the different nanoparticles: NanoDefine ParticleSizer plug-in for ImageJ for CeO_2-NM212 (Mech et al., 2019) and the ridge detection plug-in for CNT Mitsui7, which extends the ridge / line detection algorithm described in (Steger, 1998). ParticleSizer was developed as a plugin for ImageJ as a method for automated analysis of electron micrographs and has been validated for nanomaterials containing varying degrees of overlap between particles (Verleysen et al., 2019). This validation concluded that the ParticleSizer is fit for purpose for measuring size of nanoparticles and has a negligible contribution to measurement uncertainty as compared to the other sources of uncertainty across the whole measurement process (including sample preparation and analysis). To account for agglomeration artefacts when drying the sample to the grid, the irregular watershed technique was used to determine the primary constituent particle size where there is overlap between particles as recommended in the NanoDefine SOP. Additional details on the protocol used for automated measurement of the two nanoparticles can be found in SI Section 3.

2.2. Statistical evaluation of the data

2.2.1. Assessment of repeatability and reproducibility

For each analytical method, the selected representative test materials were measured in at least triplicate at each participating laboratory under repeatability conditions. To assess the repeatability and reproducibility of each test method, data was subset by both the representative test material, but also the analyte/measurand (where multiple analytes were measured in a single test material, for example multiple impurities analysis by ICP-MS, or multiple measurements such as both length and diameter with TEM). The basic statistical model is a one-way analysis of variance (ANOVA), from which the between group variance (or model mean squares MS_b) and the within group variance (or residual mean squares MS_w) can be extracted.

A statistical F-test can be used to determine if between-group variability is significantly larger than the within-group variability, i.e., if the results from the different laboratories are significantly different. The F variable is calculated from:

$$F = \frac{MS_b}{MS_w} \tag{1}$$

If the calculated F is smaller than the critical F value (Fcrit) for the current degrees of freedom, the difference between the laboratories is not statistically different.

If assumptions of normality and homogeneity of variance are not met for the dataset, Welch's F-ratio adjustment was used.

Following the guidance in the Eurachem Guide: The fitness for purpose of analytical methods (Magnusson and Örnemark, 2014), the repeatability and intermediate precision can be calculated using the mean squares from the ANOVA. Briefly, the repeatability standard deviation is calculated following Eq. (2) where S_r is the repeatability standard deviation within groups, and MS_w is the residual mean squares within groups (i.e. the standard deviation within repeat measurements across laboratories).

$$S_r = \sqrt{MS_w} \tag{2}$$

The between-laboratory standard deviation (S_b) is calculated in Eq. (3) where MS_b is the mean squares between groups, MS_w the residual mean squares within groups and n is the mean number of replicate observations when reporting results.

$$S_b = \sqrt{\frac{MS_b - MS_w}{n}} \tag{3}$$

From these two equations, the intermediate precision (also known as the reproducibility standard deviation) S_I can be calculated as in Eq. (4).

$$S_I = \sqrt{S_r^2 + S_b^2} \tag{4}$$

The intermediate precision (S_I) represents the sum of the within group variance and between group variance. Intermediate precision refers to conditions mimicking those under which the method will be used routinely (i.e. different analysts, equipment and days, Magnusson and Örnemark, 2014). This is a useful parameter as it is a measure of the absolute value of reproducibility for the measurand. In particular relative standard deviations (%) are inflated for values closer to the analytical limits of detection and so comparison of the absolute standard deviation for different NFs in a similarity assessment can be more meaningful in these instances.

From S_r and S_I the relative standard deviation (%RSD) may be calculated for both the repeatability (%RSD_r) and reproducibility (%RSD_R) demonstrated in Eqs. (5) and (6), where $\bar{y}$ is the grand mean across all data for the analyte/measurand.

$$\%RSD_r = \frac{S_r}{\bar{y}} * 100 \tag{5}$$

$$\%RSD_R = \frac{S_I}{\bar{y}} * 100 \tag{6}$$

The repeatability and reproducibility relative standard deviations are discussed for each method and interpreted in the context of the required accuracy of measurements to distinguish between nanoforms. The achievable accuracy, as indicated by the RSD_R, must be below the required accuracy of rules for similarity (where these exist), in order for data to be considered fit for purpose for similarity assessment on individual properties.

2.2.2. Fold difference between laboratories

The maximum fold difference between the means calculated by each participating laboratory is also reported as fold-differences are often utilised as a first assessment of similarity on single properties (for example, density as described in Park et al., 2018) or the median size D50, specific surface area, and other properties in the ECETOC NanoApp (Janer et al., 2021a, 2021b), and even more complex interactions in dissolution rates or in vitro NOAELs (Jeliazkova et al., 2022). For an analytical approach to be of the required accuracy it must be able to distinguish fold differences smaller than those suggested in existing rules for similarity and grouping approaches. In this instance we calculate maximal fold differences as the maximum measure (maximum value of $laboratory_{a,\ b}$ etc.) divided by the minimum measure, (minimum value of $laboratory_{a,\ b}$ etc.) to derive a fold-difference > 1 as demonstrated in Eq. (7).

$$Fold\ difference = \frac{max(laboratory_a, laboratory_b, etc.)}{min(laboratory_a, laboratory_b, etc.)} \tag{7}$$

2.2.3. Reproducibility as a function of concentration

It is established that reproducibility can be a function of the concentration of the analyte measured, with increased RSD_R as concentration of the measured analyte as a proportion of the material decreases, irrespective of the analytical method being used. As a useful guide, Eq. (8) allows estimation of the Expected RSD_R as a function of the concentration of the analyte, expressed as a decimal fraction (Heyden and Smeyers-Verbeke, 2007). Here, C is the concentration of the analyte expressed as a decimal fraction, i.e. for a pure sample $C = 1$.

$$Expected\ RSD_R(\%) = 2C^{-0.1505} \tag{8}$$

For ICP-MS and TGA analysis, this equation can act as a useful guide to provide context for the RSD_R calculated for these two methods. Eq. (8) has been proposed to be independent of the analyte or method used,

derived as it is from examination of over 50 interlaboratory studies conducted by the AOAC (Association of Official Analytical Chemists) on various commodities, ranging from a few percent concentration to ng/g concentrations (Horwitz et al., 1980; Horwitz and Albert, 1995). This relationship has been revisited more recently, and whilst at low concentrations ($<1.2 \times 10^{-7}$ kg/kg) this relationship seems to overestimate the Expected RSD_R, it has been confirmed as appropriate for relative concentrations $\geq 1.2 \times 10^{-7}$ and ≤ 0.138 (Thompson, 2000), the range assessed using ICP-MS and TGA in this contribution.

3. Results and discussion

3.1. Composition

This section considers the results for assessment of composition, both of metal impurities (through ICP-MS) and composition of what is often attributed to surface treatments of the nanoform (through TGA), two of the basic descriptors required for registering a nanoform.

3.1.1. Reproducibility of analysis of impurities: ICP-MS

Analysis of impurities by ICP-MS in both SiO_2-NM200 and the silica_silane NF found remarkably low variation between repeat measurements within each laboratory. Good reproducibility was also found between laboratories across all of the measured impurities. This was the case even when the absolute levels of these impurities were low and despite the many potential sources of variation that arise through replication of measurement between laboratories (Fig. 1). To assess the trueness of the ICP-MS results, the samples were also analysed with ICP-OES and XRF. The mean concentrations of Mg, Cr, Mn, Ni, Sr and Pb in both SiO_2_NM200 and Silica_silane measured by ICP-MS at Univie and Nouryon were all within 95–117% of the concentrations measured by ICP-OES (SI Table 1 and Table 2). For XRF, the concentrations of Na, Al,

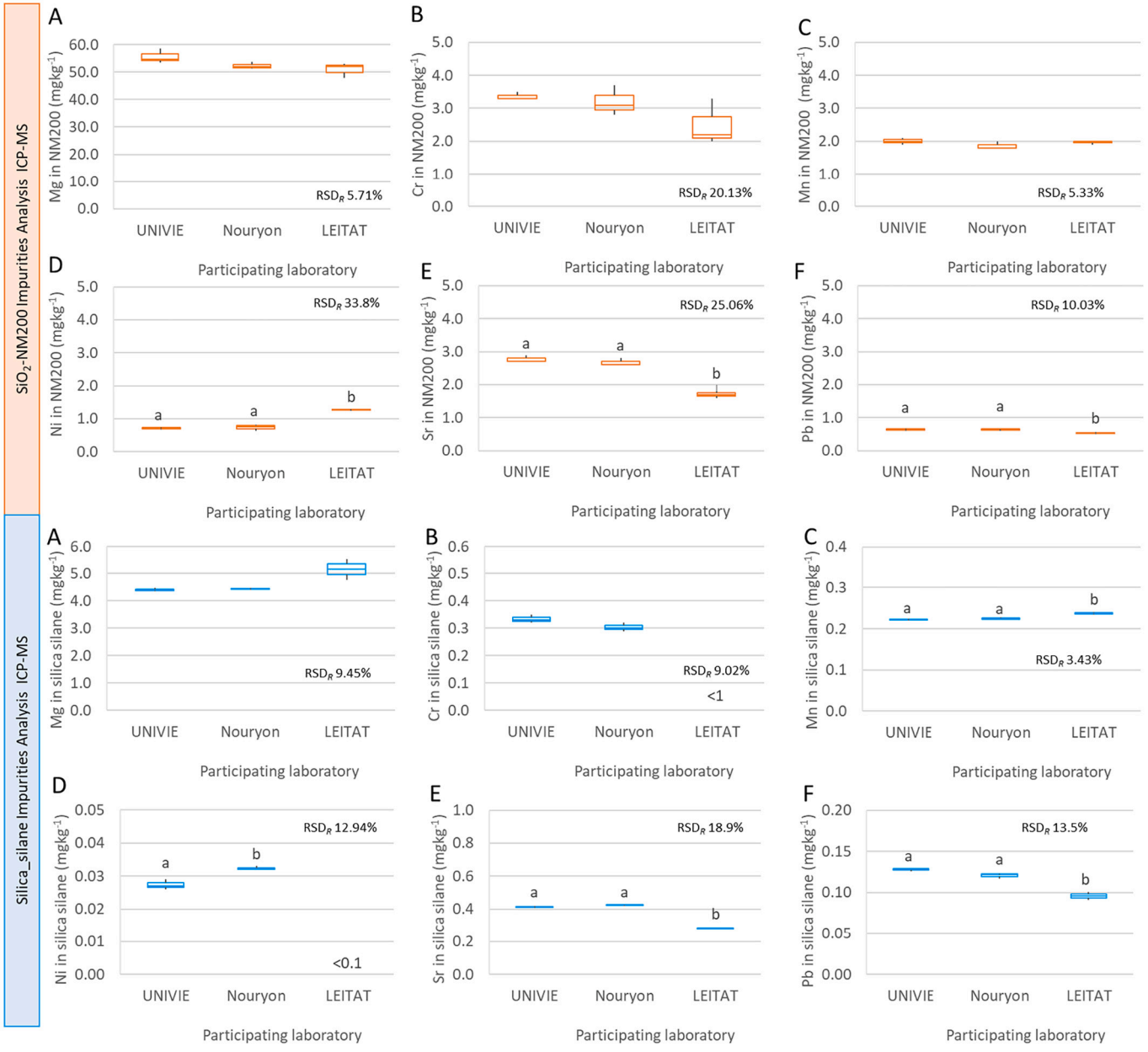


Fig. 1. Summary of impurities elemental analysis in SiO_2-NM200 (orange plots) and silica_silane (blue plots). Isotopes measured for each element are represented in figures A) ^{24}Mg, B) ^{52}Cr, C) ^{55}Mn, D) ^{60}Ni, E) ^{88}Sr and F) ^{208}Pb. Box plots represent the median, interquartile ranges and the minimum and maximum for replicate samples measured at each laboratory. Significant differences (ANOVA, $p < 0.05$), where demonstrated, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Ca and Fe were used for comparison, since the other elements were below the detection limit. The concentrations of Na (in both SiO_2_NM200 and Silica_silane), Ca and Fe (in SiO_2_NM200) measured by ICP-MS at Nouryon were within 99–108% of the concentrations measured by XRF, while the concentration of Al (in SiO_2_NM200) was 17% higher when measured by ICP-MS (SI Table 1 and Table 2). Mg was also qualitatively assessed in the JRC's own characterisation report for this representative material SiO_2-NM200 and found to be a similar levels between 0.001 and 0.005% (Rasmussen et al., 2013).

The assessment of impurities in these two silica NFs demonstrates excellent repeatability of this method, even at these low concentrations of <0.01%. The RSD_r in SiO_2-NM200 was consistently <10% with the exception of Cr which had an RSD_r of 16.13% (SI Table 3). Likewise for silica_silane, repeatability was even better, with the highest RSD_r being that of Cr, at 5.6% (SI Table 4).

In SiO_2 NM-200, only Ni and Sr had a much higher RSD_R than expected, given the relationship between RSD_R and the proportional concentration of the impurity in the material, whilst in silica_silane, all impurities had a similar or improved RSD_R than expected from Eq. (8) (Horwitz and Albert, 1995). This indicates that for the majority of impurities evaluated, reproducibility of the measurements was good, and ICP-MS measures these elements with high accuracy and precision.

No significant differences in measured concentrations of Mg and Cr were found between laboratories (ANOVA, $p > 0.05$) in either SiO_2-NM200 or the silica_silane NFs (orange plots and blue plots respectively in Fig. 1A and B). For the remaining impurities, all measurements across laboratories were within 2-fold.

The good reproducibility between laboratories for all impurities indicates that the RSD_R calculated for these elements is representative of the achievable accuracy with which impurities may be determined. Given this assumption, variation within 30% might be considered an appropriate measure of achievable accuracy (being the highest RSD_R observed for any of the impurities evaluated, SI Table 3), equating to a 1.8 fold difference in impurities, thus defining the minimum difference that must be observed between NFs to consider them dissimilar in their impurity composition. Other considerations such as the biological implications of >1.8 fold differences in impurities must also be considered when justifying similarity between NFs, but this is beyond the scope of this contribution to the topic.

In the context of existing grouping approaches which have decision rules based on impurities, this data is supportive of existing similarity rules and demonstrates that ICP-MS is a sensitive method which can achieve the required reproducibility to allow for differentiation between nanoforms according to exiting rules for similarity in impurities content. For example, The ECETOC NanoApp considers that for NFs to be considered in the same set, the maximal content of impurities should be 3% in all NFs under comparison (Janer et al., 2021b). For abundant impurities (>2% relative content), the impurity should be within 2-fold across NFs. Even for very low-level impurities analysed in this study, a maximal fold difference of ~1.8 was achieved. Therefore, the achievable accuracy for ICP-MS measurement of impurities is considered sufficient to meet the requirements for existing similarity assessment rules.

3.1.2. *Reproducibility of thermogravimetric analysis (TGA) screening of loss on ignition and water content, attributed to the surface composition of the nanoforms*

Reproducibility testing of the four nanomaterials (SiO_2-NM200, silica_silane, $BaSO_4$-NM220, and CeO_2-NM212) was performed to test the ability of TGA to detect and quantify the loss on ignition (a screening assessment for potential organic additives or surface treatments) and adsorbed water across three laboratories, performed in triplicate. The TGA-MS exploits the mass-loss occurring during controlled heating of a sample. Initially, water and smaller organic molecules will evaporate and later when temperature is above the flash point further organic molecules, including organic surface treatments may be combusted releasing CO_2 and water that can be detected by the MS. Some organic molecules may survive the heating process and when volatilised, be visible in the mass spectra. Mass loss at 30–100 °C is considered to be due to water only. Mass loss above this temperature is referred to as "loss on ignition" and may be assumed to be attributable to organic material including any intentionally added organic coating. However, it should be noted that attribution to potential organic coating is purely from a screening perspective. Several other phenomena can lead to mass-losses above 100 °C and further elucidation requires further analysis coupling mass spectrometry to the TGA which was beyond the scope of this study.

Neither the loss on ignition nor H_2O content (Fig. 2) of either silica NF (SiO_2-NM200 or silica_silane) differed in the measured concentrations between laboratories (ANOVA, $p > 0.1$). The RSD_R(%) for loss on ignition in the two silica materials was ~15%. Loss on ignition in the two remaining materials was more variable, with RSD_R of 38.7 and 54.2% respectively for $BaSO_4$-NM220 and SiO_2-NM212 (SI Table 5).

Similar to ICP-MS, the RSD_R is expected to be a function of the relative concentration of the analyte as a proportion of the total mass of the material. This may go some way to explain the higher RSD_R for both loss on ignition and H_2O content in CeO_2-NM212, 54.2 and 116.4% respectively, as compared to the other NFs, with both loss on ignition and H_2O content representing <1% of the total mass of the material.

For several materials, the variance arising within laboratories was greater than that from between laboratory differences, meaning a RSD_R could not be calculated. In these cases we assume the RSD_r (%) is equivalent to the RSD_R, with negligible contribution of between laboratory differences to the overall reproducibility. RSD_r (%) was consistently <16% for all materials except for CeO_2-NM212. Large variation in results of CeO_2-NM212 may be due to several factors including strongly reduced dehydrated surfaces after long-term storage under argon, mixture of Ce valence states in the material, and nonstoichiometric composition, which can result in both mass-gains and losses during transfer to ambient air and heating. As well as the presence of organic surface treatments, impurities or additives, the loss on ignition may be attributed to baking out of structural water or other anions annealing during the heating. Indeed, it was the water content of CeO_2-NM212 that was the least reproducible measurement (RSD_R 116.4%), likely due to a combination of factors including those listed above, and the low proportional mass of H_2O for this material, 0.07%.

The variability in the TGA-MS data on CeO_2_NM212 calls for additional investigations to understand which phenomena (e.g., sample inhomogeneity, impurities, non-stochiometric composition, oxidative-reductive reactions etc.) may explain the observed TGA-MS results. Several other mass-loss phenomena require further analysis considering also effect of previously reported impurity phases (OECD, 2015).

3.1.3. *Expected achievable accuracy can be estimated as a function of the proportional mass of the analyte – Implications for fold difference similarity rules as a function of concentration*

For properties relating to compositional differences between NFs (metal impurities, loss on ignition and H_2O content etc.) the effect of the relative mass of the analyte compared to the total material must be considered, with poorer reproducibility observed as the proportional contribution of the analyte to the total mass of the NF decreases. RSD_R as a function of concentration has been described in Eq. (8) and this relationship between the Expected RSD_R and concentration is plotted in Fig. 3. The majority of measurements for ICP-MS achieved an RSD_R similar to or lower than the expected RSD_R indicating high reproducibility. The blue and red highlighted areas represent where we would expect differences >2-fold and > 3-fold respectively to be necessary to distinguish between NFs (i.e. where the expected RSD_R is equivalent to a 2-fold and 3-fold deviation about the mean).

One of the possible algorithms for similarity assessment evaluated in (Jeliazkova et al., 2022) is the x-fold approach to quantifying similarity between NFs on individual properties. With such an approach, the acceptable limits of similarity, i.e. the fold difference within which NF may still be considered similar on assessment of a single property is often

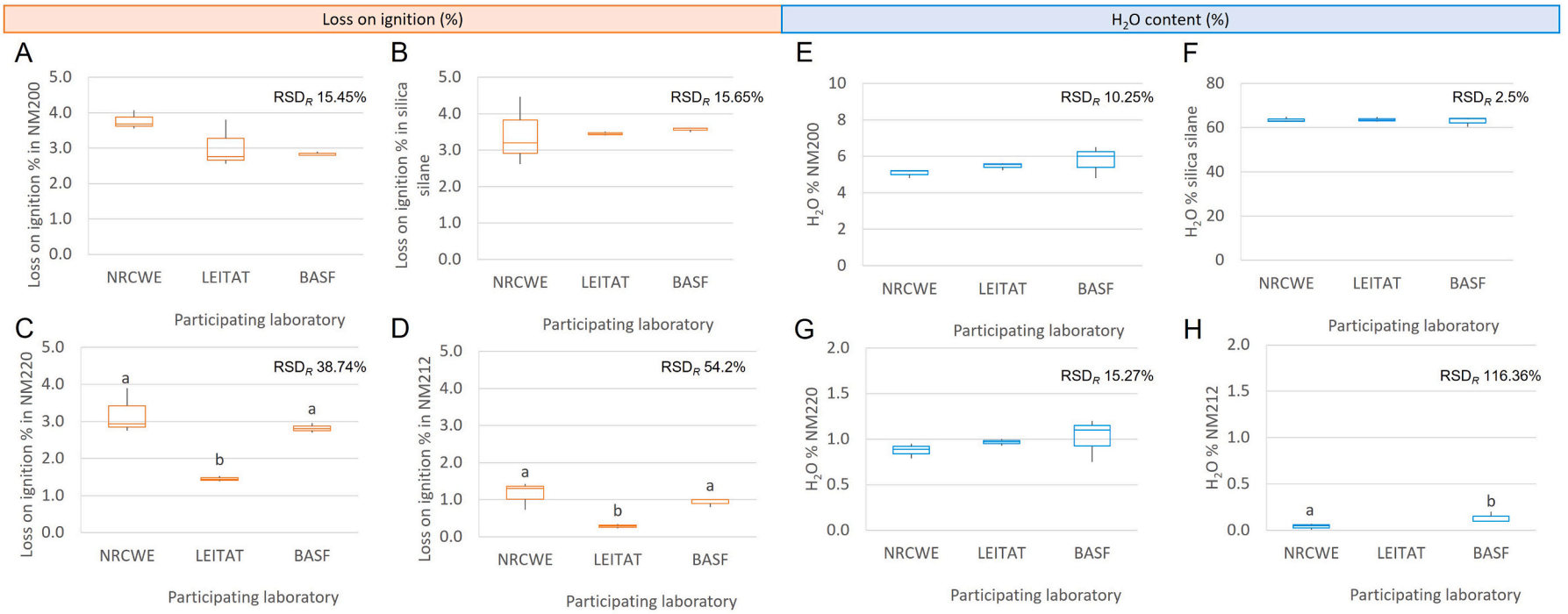


Fig. 2. Summary of the loss on ignition (% mass, orange box plots) measured by TGA for A) SiO_2-NM200, B) Silica_silane, C) $BaSO_4$-NM220 and D) CeO_2-NM212 and of the H_2O content (% mass, blue box plots) for E) SiO_2-NM200, F) Silica_silane, G) $BaSO_4$-NM220 and H) CeO_2-NM212. Box plots represent the median, interquartile ranges and the minimum and maximum for replicate samples measured at each laboratory. Significant differences (ANOVA, $p < 0.05$), where demonstrated, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

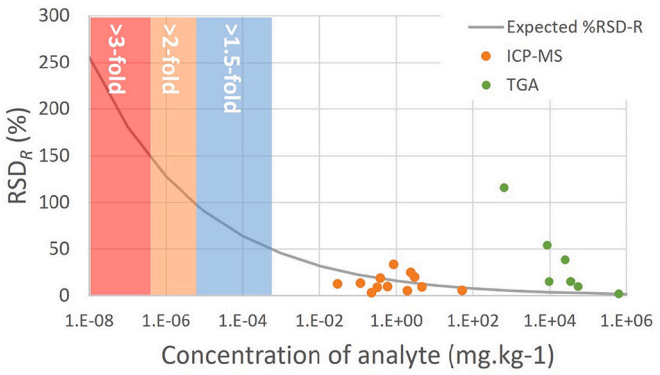


Fig. 3. The expected RSD_R (%) as a function of concentration as calculated by Eq. (8) (grey line). Orange data points represent the two dimensional data points from ICP-MS whilst blue dots TGA analysis for the grand mean concentration of analyte and RSD_R. The three shaded areas represent regions where 1.5-, 2- and 3-fold differences would need to be demonstrated to interpret differences between nanoforms based on the expected relationship between RSD_R and concentration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the range of 2 to 5 fold, or in some cases even greater. For example, the PATROLS project determined that when estimating the deposited dose for in vitro cell cultures an uncertainty factor between 2 and 10-fold can occur even for well dispersed NFs (Keller et al., 2021). The implication is that differences >2 and in some cases even 10 fold may be necessary to conclude that NFs are different for this endpoint.

All impurities measured had an RSD_R equivalent to a maximum of 1.8-fold differences between participating laboratories. As there was good agreement between the calculated and expected RSD_R for impurities measured by ICP-MS, the achievable accuracy for any impurity (beyond just those tested in this exercise) can be estimated for this method on the basis of the function between expected RSD_R and concentration of the analyte.

For TGA, the expected RSD_R calculated from Eq. (8) would not be a suitable estimate for achievable accuracy for this method. RSD_R was usually higher than the expected RSD_R for both loss on ignition and water content of the four tested nanoforms (Fig. 2). This method has been categorised as of medium technology readiness level (Loosli et al., 2021), which may in part explain the poorer reproducibility than that demonstrated for the well-established ICP-MS method. It is interesting to consider whether this poorer reproducibility for TGA than ICP-MS precludes its use as an approach to generate data which could group NFs on the relative content of organic surface treatments. The ECETOC NanoApp requires surface treatments, impurities or additives with a relative content of >2% of the total NF to also demonstrate below 2-fold difference between two NFs under comparison. Assuming that loss on ignition can be used as a screen for organic surface treatments, impurities and additives, this is achieved for SiO_2-NM200 and silica_silane, but not for $BaSO_4$-NM220 (RSD_R 38.74%, equivalent to a maximal 2.2 fold difference in loss on ignition between participating laboratories). This would suggest that further optimisation or standardisation of this method could be beneficial if it is to be used to generate data suitable for existing grouping frameworks. It should be noted that the RSD_R is representative of the total loss on ignition, and so is a screen for organic surface treatments, impurities or additives. Additional analysis using TGA-MS would identify the components that are lost on ignition, to assist in a more specific further assessment of similarity in the chemical identity and content of any organic surface treatment, impurities or additives between nanoforms that should be the focus of future studies.

3.2. Surface potential

Characterisation of surface functionalisation or treatments is one of the basic descriptors required for the registration of nanoforms (NFs) of substances according to REACH (ref to 1907/2006 and 2018/1881). We have demonstrated the reproducibility of TGA as part of the compositional assessment of surface treatments in the previous section. Measurement of the surface potential using electrophoretic light scattering, whilst not a direct measurement of the functionalisation or composition of surface treatments for a nanoform, is still a useful descriptor to take into account for grouping approaches (ECHA, 2017). Indeed, in the GRACIOUS Framework, whilst Zeta potential is not called for directly as a descriptor for similarity assessment, it is indirectly considered in the integrated approaches to testing and assessment (Loosli et al., 2021). Previous developments in similarity assessment have highlighted concerns that as Zeta potential is a function of the nanoform itself, but also the surrounding media, assessment of similarity would only be suitable on the basis of the individual property of Zeta potential if the measured Zeta potential was derived under exactly the same experimental conditions for all nanoforms in the assessment (Park et al., 2018). As it is unlikely that available data on Zeta potential of NFs would be conducted at the same pH or under the same media chemistry, we report here the reproducibility of the scalar descriptor isoelectric point (IEP), which would be a useful alternative to Zeta potential for similarity assessment if required in future grouping approaches, as it does not rely on the same pH conditions to have been tested between existing data sources.

3.2.1. Isoelectric point as a descriptor for surface potential

The reproducibility test dealing with the determination of the-NM Isoelectric point (IEP) and zeta potential (ZP) was investigated using the representative test material CeO_2-NM212 (JRC). Fig. 4A represents the average zeta potential value as a function of pH determined by all participants. The grand mean IEP across all participants is reported in the upper left corner of the figure, whilst average IEP per participating laboratory is summarised in the lower left corner of the figure. The IEP was determined by linear interpolation between the ZP values corresponding to the closest lower (ZP > 0 mV) and upper (ZP < 0 mV) pH unit around the IEP (ZP = 0 mV) value.

Fig. 4B graphically represents the calculated IEP for each participating laboratory. Lessons learnt from previous interlaboratory comparisons for Zeta potential (e.g. Lamberty et al., 2011 for silica nanoparticles), such as allowing equilibration prior to measuring led to good repeatability within laboratories. Whilst a significant difference in IEP was found for UNIVIE compared to the other participating laboratories ($F(3,10) = 24.12$, $p = 6.77E\text{-}05$, Tukey's HSD <0.05), this was still within a 1.12 fold difference to the pH of other calculated IEPs. Good reproducibility is found, with RSD_R of 5.89%, meaning the IEP can be measured to an accuracy of within a pH range of ~0.45 for the same nanoform. The logarithmic nature of pH means that differences between NFs with IEPs further from a neutral pH may have a lower RSD_R and so be easier to distinguish from each other. However, it should also be remembered that a change in 1 pH unit represents a 10-fold change in hydrogen ion activity and so smaller differences in the pH of the IEP towards the extremes of acidic or basic conditions need to be carefully considered when evaluating similarity between nanoforms for the purpose of risk assessment, taking into account the context of the likely conditions of the surrounding environments and how this might effect the expected behaviours of the nanoforms. Full details on the reproducibility assessment of IEP can be found in SI Table 6.

3.3. Specific surface area

The reproducibility test dealing with the determination of specific surface area (SSA) was investigated using CeO_2-NM212 particles. The determination of the ceria SSA by each participant shows a high precision for SSA determination with an SD of 0.4, 1.4, 0.6, and 0.3 m^2/g for UNIVIE, JRC, LEITAT, and Nouryon, respectively. The grand mean SSA across all participants was calculated as 27.22 m^2/g (Fig. 5), in excellent agreement with the JRC's stated SSA of 27.8 m^2/g as reported for this

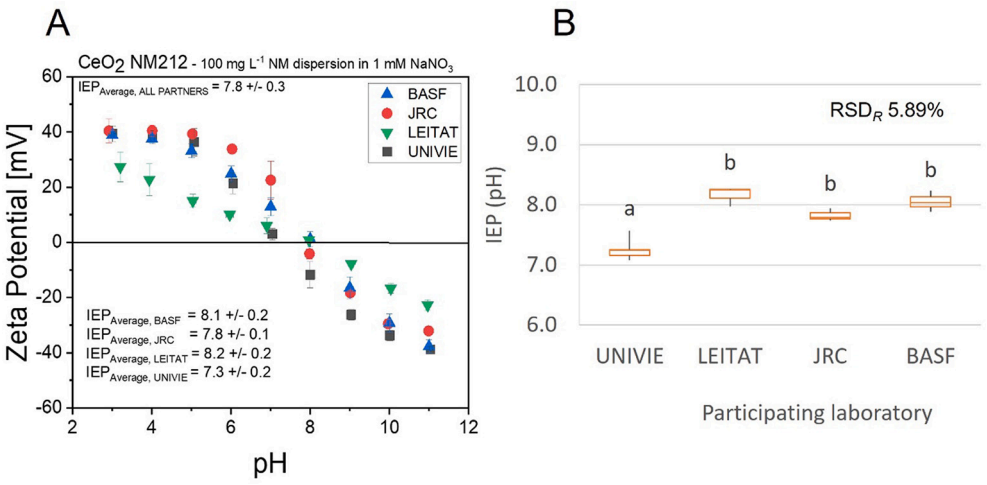


Fig. 4. A) CeO_2-NM-212 zeta potential (ZP) value as a function of pH performed by all participants. The isoelectric point (IEP) is also reported in the figure, as calculated for each participating laboratory. B) Summary of the replicated IEP calculated by each participating laboratory. Box plots represent the median, interquartile ranges and the minimum and maximum for replicate samples measured at each laboratory. Significant differences (ANOVA, $p < 0.05$), where demonstrated, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$).

nanoform in the "JRC Nanomaterials Repository:-NM-series of Representative Manufactured Nanomaterials" (Singh et al., 2014), represented as the blue line in Fig. 5.

All participating laboratories measured a similar SSA (no significant effect of participating laboratory on the SSA of $CeO_{2\text{-}NM}212$ measured, Welch's $F(3,8) = 3.465$, $p = 0.074$, SI Table 7). The achievable accuracy for BET measurements of SSA would be in the range of <5% RSD, with the RSD_R standing at 3.26%, equating to a total reproducibility standard deviation (S_I) of 0.9 m^2/g. This is in good agreement with a previous interlaboratory comparison study on TiO_2 P25 (Eonik Industries AG, Essen, Germany) which found a worst case RSD_r within laboratory variation of 3.96% (Hackley and Stefaniak, 2013), as compared to the 2.77% demonstrated here for CeO_2-NM212.

3.4. Particle size and shape

The following section examines the reproducibility of TEM measurement of the spheroidal nanomaterial CeO_2-NM212 and SEM measurements of the high aspect ratio nanomaterial, the carbon nanotube Mitsui7. Both the manual and automatic approaches to measuring particle size using these two techniques are first compared to evaluate these two approaches and their application when measuring mean particle diameter (TEM, CeO_2-NM212) and length (SEM, CNT Mitsui7). Further assessment of the reproducibility of these two analytical techniques in calculating D10, D50 and D90, three descriptors considered under the basic information requirements in the GRACIOUS Framework, are performed on data measured using the automated approaches only.

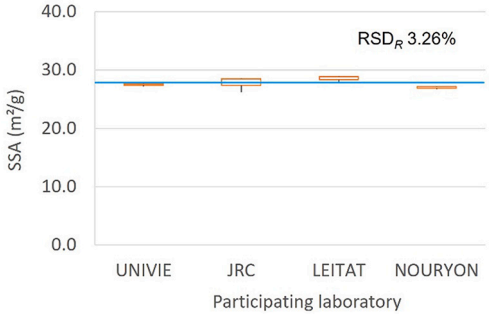


Fig. 5. Summary of the replicate BET measurements of specific surface area (SSA) for the representative test materials CeO_2-NM212. Box plots represent the median, interquartile ranges and the minimum and maximum for replicate samples measured at each laboratory. The blue line in represents the expected SSA based on the characterisation by the JRC in its report on characterisation and test item preparation for cerium dioxide-NM series. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4.1. Both manual and automated approaches to measuring mean particle dimensions using TEM and SEM are similarly reproducible

To verify that both manual and automatic analysis of particle dimensions are appropriate approaches to measuring the particle dimensions for both materials, the mean minimum Feret diameter and Feret length were evaluated for both CeO_2-NM212 (TEM, manual method versus NanoDefine automated approach) and for CNT Mitsui7 (SEM, manual method versus ridge detection plugin automated approach in ImageJ). Whilst mean minimum Feret diameter is not required for the registration of a nanomaterial, it is a descriptor that is often used in the literature to define the particle size of a nanoform and so is included in this analysis. The results of this comparison are presented in Fig. 6. Example micrographs of CeO_2-NM212 and CNT Mitsui7 are presented in SI Figs. 2 & 3.

Both the mean calculated for each replicate sample, but also the standard deviation that is due to each replicate measurement being an average of the characterisation of the dimensions of >100 individual particles, is presented in the figure. Neither approach (manual nor automatic) consistently outperformed the other, with consistent RSD_R across each approach. The within laboratory repeatability was also very high when comparing the mean dimension calculated for each replicate, with the RSD_r consistently <10%. This is in agreement with a previous interlaboratory evaluation of the ParticleSizer software, where between-laboratory precision was estimated at between 2 and 14% (Verleysen et al., 2019).

There was no significant effect of participating laboratory on the mean Feret minimum diameter measured for CeO_2-NM212 when measured manually ($F(2,6) = 3.2$, $p = 0.113$) or automatically ($F(2,5.52) = 7.51$, $p = 0.054$, SI Table 8). RSD_R for manually calculated CeO_2-NM212 was 8.18% whilst for the automated approach it was 10.93% (Fig. 6A and B). It should be noted that for the automated approach, the variance attributed to within laboratory variation was greater than between laboratory variation, with an RSD_R 13.27%. Therefore, an achievable accuracy of within ~10% of the mean particle diameter of spheroidal nanoforms might be considered appropriate when considering whether another NF differs in size from CeO_2-NM212. Differences less than this should be carefully interpreted.

When comparing the mean Feret diameter calculated for Mitsui 7 per triplicate measurement at each laboratory, there was a significant effect of participating laboratory, whether calculated by the manual approach ($F(2,6) = 14.69$, $p = 0.005$) or the automated approach using the ridge detection plugin ($F(2,6) = 126.2$, $p = 1.25E\text{-}05$). However, when these mean diameters per replicate are placed in the context of the underlying

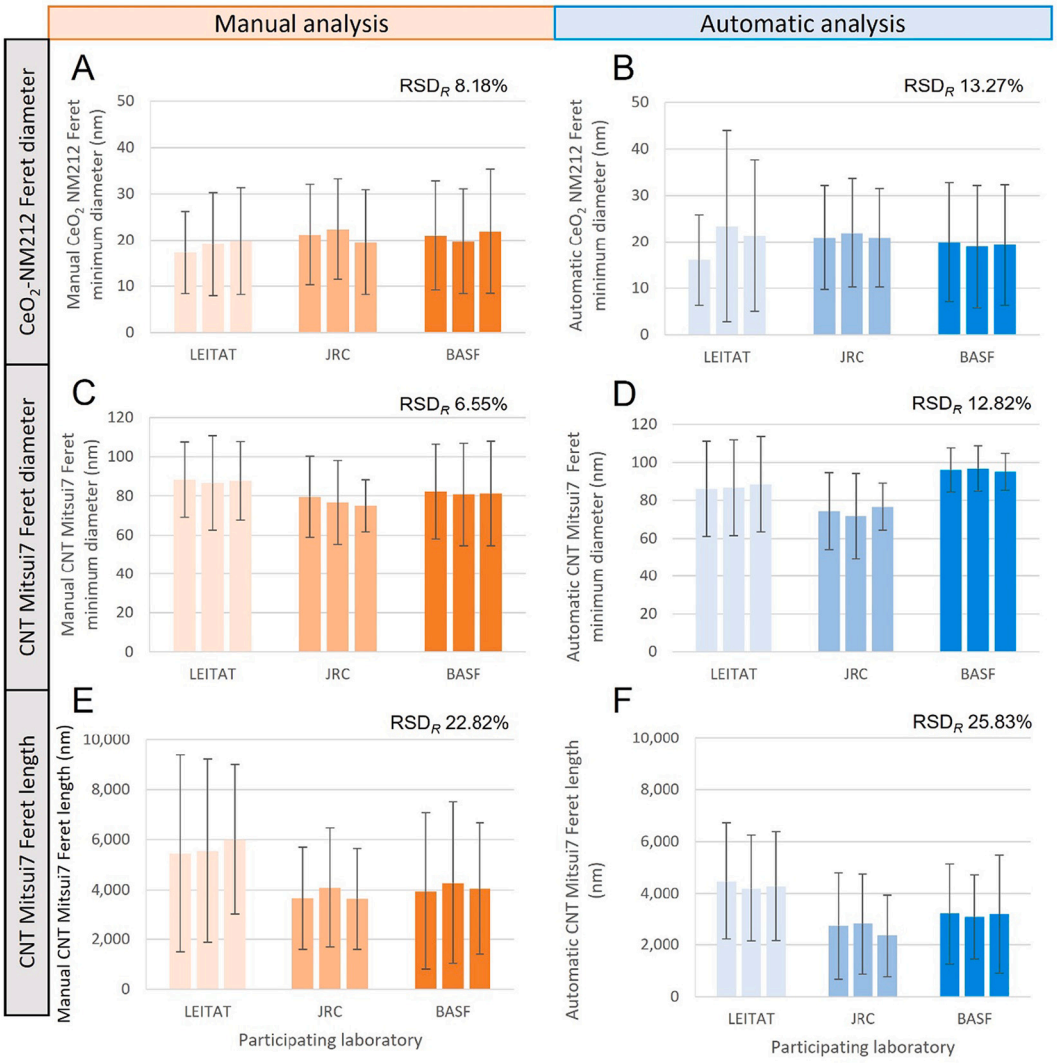


Fig. 6. Comparison between manual and automatic approaches to measurement respectively for CeO_2-NM212 minimum ferret diameter (A and B), CNT Mitsui7 minimum Feret diameter (C and D) and CNT Mitsui7 Feret length (E and F). Error bars represent the standard deviation of >100 measurements of individual particles. RSD_R is reported on each inset figure. Note, for CeO_2-NM212 the automated procedure used the NanoDefine ParticleSizer plug-in, whilst for CNT Mitsui7, the automated approach employed the ridge detection plug-in.

size distributions of the population of CNTs from which the mean diameter is calculated (Fig. 6C and D) it is apparent that there is significant overlap in the standard deviations across participating laboratories.

It should be noted that whilst the RSD_r and RSD_R were good when comparing the scalar descriptors (mean Feret minimum diameter and Feret length) across approaches, this does not reflect the actual diversity of the size distributions of the particle populations that were measured to derive these scalar descriptors. For example, whilst the RSD_r within laboratories was consistently <10%, it is clear from Fig. 6 that the standard deviations about the mean for each replicate are far higher than this, with a coefficient of variance usually >50%. This means it is likely that the reduction of particle population size distributions to scalar descriptors will be more sensitive to differences between nanoforms, than comparison of the size distribution of the particles as a whole. This will be discussed in further detail in reference to the reproducibility of the scalar descriptors which are used as part of a nanomaterial registration, D10, D50 and D90 in the following sections.

3.4.2. TEM measurement of D10, D50 and D90 (spheroidal particles, CeO_2-NM212)

TEM analysis indicates that primary CeO_2-NM-212 particles appeared aggregated, but primary particle constituent size could be measured using the irregular watershed approach in ParticleSizer for overlapping particles (SI Fig. 2). They show a polyhedral with irregular morphology and a broad size distribution, ranging from below 10 nm to well over 100 nm for all three participating laboratories. The pixel resolution at 20,000 x magnification was 0.3 nm. The calculated D10, D50 and D90 were < 100 nm when measured across all three participating laboratories (Fig. 7). Differences between laboratories were only statistically significant for D10 (ANOVA, Welch's $F(2,2.92) = 469.9$, $p < 0.05$).

Repeatability was high for D10 and D50, with RSD_r 6.55 and 6.46% respectively (SI Table 8). Reproducibility was best for D50 (RSD_R 9.49%) whilst it was poorer for both D10 and D90, RSD_R 43.53 and 28.3% respectively. This is to be expected, as the both the D10 and D90 will be influenced more by outliers or extremes in the particle population distribution and so reproducibility is expected to be worse, whereas the D50 of 16.79 nm is closer to the grand mean diameter across all measured particles of 20.34 ± 2.22 nm, as would be expected for

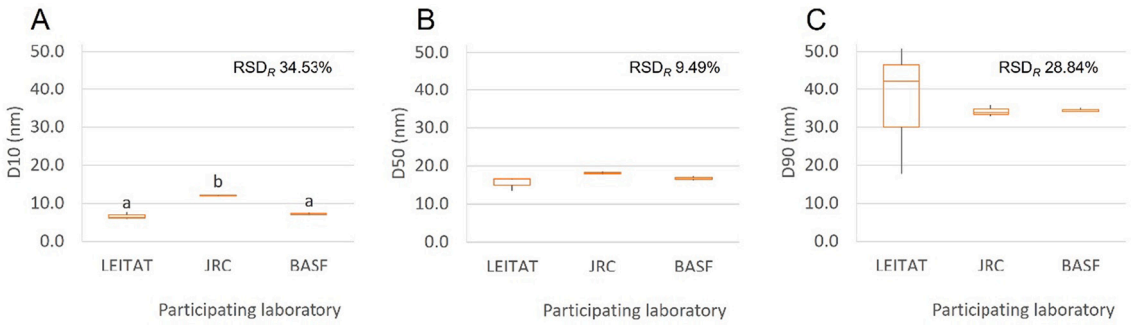


Fig. 7. Automated TEM analysis of CeO_2-NM212 using the automated NanoDefine ParticleSizer plug-in to calculate A) D10, B) D50 and C) D90. Significant differences (ANOVA, $p < 0.05$), where demonstrated for D10, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$).

spheroidal particles. A maximum fold difference of 1.17-fold was observed between laboratories measuring D50 for CeO_2-NM212. The high reproducibility of D50 and low fold difference between laboratories indicates this would be preferred as a single descriptor of size for characterisation of NFs, over the mean diameter or D10/90, as it would be less prone to type 1 errors, where statistically significant differences may be observed between measurements even on the same nanoform, such as observed for D10. If a similarity assessment between NFs is to be performed based on the individual property of particle size, additional descriptors such as other percentile values or information on the shape of the distribution may be needed to fully justify the assessment. In this way, the reproducibility of D10, D50 and D90 calculated here informs the achievable accuracy for each scalar descriptor, were they to be used in combination to assess the similarity between nanoforms based on the single property of size. For all laboratories, the percentage of particles <100 nm on a number basis were between 98 and 100%, correctly identifying these particles as a nanomaterial.

3.4.3. SEM measurement of D10, D50 and D90, length and aspect ratio (elongated particles, CNT Mitsui7)

As part of a nanoform registration, D10, D50 and D90 are required, whilst for elongated particles, the average length and aspect ratio must also be reported. NanoDefine ParticleSizer plug-in for ImageJ was first used to measure particle dimensions for CNT Mitsui7. This is in line with several grouping frameworks, for instance it is the recommended methodology for the nanoGRAVUR framework (Wohlleben et al., 2019), and is adopted in the GRACIOUS Framework also. However, use of the NanoDefine ParticleSizer plug-in, whilst validated for ellipsoidal and irregular particles (Verleysen et al., 2019), was found not to be suitable for this high aspect ratio, fibrous material. No efficient particle labelling was possible, due to the NFs presence as both individual and intertwined fibres, creating bundles. Rather, an automated standardized protocol was used named the "ridge detection plug-in" freely available in ImageJ, which extends the ridge / line detection algorithm described in (Steger, 1998). Example SEM micrographs and the ridge detection overlay are presented in SI Fig. 3.

The reported values for D10, D50 and D90 are all based on the diameter of the fibres derived from this automated ridge detection plugin. The mean length and aspect ratios were measured using both the automated ridge detection plugin, but also using a manual approach. Because of the high aspect ratio of CNT Mitsui7, the length and diameter of the particles were not measured at the same magnification. The reported aspect ratio is calculated as described in the NanoDefine method, from the ratio between the long side minimum bounding rectangle and the short side minimum bounding rectangle and is calculated for each replicate performed at each participating laboratory. The statistical output and calculations of relative standard deviations for D10, D50 and D90 for diameter measurements are summarised in SI Table 9 whilst length and aspect ratio are reported in SI Table 10.

SEM measurements of particle dimensions requires measurement of >100 individual particles, the distribution of which is represented by scalar descriptors such as the median (D50). It is these scalar descriptors which are compared in the reproducibility assessment, as it is these scalar descriptors which are called often upon, either in the registration of a nanoform, or in similarity or grouping approaches. It should be noted that when conducting a reproducibility assessment on these scalar descriptors, the underlying measurement uncertainty (e.g. the standard deviation about the mean length for each replicate, for example that demonstrated in Fig. 6) is not considered in these calculations. In this way, scalar descriptors are a sensitive endpoint with which to compare pairwise similarity between NFs. As with the other techniques, the achievable accuracy of the method is specific to the test material and SOPs followed here.

3.4.3.1. Diameter: D10, D50, D90. Diameter of the CNT Mitsui7 fibres was most reproducible when expressed as the D50 (Fig. 8B). Using D10 as a descriptor of fibre diameter was the least reproducible of the three scalar descriptors evaluated, with an RSD_r of 10.34% and RSD_R of 29.53%, whist D50 and D90 were similar, indicating negative skew in the data. Interestingly, statistically significant differences were observed in calculated D10, D50 and D90 between the participating laboratories. Nonetheless, the D50 measurements from all laboratories correctly identified the CNT Mitsui7 as a nanomaterial according to Recommendation 2011/696/EU on the definition of nanomaterial. These results indicate slightly improved reproducibility of the CNT measurement than in a wider recent interlaboratory study aiming to develop a specific OECD test guideline on particle size distributions of nanomaterials (Schmidt and Bresch, 2021). Our findings for the CNT Mitsui-7 support the report's conclusion that RSD for SEM measurements of fibre diameter will be specific to the material tested. They report relative error in the D50 of metal and metal oxide NFs as low as 20% but for multi-walled CNTs being much more variable at 42%, whilst here we report an RSD_R of 29.5% for the multi-walled Mitsui-7 CNT. We therefore recommend that best practice when using RSD_R to estimate the achievable accuracy of the method is to calculate this for the specific materials that are to be compared.

It is also important to recognise that the reproducibility analysis is performed only on the scalar descriptors of D10, 50 and 90 themselves, as calculated in triplicate by each participating laboratory. Whilst the repeatability of SEM to calculate these descriptors for particle size distributions is excellent (RSD_r of 10.34, 2.56 and 3.25% respectively for D10, D50 and D90, SI Table 9), this only considers variation between the calculated scalar descriptor for each replicate, not the standard deviation within each replicate measurement across all the particles measured. For example, in Fig. 6, the mean Feret minimum diameter is presented as calculated by the manual and automatic ridge detection approaches. Wide standard deviations exist for each replicate. These standard deviations are much greater than the repeatability RSD_r calculated for the scalar descriptors of diameter D10, 50 and 90. These scalar descriptors are more sensitive to differences between particle

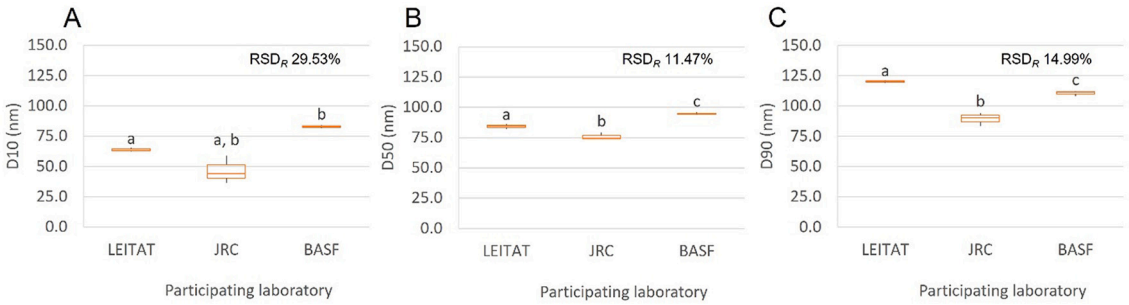


Fig. 8. Automated SEM analysis using automated ridge detection to calculate A) D10, B) D50 and C) D90 for Mitsui-7 CNT. Significant differences (ANOVA, $p < 0.05$), where demonstrated, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$).

populations, as observed in Fig. 8, where significant differences were observed between the calculated D50 by each laboratory, which would otherwise not be apparent when just comparing the mean diameter of the particles and the standard deviations within each replicate (Fig. 6). All descriptors (D10, 50, 90 and mean diameter) were found to be within ~1.78 fold of each other between participating laboratories. This is consistent with previous efforts to understand the reliability with which a material can be classified as a nanomaterial using existing methodologies, which found minimum Feret diameters measured by electron microscopy to be within a factor of 2 fold as a worst case (Babick et al., 2016). Importantly, in the current assessment we go beyond the single descriptor of minimum Feret diameter and also evaluate D10, 50 and 90, which have relevance for the registration of nanoforms. D50 was again the most repeatable and reproducible scalar descriptor for width of CNT Mitsui7, as it was for CeO_2 NM-212.

3.4.3.2. Length. When comparing the mean Feret length calculated for Mitsui 7 per triplicate measurement at each laboratory (Fig. 9A and B), there was a significant effect of participating laboratory (SI Table 10), whether calculated by the manual approach ($F(2,6) = 50.0$, $p = 0.0002$) or the automated approach using the ridge detection plugin ($F(2,6) = 80.32$, $p = 4.67E\text{-}5$). However, once again, when these mean lengths per replicate are placed in the context of the underlying size distributions of the population of particles from which the mean length is calculated (Fig. 6E and F) it is apparent that there is significant overlap in the standard deviations across participating laboratories and these significant differences between labs are therefore perhaps an artefact of the data reduction (from mean measurements for triplicates with associated standard deviations pre replicate, to the standard deviation for each laboratory representing the standard deviations across means for each replicate), thus ignoring some of the underlying variation that is attributable to the wide size distribution of fibre lengths in these materials.

3.4.3.3. Aspect ratio. When comparing the mean aspect ratio calculated for Mitsui 7 per triplicate measurement at each laboratory (Fig. 9C and D), there was a significant effect of participating laboratory, whether calculated by the manual approach (Welch's $F(2,3.51) = 14.76$, $p = 0.019$) or the automated approach using the ridge detection plugin ($F(2,6) = 12.17$, $p = 0.008$). Overall, the manual approach calculated an aspect ratio of 59.09 for Mitsui 7, whilst the automated approach calculated a lower aspect ratio of 40.23 ($F(1,5) = 7.99$ $p = 0.047$). The RSD$_R$ for the manual and automated approach to calculated aspect ratio was 18.12 and 22.85% respectively, indicating that for these quite

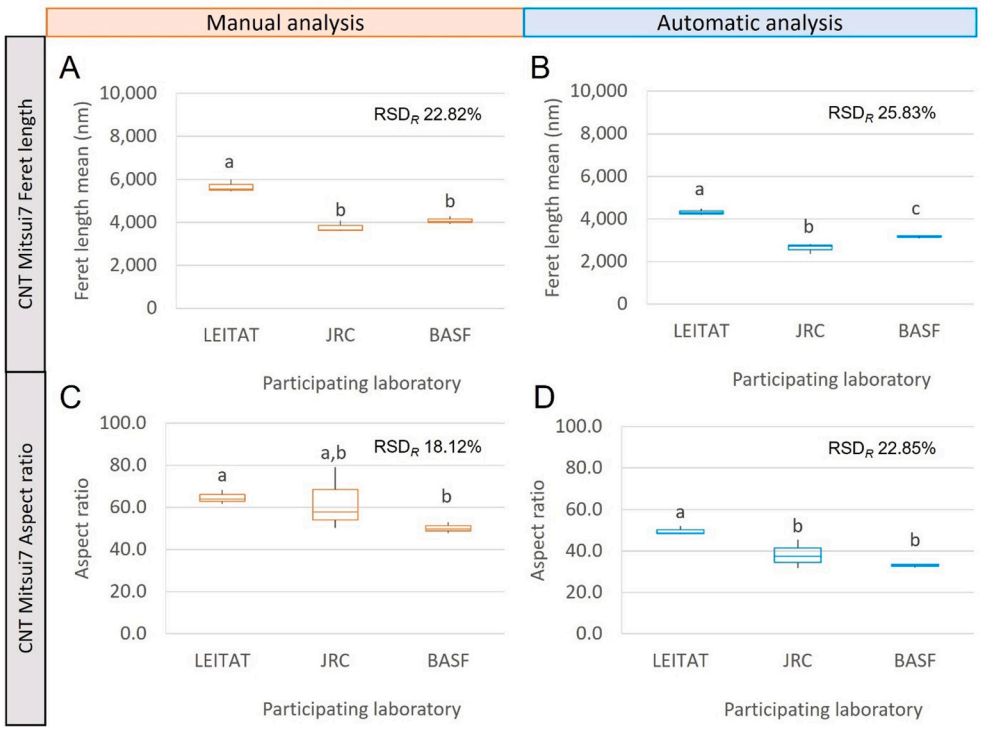


Fig. 9. Summary comparison of the SEM Mitsui 7 length (A and B) or aspect ratio (C and D) calculated for triplicate samples using either a manual (orange plots) or automated ridge detection approach to calculation (blue plots). Box plots represent the median, interquartile ranges and the minimum and maximum for replicate samples measured at each laboratory. Significant differences (ANOVA, $p < 0.05$), where demonstrated, are represented by different letters between laboratories (Tukey's HSD, $p < 0.05$). The RSD$_R$ is presented inset into each figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

heterogeneous materials, differences in aspect ratio should be considered meaningful between nanoforms compared to this material when the difference is >20% of the mean.

4. Conclusions

For all basic NF descriptors, the achievable accuracy can already be estimated for the materials tested using established methods, informed by the summary results for the RSD_R in Table 2. We recommend this approach is taken uniquely each time when performing similarity assessments between new nanoforms to ensure that any difference of a measured physicochemical property is of metrological significance.

ICP-MS is well established and reproducibility assessment of impurities measurements performed close to the expected RSD_R as a function of proportional content of the analyte measured. This function can therefore be used to estimate the achievable accuracy of impurities beyond those tested in this exercise. When considering impurities, specific representative test materials are lacking that are designed with consistent levels of impurities in mind. However, even given this, RSD_R was relatively good, with the highest variation observed for nickel of ~33%. When interpreting differences in impurities between nanoforms, both the relevance of the concentration of the impurity towards toxicity and the possible batch to batch variation should be considered.

TGA was effective at screening for the proportional mass contribution from loss on ignition (a screen for organic surface treatments, impurities and additives) and H_2O to the tested nanoforms, but may need further optimisation to achieve the accuracy required by some existing grouping frameworks.

As an aspect of surface chemistry, zeta potential measurements are found to have high reproducibility but are an extrinsic property of an NF, that describe the surface charge of a material as a function of the surrounding chemistry of the media. The isoelectric point is a scalar descriptor of this extrinsic property that is better suited to similarity assessment between NFs and is demonstrated to be highly reproducible in this investigation (RSD_R of 5.89%).

BET specific surface area was the best performing method for the representative test material evaluated (CeO_2 NM212), with the lowest RSD_R of 3.26%.

For measurement of size and shape using electron microscopy, it was found that the automated NanoDefine ParticleSizer plug-in for ImageJ was not suitable for characterising the high aspect ratio carbon nanotube Mitsui7. Rather, the Ridge Detection plug-in is recommended as an alternative where necessary for this class of material. Both manual and automated approaches to measuring the dimensions of NFs from TEM and SEM images were considered similar in their reproducibility. D50 was the preferred descriptor for diameter due to it having the highest reproducibility and being less sensitive to variations at the extremes of the size distribution, which resulted in statistically significant differences in D10 between laboratories for the same representative test material.

Overall, the methods performed well and would be considered to have the required accuracy for use in existing grouping frameworks. TGA was effective in screening organic and water content of NFs, but may need further standardisation to improve RSD_R to attain the required accuracy for existing grouping frameworks and approaches to similarity assessment for this physicochemical property.

Estimating the achievable accuracy on the basis of the reproducibility standard deviation is an important component of similarity assessment: only NFs that differ by more than the method reproducibility should be considered as different NFs. Thus, for example, our results indicate that differences in D10 up to 34% should not be interpreted for CeO_2-NM212 measured by TEM whist differences in D50 above 10% would be considered meaningful. Only differences larger than the achievable accuracy are metrologically robust. For the GRACIOUS grouping framework, similarity assessment is meaningful when the values of both NFs are in the biologically relevant range, and when the pairwise NF distance is metrologically significant (Hunt, 2021).

It is important to note that the method reproducibility will be dependent on the material and the exact protocol followed. Best practice will require the achievable accuracy to be estimated from a reproducibility assessment of the property measured for the nanoforms you wish to compare. Whilst this is straightforward in cases where you may be measuring the nanoforms directly and so have access to the underlying distributions of the results, this will be more challenging for similarity assessment comparing nanoforms from a database for example, where the raw data or associated uncertainties may not be accessible. In this case, RSD for the method for nanomaterials with the same core constituent composition should be sought from existing literature. Databases containing reproducibility standard deviations for different materials and methods would be a valuable contribution to the scientific community in this respect. The complete tables of relative standard

Table 2
Summary of the relative standard deviation for repeatability (within laboratory variance, RSD_r) and total reproducibility (RSD_R) found across laboratories covering common analytical methods used to measure basic NF descriptors of nanoforms: composition, surface chemistry, surface area, size and shape.

Basic NF descriptors	Analytical method	Descriptors/ effect level endpoints	Nanoforms tested	RSD_r (%)	RSD_R (%)	Maximal x-fold difference
Composition	ICP-MS	Impurities: (atomic percentage)	SiO_2-NM200 and silica_silane (Mg, Cr, Mn, Ni, Sr and Pb)	0.8% - 16.1%	3.4% - 33.8%	1.07–1.78
Surface chemistry	TGA	Loss on ignition (percentage)	SiO_2-NM200, silica_silane, $BaSO_4$-NM220 and-NM212	13.7% - 29.2%	15.4% - 54.2%	1.04–4.05
		H_2O content (percentage)	SiO_2-NM200, silica_silane, $BaSO_4$NM220 and-NM212	2.5% - 66.2%	2.5% - 116.4%[a]	1.01–3.33
	Zetasizer Iso-electric point	Iso-electric point (IEP)	CeO_2-NM212	2.1%	5.9%	1.12
Specific surface area	BET	Specific surface area	CeO_2-NM212	2.8%	3.3%	1.06
Size	TEM	Feret minimum diameter	CeO_2-NM212 (manual and automated approach)	6.2%; 10.9%	8.2%; 10.9%	1.12
		D10, D50 and D90	CeO_2-NM212 (automated approach)	6.5–28.3%	9.5–34.5%	1.17 (D50)
Shape	SEM	Feret minimum diameter	Mitsui 7 (manual and automated approach)	1.8%; 2%	6.6%; 12.8%	1.37
		D10, D50 and D90	Mitsui 7 (automated approach)	2.6–10.3%	11.5–29.5%	1.25 (D50)
		Feret Length	Mitsui 7 (manual and automated approach)	5.5%; 4.9%	22.8%; 25.8%	1.64
		Aspect ratio	Mitsui 7 (manual and automated approach)	15.2%; 10.5%	18.1%; 22.9%	1.51

[a] Note the high maximum RSD_r and RSD_R for H_2O content of CeO_2-NM212 measured by TGA may be due to several factors including strongly reduced dehydrated surfaces after long-term storage under Argon, mixture of Ce valence states in the material, and nonstoichiometric composition, which can result in both mass-gains and losses during transfer to ambient air and heating. RSD_R was more consistent, ranging between 2.5 and 15.27% for the remaining materials

deviations reported in the supplementary information for each material and method can be considered a demonstration of this principle.

CRediT authorship contribution statement

Richard K. Cross: Writing – original draft, Visualization, Formal analysis, Conceptualization. **Nathan Bossa:** Writing – original draft, Investigation, Data curation. **Björn Stolpe:** Methodology, Writing – original draft, Formal analysis, Supervision. **Frédéric Loosli:** Writing – original draft, Investigation, Formal analysis. **Nicklas Mønster Sahlgren:** Investigation. **Per Axel Clausen:** Investigation, Writing – original draft. **Camilla Delpivo:** Investigation. **Michael Persson:** Investigation. **Andrea Valsesia:** Investigation. **Jessica Ponti:** Investigation. **Dora Mehn:** Investigation. **Didem Ag Seleci:** Supervision. **Philipp Müller:** Supervision, Data curation. **Frank von der Kammer:** Supervision. **Hubert Rauscher:** Supervision, Writing – review & editing. **Dave Spurgeon:** Writing – review & editing, Formal analysis. **Claus Svendsen:** Writing – review & editing, Formal analysis. **Wendel Wohlleben:** Conceptualization, Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The GRACIOUS project has received funding from the European Union's Horizon 2020 research and 1063 innovation program under grant agreement No. 760840.

We would like to also acknowledge the contribution of several individuals to the execution of this study: Marianne Matzke from UKCEH for support in management and conception; Darren Sleep from UKCEH for advice on reproducibility assessments; Jose Daniel Granero Löfgren and Kristoffer Amiridis from Nouryon for BET analysis: and from BASF Lutz Höring (TGA measurements), Thorsten Wieczorek (TEM, SEM measurements) and Stefan Herrmann (zeta-potential measurements).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.impact.2022.100410.

References

Babick, F., Mielke, J., Wohlleben, W., Weigel, S., Hodoroaba, V.D., 2016. How reliably can a material be classified as a nanomaterial? Available particle-sizing techniques at work. J. Nanopart. Res. Springer Netherlands. https://doi.org/10.1007/s11051-016-3461-7.

ECHA, 2017. Appendix R. 6–1 for nanomaterials applicable to the Guidance on QSARs and Grouping of Chemicals. https://doi.org/10.2823/884050.

ECHA, 2019. Appendix for Nanoforms Applicable to the Guidance on Registration and Substance Identification. https://doi.org/10.2823/832485.

European Commission, 2018. Commission Regulation (EU) 2018/1881 of 3 December 2018 amending Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards Annexes I, III,VI, V. Off. J. Eur. Union 2016, pp. 48–119.

European Parliament and Council, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/4. OJ L 396, pp. 1–849.

Hackley, V.A., Stefaniak, A.B., 2013. "Real-world" precision, bias, and between-laboratory variation for surface area measurement of a titanium dioxide nanomaterial in powder form. J. Nanopart. Res. 15 https://doi.org/10.1007/s11051-013-1742-y.

Heyden, Y., Smeyers-Verbeke, J., 2007. Set-up and evaluation of interlaboratory studies. J. Chromatogr. A 1158, 158–167. https://doi.org/10.1016/j.chroma.2007.02.053.

Holzwarth, U., Ponti, J., 2020. 44Ti diffusion labelling of commercially available, engineered TiO2 and SiO2 nanoparticles. J. Nanopart. Res. 22. https://doi.org/10.1007/s11051-020-04978-5.

Horwitz, W., Albert, R., 1995. Precision in analytical measurements: expected values and consequences in geochemical analyses. Fresenius J. Anal. Chem. 351, 507–513. https://doi.org/10.1007/BF00322724.

Horwitz, W., Kamps, L.R., Boyer, K.W., 1980. Quality assurance in the analytical analysis of foods for trace constituents. J. Assoc. Off. Anal. Chem. 63.

Hunt, N., 2021. Guidance in a Nutshell - GRACIOUS Framework for Grouping and Read-across of Nanomaterials and Nanoforms [WWW Document]. https://doi.org/10.5281/zenodo.5534105.

Janer, G., Ag-Seleci, D., Sergent, J.A., Landsiedel, R., Wohlleben, W., 2021a. Creating sets of similar nanoforms with the ECETOC NanoApp: real-life case studies. Nanotoxicology. https://doi.org/10.1080/17435390.2021.1946186.

Janer, G., Landsiedel, R., Wohlleben, W., 2021b. Rationale and decision rules behind the ECETOC NanoApp to support registration of sets of similar nanoforms within REACH. Nanotoxicology 15, 145–166. https://doi.org/10.1080/17435390.2020.1842933.

Jeliazkova, N., Bleeker, E., Cross, R., Haase, A., Janer, G., Peijnenburg, W., Pink, M., Rauscher, H., Svendsen, C., Tsiliki, G., Zabeo, A., Hristozov, D., Stone, V., Wohlleben, W., 2022. How can we justify grouping of nanoforms for hazard assessment? Concepts and tools to quantify similarity. NanoImpact 25, 100366. https://doi.org/10.1016/j.impact.2021.100366.

Keller, J.G., Quevedo, D.F., Faccani, L., Costa, A.L., Landsiedel, R., Werle, K., Wohlleben, W., 2021. Dosimetry in vitro – exploring the sensitivity of deposited dose predictions vs. affinity, polydispersity, freeze-thawing, and analytical methods. Nanotoxicology 15, 21–34. https://doi.org/10.1080/17435390.2020.1836281.

Lamberty, A., Franks, K., Braun, A., Kestens, V., Roebben, G., Linsinger, T.P.J., 2011. Interlaboratory comparison for the measurement of particle size and zeta potential of silica nanoparticles in an aqueous suspension. J. Nanopart. Res. https://doi.org/10.1007/s11051-011-0624-4.

Loosli, F., 2020. Iso Electric Point (IEP), v1.0. Zenodo. https://doi.org/10.5281/ZENODO.5519851.

Loosli, F., Cross, R.K., Bossa, N., Rasmussen, K., Rauscher, H., Peijnenburg, W., Arts, J., Clausen, P.A., Wohlleben, W., Ruggiero, E., von der Kammer, F., 2021. Refinement of the selection of physicochemical properties for grouping and read-across of nanoforms. NanoImpact. https://doi.org/10.1016/j.impact.2021.100375.

Magnusson, B., Örnemark, U., 2014. Eurachem guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics (2nd ed. 2014), Eurachem Guide.

Mech, A., Rauscher, H., Babick, F., Hodoroaba, V.-D., Wohlleben, W., Marvin, H., Weigel, S., Brüngel, R., Friedrich, C., 2019. The NanoDefine methods manual. In: Part 3: Standard Operating Procedures (SOPs). Publications Office of the European Union. https://doi.org/10.2760/02910.

Mülhopt, S., Diabaté, S., Dilger, M., Adelhelm, C., Anderlohr, C., Bergfeldt, T., de la Torre, J.G., Jiang, Y., Valsami-Jones, E., Langevin, D., Lynch, I., Mahon, E., Nelissen, I., Piella, J., Puntes, V., Ray, S., Schneider, R., Wilkins, T., Weiss, C., Paur, H.R., 2018. Characterization of nanoparticle batch-to-batch variability. Nanomaterials 8. https://doi.org/10.3390/nano8050311.

NANoREG, 2018. NANoREG D2.04 FS Protocol for Quantitative Analysis of Inorganic and Organic MNM Surface Coatings [WWW Document]. URL. https://www.rivm.nl/en/documenten/nanoreg-d204-fs-protocol-for-quantitative-analysis-of-inorganic-and-organic-mnm-surface (accessed 1.25.22).

OECD, 2015. Dossier on Silicon Dioxide-Annex, p. 1.

Park, M., Catalán, J., Ferraz, N., Cabellos, J., Vanhauten, R., Vázquez-Campos, S., Janer, G., 2018. Development of a systematic method to assess similarity between nanomaterials for human hazard evaluation purposes – lessons learnt. Nanotoxicology 0, 1–25. https://doi.org/10.1080/17435390.2018.1465142.

Rasmussen, K., Mech, A., Mast, J., De Temmerman, P., Waegeneers, N., Van Steen, F., Pizzolon, J.C., De Temmerman, L., Van Doren, E., Jensen, A., Birkedal, R., Levin, M., Nielsen, H., Koponen, I.K., Axel, P., Kembouche, Y., Thieriet, N., Spalla, O., Giuot, C., Rousset, D., Bau, S., Bianchi, B., Gilliland, D., Pianella, F., Ceccone, G., Cotogno, G., Gibson, N., Stamm, H., 2013. Synthetic Amorphous Silicon Dioxide (NM-200, NM-201, NM-202, NM-203, NM-204): Characterisation and Physico-Chemical Properties, JRC Repository: NM-series of Representative Manufactured Nanomaterials, EUR - Scientific and Technical Research Reports. https://doi.org/10.2788/57989.

Sahlgren, N.B., Clausen, P.A., Jensen, K.A., 2019. TGA-MS-Screening, v1.0. Zenodo. https://doi.org/10.5281/ZENODO.5519861.

Schmidt, A., Bresch, H., 2021. Development of a Specific OECD Test Guideline on Particle Size and Particle Size Distribution of Nanomaterials, Environmental Research of the Federal Ministry for the Environment, Nature Conservation. Nuclear Safety and Consumer Protection, Berlin.

Singh, C., Friedrichs, S., Ceccone, G., Gibson, N., Jensen, K.A., Levin, M., Goenaga Infante, H., Carlander, D., Rasmussen, K., 2014. Cerium Dioxide, NM-211, NM-212, NM-213: Characterisation and Test Item Preparation. https://doi.org/10.2788/80203.

Steger, C., 1998. An unbiased detector of curvilinear structures. IEEE Trans. Pattern Anal. Mach. Intell. 20. https://doi.org/10.1109/34.659930.

Stone, V., Gottardo, S., Bleeker, E.A.J., Braakhuis, H., Dekkers, S., Fernandes, T., Haase, A., Hunt, N., Hristozov, D., Jantunen, P., Jeliazkova, N., Johnston, H., Lamon, L., Murphy, F., Rasmussen, K., Rauscher, H., Jiménez, A.S., Svendsen, C., Spurgeon, D., Vázquez-Campos, S., Wohlleben, W., Oomen, A.G., 2020. A framework

for grouping and read-across of nanomaterials- supporting innovation and risk assessment. Nano Today 35, 1–15. https://doi.org/10.1016/j.nantod.2020.100941.

Thompson, M., 2000. Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing. Analyst 125, 385–386. https://doi.org/10.1039/b000282h.

Verleysen, E., Wagner, T., Lipinski, H.G., Kägi, R., Koeber, R., Boix-Sanfeliu, A., De Temmerman, P.J., Mast, J., 2019. Evaluation of a TEM based approach for size measurement of particulate (nano)materials. Materials (Basel). 12, 1–20. https://doi.org/10.3390/ma12142274.

Wohlleben, W., Hellack, B., Nickel, C., Herrchen, M., Hund-Rinke, K., Kettler, K., Riebeling, C., Haase, A., Funk, B., Kühnel, D., Göhler, D., Stintz, M., Schumacher, C., Wiemann, M., Keller, J., Landsiedel, R., Broßell, D., Pitzko, S., Kuhlbusch, T.A.J., 2019. The nanoGRAVUR framework to group (nano)materials for their occupational, consumer, environmental risks based on a harmonized set of material properties, applied to 34 case studies. Nanoscale 11, 17637–17654. https://doi.org/10.1039/c9nr03306h.