JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

Received 23 June 2023
Accepted 21 September 2023

Edited by T. J. Sato, Tohoku University, Japan

Keywords: round robins; data analysis; small-angle scattering; nanomaterials; interlaboratory comparability; nanostructure quantification

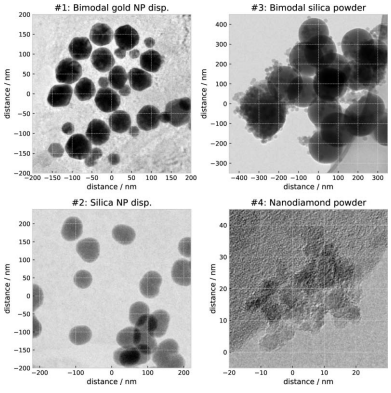
Supporting information: this article has supporting information at journals.iucr.org/j

The human factor: results of a small-angle scattering data analysis round robin

Brian R. Pauw,[a]* Glen J. Smales,[a] Andy S. Anker,[b] Venkatasamy Annadurai,[c] Daniel M. Balazs,[d] Ralf Bienert,[e] Wim G. Bouwman,[f] Ingo Breßler,[a] Joachim Breternitz,[g] Erik S. Brok,[h] Gary Bryant,[i] Andrew J. Clulow,[j] Erin R. Crater,[k] Frédéric De Geuser,[l] Alessandra Del Giudice,[m] Jérôme Deumer,[n] Sabrina Disch,[o,p] Shankar Dutt,[q] Kilian Frank,[r] Emiliano Fratini,[s] Paulo R. A. F. Garcia,[t] Elliot P. Gilbert,[u] Marc B. Hahn,[a] James Hallett,[v] Max Hohenschutz,[w] Martin Hollamby,[x] Steven Huband,[y] Jan Ilavsky,[z] Johanna K. Jochum,[aa] Mikkel Juelsholt,[bb] Bradley W. Mansel,[cc] Paavo Penttilä,[dd] Rebecca K. Pittkowski,[b] Giuseppe Portale,[ee] Lilo D. Pozzo,[ff] Leonhard Rochels,[o,p] Julian M. Rosalie,[a] Patrick E. J. Saloga,[gg] Susanne Seibt,[j] Andrew J. Smith,[hh] Gregory N. Smith,[ii] Glenn A. Spiering,[jj] Tomasz M. Stawski,[ff] Olivier Taché,[kk] Andreas F. Thünemann,[a] Kristof Toth,[ll] Andrew E. Whitten[u] and Joachim Wuttke[mm]

[a]BAM Federal Institute for Materials Research and Testing, 12205 Berlin, Germany, [b]Department of Chemistry, University of Copenhagen, 2100 Copenhagen, Denmark, [c]University of Mysore, NIE First Grade College, Mysore – 570008, India, [d]Institute of Science and Technology Austria (IST Austria), Am Campus 1, Klosterneuburg 3400, Austria, [e]BAM Federal Institute for Materials Research and Testing, Richard-Willstätter-Straße 11, 12489 Berlin, Germany, [f]Delft University of Technology, The Netherlands, [g]Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Germany, [h]FORCE Technology, Park Alle 345, 2605 Brøndby, Denmark, [i]Physics, School of Science, RMIT University, Melbourne, Australia, [j]Australian Synchrotron, ANSTO, 800 Blackburn Road, Clayton, VIC 3168, Australia, [k]Virginia Polytechnic Institute and State University, Blacksburg, VA 24060, USA, [l]Université Grenoble Alpes, CNRS, Grenoble INP, SIMAP, F-38000 Grenoble, France, [m]Department of Chemistry, Sapienza University of Rome, P.le A. Moro 5, I-00185 Rome, Italy, [n]Physikalisch-Technische Bundesanstalt, Abbestraße 2, 10587 Berlin, Germany, [o]Department für Chemie, Universität zu Köln, Greinstraße 4–6, 50939 Köln, Germany, [p]Institute for Inorganic Chemistry and Center for Nanointegration Duisburg-Essen (CENIDE), University of Duisburg-Essen, Universitaetsstraße 5–7, 45117 Essen, Germany, [q]Department of Materials Physics, Australian National University, ACT, Australia, [r]Faculty of Physics and CeNS, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany, [s]Department of Chemistry 'Ugo Schiff' and CSGI, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy, [t]Brazilian Synchrotron Light Laboratory (LNLS) – Giuseppe Máximo Scolfaro 10000, 13083-100 Campinas, São Paulo, Brazil, [u]Australian Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, New Illawarra Road, Lucas Heights, NSW 2234, Australia, [v]Department of Chemistry, School of Chemistry, Food and Pharmacy, University of Reading, Whiteknights, PO Box 224, Reading RG6 6AD, United Kingdom, [w]Institute of Physical Chemistry, RWTH Aachen University, Landoltweg 2, D-52056 Aachen, Germany, [x]School of Chemical and Physical Sciences, Keele University, Staffordshire ST5 5BG, United Kingdom, [y]Department of Physics, University of Warwick, Warwick CV4 7AL, United Kingdom, [z]APS, ANL, 9700S Cass Avenue, Lemont, IL, USA, [aa]Heinz Maier-Leibnitz Zentrum, Technische Universität München, D-85748 Garching, Germany, [bb]Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom, [cc]National Synchrotron Radiation Research Center, 101 Hsin-Ann Road, Hsinchu Science Park, Hsinchu 30076, Taiwan, People's Republic of China, [dd]Department of Bioproducts and Biosystems, Aalto University, PO Box 16300, FI-00076 Aalto, Finland, [ee]University of Groningen, The Netherlands, [ff]University of Washington, Chemical Engineering, Box 351750, Seattle, WA 98195-1750, USA, [gg]Independent Researcher, Germany, [hh]Diamond Light Source Ltd, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, United Kingdom, [ii]ISIS Neutron and Muon Source, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, United Kingdom, [jj]Department of Chemistry, Macromolecules Innovation Institute (MII), Virginia Tech, Blacksburg, VA 24061, USA, [kk]Université Paris–Saclay, CEA, CNRS, NIMBE, 91191 Gif-sur-Yvette, France, [ll]Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA, and [mm]Forschungszentrum Jülich GmbH, JCNS-MLZ, Lichtenbergstraße 1, 85747 Garching, Germany. *Correspondence e-mail: brian.pauw@bam.de

A round-robin study has been carried out to estimate the impact of the human element in small-angle scattering data analysis. Four corrected datasets were provided to participants ready for analysis. All datasets were measured on samples containing spherical scatterers, with two datasets in dilute dispersions and two from powders. Most of the 46 participants correctly identified the number of populations in the dilute dispersions, with half of the population mean entries within 1.5% and half of the population width entries within 40%. Due to the added complexity of the structure factor, far fewer people submitted answers on the powder datasets. For those that did, half of the entries for the means and widths were within 44 and 86%, respectively. This round-robin experiment highlights several causes for the discrepancies, for which solutions are proposed.



1. Introduction

The scientific method has been historically developed to eliminate human and instrumental bias from understanding of the natural world. It has been applied to a wide range of fields with various levels of success. This success may be gauged using tools such as round-robin (RR) experiments, where, for example, identical objects are circulated to various laboratories, enabling the quantification of the spread in findings. Ideally, the observations and resulting conclusions are independent of the observer and instrumentation, provided the means and methodology meet a minimum standard. Such minimum standards can be defined by national and international standards bodies such as DIN and ISO. If the observations and conclusions of such an experiment are indeed comparable, we can be confident that the results are free of bias and likely to be an accurate representation of the object or phenomenon under investigation.

Several notable RR experiments in nanomaterial analysis compare the results of different techniques (insofar that different techniques are able to provide truly comparable end parameters). To more closely capture the true human or instrumental variability, however, experiments focusing on a single technique or even a single aspect of a technique are perhaps better suited. Such focus allows us to pinpoint the larger contributors to interlaboratory variability, with the eventual goal of eliminating or minimizing such dependencies. Notable examples of such studies have been carried out in fields such as atom probe tomography (Dong *et al.*, 2019); X-ray diffraction (Madsen *et al.*, 2001; Scarlett *et al.*, 2002); neutron scattering (Rennie *et al.*, 2013); neutron powder diffraction (Whitfield, 2016); biology-specific small-angle X-ray scattering (Bio-SAXS) (Trewhella *et al.*, 2022); and surface-area determination following the Brunauer, Emmett and Teller method (Osterrieth *et al.*, 2022).

In this vein, a large RR experiment was carried out several years ago focusing on the collection of small-angle scattering (SAS) data for nanoparticle (NP) liquid suspensions. SAS is a technique for (traceable) quantification of nanostructures in bulk amounts of sample. After appropriate data correction, information might be retrieved on the scatterer morphology (form factor), its size distribution or its packing (structure factor). In most cases, one of these three may be elucidated upon the provision of information or via assumptions on the other two. In rare cases, two or more of these pieces of information can be convincingly extracted from the data. (While the data contain information proportional to the mass or volume of the scatterers, for narrow distributions this can be converted to number-weighted distributions while maintaining low uncertainties.)

In the aforementioned RR, SAXS datasets from a wide range of laboratories were collected and subsequently analysed using several programs with consistent starting parameters (Pauw *et al.*, 2017*a*). From this experiment, it was clear – at least for non-challenging samples – that most laboratories and instruments were able to collect consistent data resulting in standard deviations on the order of a percent for the mean particle size, and ∼10% for the population size distribution width, regardless of software choices. The next logical step, then, is to find out what influence the individual researchers would have on the analysis of the data (aka the 'human factor').

The influence of researchers on the results can be investigated by circulating a dataset to be interpreted and quantifying the variation on the resulting morphological parameters (Osterrieth *et al.*, 2022; Madsen *et al.*, 2001; Scarlett *et al.*, 2002). In particular, in SAS, the data analysis can be a stumbling block, so the expectation is to see a large spread in the results for this data-analysis round robin (DARR) for SAS.

Given the wide range of possible samples and analyses in this field, the challenge was to find representative datasets that would

(*a*) cover a range of relevant materials and common challenges for datasets,

(*b*) be of high quality to minimize result variation through data uncertainties,

(*c*) be from well characterized and well understood samples,

(*d*) be accompanied by the same nominal level of supplementary information as is normally provided by materials researchers.

Eventually, four experimental datasets were chosen that included common challenges. These were made available online, and repeatedly advertised on professional mailing lists, websites and through personal communication channels. After allowing considerable time and a few deadline extensions, 46 entries were received. The results inferred from the 46 entries provide a good insight into the challenges facing SAS as a materials science tool.

In this article, RR design, methods and results are presented. On the basis of the results, the following discussion highlights the thus-identified areas of improvement of the field, and provides possible avenues for doing so. Additionally, a section is spent on discussing possible improvements in future RR experiments. Lastly, while the results presented herein are necessarily limited, the anonymized results are available on Zenodo (https://zenodo.org/records/7509710), as is the *Jupyter Notebook* used for the correction, interpretation and visualization, in the hope that alternative or extended interpretations of the results may be developed.

2. Methods

2.1. The four round-robin datasets

Four datasets of 1D scattering data, representative of two dilute and two dense NP systems, were made available to willing participants in the form of tabulated three-column `.dat` files, containing Q, $I(Q)$ and $\sigma_I(Q)$. Here, Q is the scattering vector magnitude in units of nm^{-1}, $I(Q)$ is the scattering intensity in units of $(m\,sr)^{-1}$ and $\sigma_I(Q)$ is the absolute uncertainty of the intensity (one standard deviation). In addition, participants were provided with a letter describing the datasets and the task ahead, as well as an *Excel* sheet for tracking results in a standardized form (see the Zenodo repository https://zenodo.org/records/7506365).

The four datasets are shown in Fig. 1, with model fits that serve only as a suggestion of an appropriate model for the datasets. Datasets 1 and 2 were fitted in *McSAS3* (Pauw & Bressler, 2022) utilizing a spherical form factor, whilst datasets 3 and 4 were fitted in *SASfit* (Kohlbrecher & Breßler, 2022) using spherical form factors (with log-normal distributions) with appropriate structure factors (sticky hard sphere and mass fractals for datasets 3 and 4, respectively), alongside background contributions and peak functions to better describe the wide-angle data. These fits can also be found in the Zenodo repository. A subset of fitting parameters (means and widths for each population of each dataset) is shown in Table 1.

Datasets 1 and 2 originate from publicly available measurements (Deumer & Gollwitzer, 2022) performed at the SAXS beamline of the Physikalisch-Technische Bundesanstalt (PTB) at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung II mbH (BESSY II) (Krumrey & Ulm, 2001; Wernecke *et al.*, 2014), on reference samples synthesized for the European Metrology Programme for Innovation and Research (EMPIR) 'nPSize' project. These reference samples are detailed by Bartczak & Hodoroaba (2022).

Datasets 3 and 4 were measured as powders using the MOUSE project (methodology optimization for ultrafine structure exploration) (Smales & Pauw, 2021). X-rays were generated using a microfocus X-ray tube with a copper anode, and multilayer optics were employed to parallelize and monochromatize the X-ray beam to an approximate wavelength of Cu $K\alpha$ (λ = 0.154 nm). Samples were mounted as small amounts of powder in a thin laser-cut holder and held in place between two pieces of low-scattering Scotch Magic Tape. Scattered radiation was detected on an in-vacuum EIGER 1M detector (Dectris, Switzerland), which was placed at multiple distances between 55 and 2507 mm from the sample. The resulting data have been processed and scaled to absolute intensity using the *DAWN* (data-analysis workbench; Basham *et al.*, 2015) software package in a standardized complete 2D correction pipeline with uncertainty propagation (Smales & Pauw, 2021; Pauw *et al.*, 2017*b*).

Table 1
Means and widths determined from example fits to the data using *McSAS* and *SASfit* for datasets 1–4, and populations 1 and 2.

The subscript 'v' is for volume-weighted values and 'n' is for number-weighted values. SD = standard deviation. Uncertainties are in brackets where available.

			Results (nm)			
Dataset	Population	Software	μ_v	SD_v	μ_n	SD_n
1	P1	*McSAS*	30.8 (4)	3.8 (4)	–	–
		SASfit	31.2	5.0	29.1	4.7
	P2	*McSAS*	59.0 (2)	6.2 (4)	–	–
		SASfit	58.8	5.3	57.4	5.2
2	P1	*McSAS*	52.2 (6)	4.4 (1)	–	–
		SASfit	50.7	1.1	50.6	1.1
	P2	*SASfit*	56.1	8.5	52.4	7.9
3	P1	*SASfit*	28.5	3.0	27.6	2.9
	P2	*SASfit*	214.8	27.4	204.5	26.1
4	P1	*SASfit*	4.26	2.20	1.92	0.99

2.1.1. Dataset 1: bimodal gold nanoparticles. Dataset 1 is from a material designated as nPSize 1. This sample was designed to contain two populations with known concentrations of spherical gold NPs in water, with diameters of 30 and 60 nm at a 1:1 number ratio. Given this number ratio, the volume-fraction ratio is ∼1:8. In other words, the smaller population (population 1) contributes only 1/9th of the total volume fraction of scatterers which, in this case, renders its

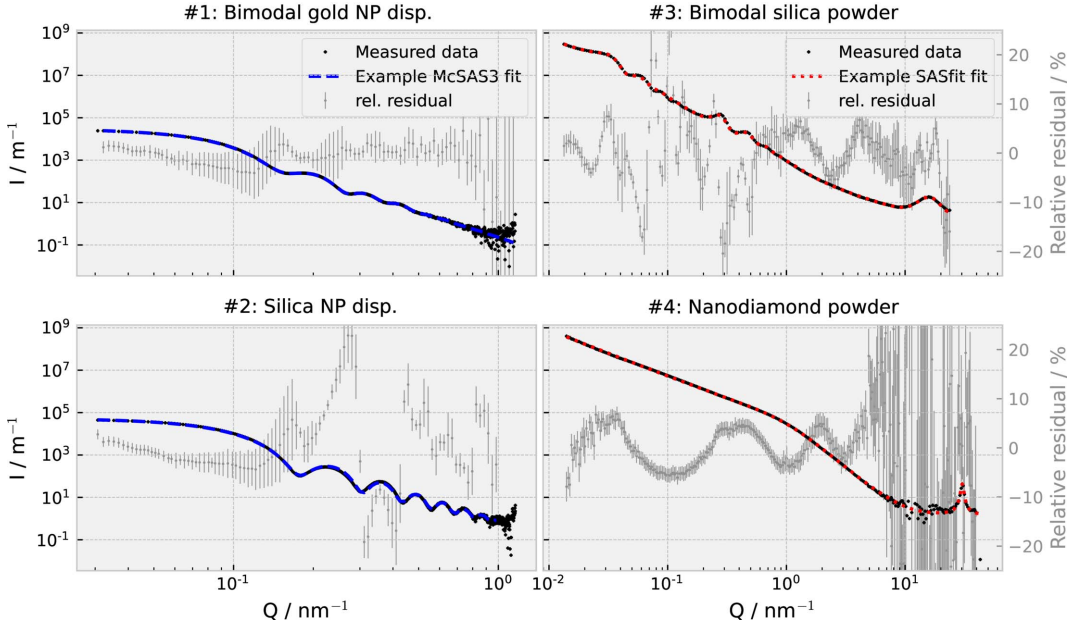


Figure 1
The four datasets (black) chosen for this RR experiment, with example fits and relative residuals (light grey, relating to the secondary right-hand side axis). The example fits have been generated using *McSAS3* (left, blue) and *SASfit* (right, red). The size distributions resulting from these fits are shown in Fig. 2.

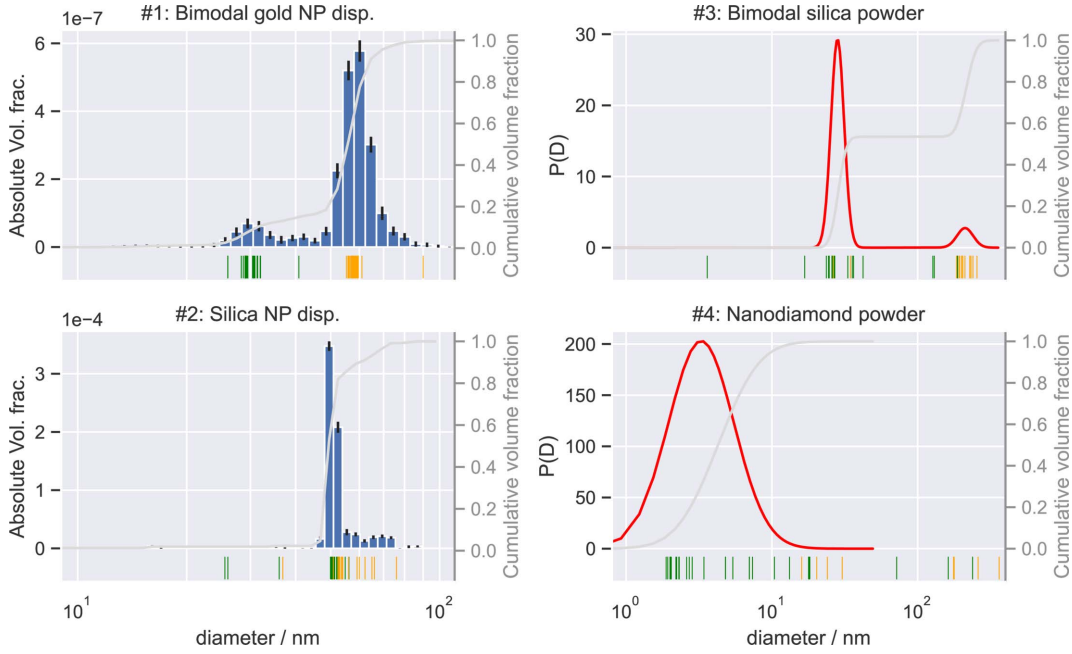


Figure 2
Example volume-weighted distributions for datasets 1 through 4, using *McSAS3* (left) and *SASfit* (right). The *McSAS* histogram bars show the integrated volume fraction for the span of each bar, while the *SASfit* results show the volume-weighted probability function. For both, the cumulative distribution function is plotted in grey on the secondary axis. The marks below indicate the participant-submitted population means for population 1 (green) and population 2 (orange).

signal nearly invisible to the eye in the presence of the larger population (population 2). In practical measurements, the modality of the populations is often not known. The existence of the second population has therefore not been explicitly revealed to the participants but merely hinted at through the design of the answer form.

One possible solution for this scattering pattern is provided in Fig. 2 using *McSAS3*. This example solution shows the presence of two populations. When these populations are analysed in the diameter ranges of $20 \leq D$ (nm) ≤ 40 and $40 \leq D$ (nm) ≤ 80, they provide the volume-weighted means and widths (standard deviations) for populations 1 and 2, as shown in Table 1. For comparison, values from example fits using *SASfit* have also been added. We cannot claim these solutions to be 'correct' but they serve as an example.

2.1.2. Dataset 2: silica nanoparticles. The second dataset is nPSize 10 from the same series, where the scatterers consist of a narrow distribution of nominally monomodal silica with a nominal diameter of 60 nm (Bartczak & Hodoroaba, 2022). Recent discussions revealed that this sample may also contain a minor fraction of a slightly larger population (*cf.* Table 1). *McSAS3* example fits (as well as *SASfit*, not shown here) do indicate the presence of a small fraction of a broad distribution of particles, and a significant number of participants found the same.

2.1.3. Dataset 3: mixture of AS-40 and 250 nm silica powders. Dataset 3 was produced in-house by mixing together two spherical silica materials in a 1:1 mass ratio. Smaller silica spheres were obtained by freeze-drying Ludox AS-40 (Sigma–Aldrich, *ca* 22 nm in diameter), whilst the larger spheres were synthesized using the Stöber process, where tetraethyl orthosilicate (TEOS, Sigma–Aldrich, 98%) was added to a solution of ethanol (Sigma–Aldrich, 96%), water and ammonium hydroxide solution (ACS reagent, *ca* 28%) and left to stir at room temperature for 24 h. The resulting suspension was then centrifuged and washed with ethanol before being dried at 60°C overnight. A *SASfit* example distribution is show in Fig. 2, with means and widths detailed in Table 1.

2.1.4. Dataset 4: nanodiamond powder. Dataset 4 was measured from a commercial sample of nanodiamonds obtained from PlasmaChem GmbH in Berlin, catalogue number PL-D-G02. These are globular diamond particles with a nominal diameter between 4 and 6 nm, supplied as a dry powder. Some technical details and additional references demonstrating their use are available for this material on the PlasmaChem website (https://shop.plasmachem.com/nanodiamonds-and-carbon/32-112-nanodiamonds-extra-pure-grade-g02.html). A *SASfit* example distribution is shown in Fig. 2, with means and widths in Table 1.

2.2. Electron micrographs

Electron micrographs showing the scatterers underpinning the four datasets are shown in Fig. 3. While these were not provided with the original challenge, they here serve to show the imposed assumption of roughly spherical scatterers. The images from datasets 1 and 2 were measured at the Commissariat à l'énergie atomique et aux énergies alternatives (CEA) and deposited in a Zenodo repository (Pollen Metrology, 2021).

Images for the samples of datasets 3 and 4 were recorded using an electron microscope available on site. For these two, the experimental details are as follows. Transmission electron

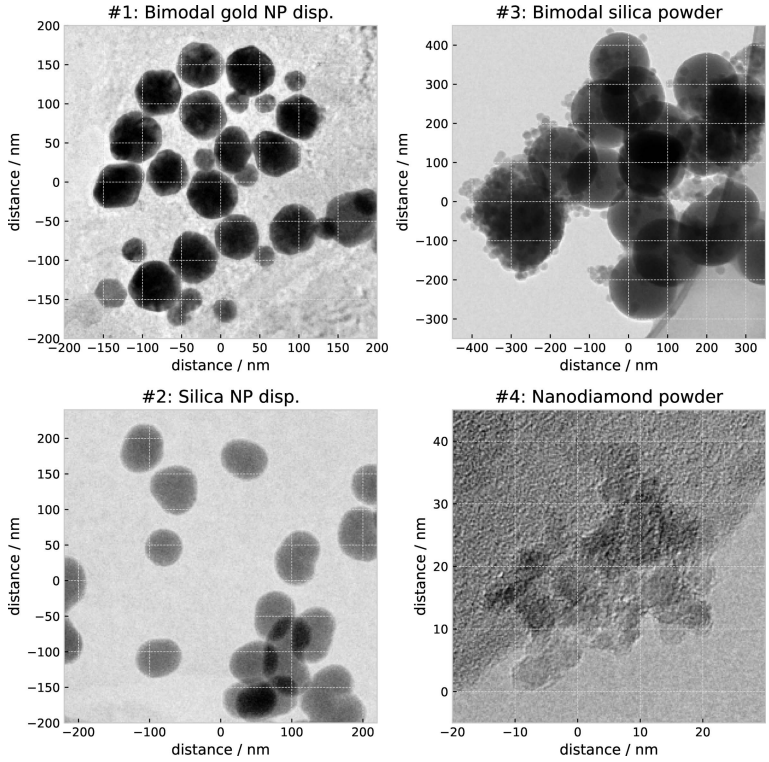


Figure 3
Electron micrographs of the scatterers that make up the four datasets. The images from datasets 1 and 2 originate from the nPSize project repository (Pollen Metrology, 2021). Images for the samples of datasets 3 and 4 were recorded on site.

microscopy (TEM) samples were prepared by dispersing the powders via ultrasonication for a minimum of 5 min in ethanol. One to three droplets of the resulting suspensions were placed on Cu TEM grids coated with lacey- or holey-carbon films. TEM observations were conducted on a JEOL 2200FS instrument, operating at 200 kV. Images were acquired in bright-field TEM mode, using high-contrast apertures to enhance the visibility of the particles, except for the ∼10 nm nanodiamonds of dataset 4, where high-resolution TEM images were obtained instead.

2.3. Answer-sheet design, data read-in and corrections

The answer-sheet design imposed practical limitations on the answer space: a machine-readable answer sheet needed to be developed that would present the resulting morphological parameters for variation analysis without accidental telegraphing of a desired result or answer space. Several entry fields of the answer sheet were deliberately left vague, in an attempt to eke out further information on what the small-angle scatterer might understand for common but confusing terms (one example of this is the 'mean size' of the spherical scatterers not specifying whether diameter or radius was meant). The final design of the answer-sheet template can be found in the repositories.

The submitted answer sheets had to undergo several processing steps before they could be compared. Author information and reported analysis results were read in separately to aid anonymization. The following procedures were applied to the evaluation data in this order:

(1) Manual corrections were applied to some sheets to ensure the entries were in the right column for reading, to remove extraneous information, to change decimal commas to periods, to fill in missing information (after communication with the author) *etc*.

(2) The software package names were sorted and shortened to their minimal identifying names. For software packages that were only used once or twice, they were categorized under 'Other', in order to not compromise anonymization.

(3) The weighting category was forced into either 'volume', 'number' or 'not defined'.

(4) At this point, a check and correction for common pitfalls was performed. This included differences in interpretation of 'size' (*i.e.* radius or diameter), and missing unit conversions in read-in or reporting. Details are provided in Section 3.1.

(5) Due to the limited number of entries, no outlier test was applied to further exclude submitted values.

(6) Lastly, the entries of the set of concatenated data were randomized, so that they were no longer in the sequence in which they were ingested.

3. Results

3.1. Overall statistics

In total, 46 answer sheets were received. During the read-in of these answer sheets, common pitfalls were compensated for with the following frequencies:

(i) 72% interpreted 'size' as 'radius' (set correction factor to 2).

(ii) 4% missed the dataset unit information (additional correction factor of 10).

(iii) 4% mis-corrected the dataset units or reported information in ångström (additional correction factor of 0.1).

(iv) 40% used the *SasView* software package but reported the size distribution width in *SasView*'s 'polydispersity' units rather than in a population width in standard deviation (set correction factor to the mean radius). Other software packages might also report the distribution width in other ways, but this is not known and thus not corrected for.

Not all four datasets were fitted by all participants, and not all participants identified the same number of populations in the datasets. While it was telegraphed through the answer sheet that there might be more than one population present, it was not indicated for which sample(s) that would be the case. Fig. 4 shows how many participants had entered values for a given population for each dataset and what software they used to detect this population.

Most participants recognized a bimodal population in datasets 1 and 3, about half added a second population to dataset 2, and almost none saw a second population in dataset 4. Datasets 3 and 4 were more of a challenge than 1 and 2, with only about half of the participants entering results for these.

The software packages used for these analyses consisted of four main packages, in alphabetical order: *Irena* (Ilavsky & Jemian, 2009), *McSAS* (Bressler *et al.*, 2015) and/or *McSAS3* (Pauw & Bressler, 2022), *SASfit* (Kohlbrecher & Breßler, 2022), and *SasView* (https://www.sasview.org). Some packages such as *Irena* offer multiple methods for optimization. Other users used more uncommon software, among which were one or two uses of either:

(*a*) *autoSAXS* (in-house software, PTB),
(*b*) *GNOM* via *BioXTAS RAW* (Hopkins *et al.*, 2017),
(*c*) *pySAXS* (Taché *et al.*, 2017),
(*d*) SAXS numerical inversion (Bender *et al.*, 2017),
(*e*) *XSACT* (Xenocs, 2022),
as well as several in-house developed programs. When software names are abbreviations, they have not been spelled out here, for reasons of legibility, and because they are known by their names, not their spelled-out (b)ac(k)ronyms.

Given the range of challenges posed by the datasets, the analysis likewise could benefit from leveraging different approaches implemented in the software packages. Fig. 4 displays the software packages used for each dataset, showing a fairly even split between software packages. This indicates a healthy ecosystem, on the one hand, but also complicates comparison as each software package may report parameters, such as distribution widths, volume fractions and goodness of fit, in their own unique way.

3.2. Dataset 1

Dataset 1 is, perhaps, the most straightforward, and thus serves well as a starting point for discussion of the actual results. Fig. 5 attempts to show as much relevant information as possible in a single figure. As the size distribution is reasonably narrow, the volume-weighted mean and the number-weighted mean are sufficiently proximate to be shown on the same plot.

The visualization shows a breadth of entries that spans ∼±7% for a 95% confidence interval on the reported mean scatterer dimension, and a remarkable 50% of the entries fall within 1.5% of the median mean. The reported widths deviate much more, with 50% of the entries within 44% of the median width. This is very likely due to the inconsistent reporting of distribution widths by the various software packages. Despite

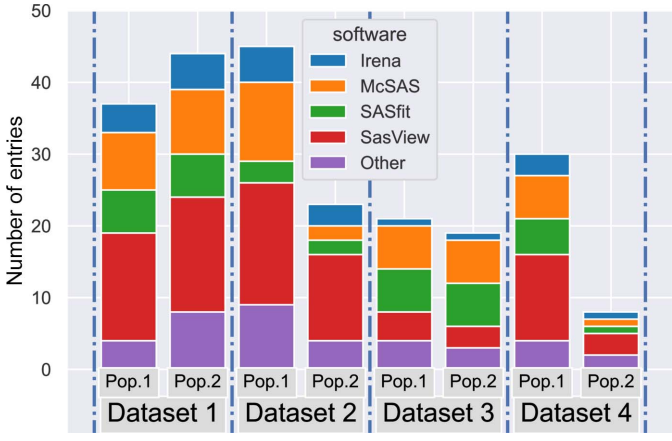


Figure 4
The number of entries for each population of each dataset (*y* axis), subdivided by software package. Datasets 1 and 3 were from samples with a nominally bimodal distribution, and datasets 2 and 4 from samples with a nominally monomodal distribution.

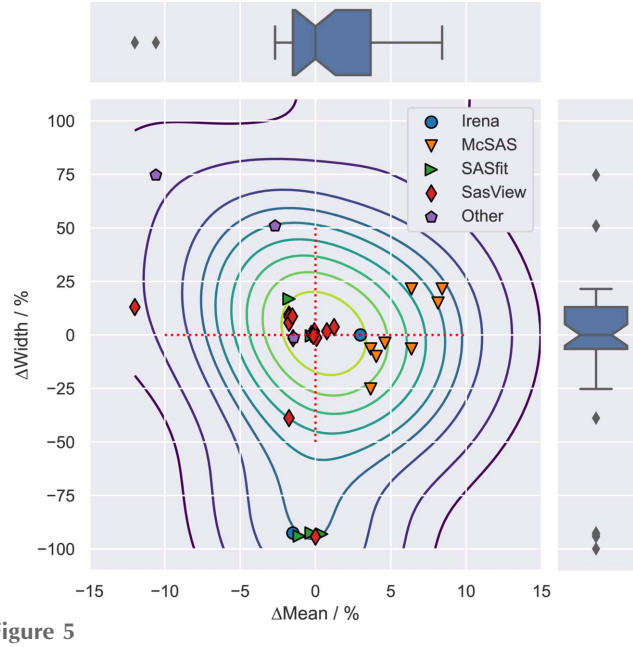


Figure 5
The deviation from the median in percent, for both the mean and width of the smallest population in dataset 1, separated by software. The 2D kernel density estimate (KDE) is shown as contour lines, the median is highlighted with a red dashed line cross, and the notched box plots show the quartiles for their respective mean or width values, with the whiskers indicating the 95% confidence interval.

the instructions specifying that the widths should be reported as a standard deviation, many values were well outside the viable range for this specification, often hovering between 0 and 1. One intermediate conclusion from this is that, due to this reporting inconsistency, the reported widths are largely unusable for the purposes of comparison. In lieu of an acceptable solution, we will thus concentrate mainly on the reported population means for the remainder of the article.

Another interesting aspect is the clustering of the various software packages. For example, there is a cluster of *McSAS* results, slightly to the right of the mean. The cluster is offset slightly to the right probably because of the volume weighting of the results having an effect on the population means. This argument does not hold universally, however, with reported number- and volume-weighted values spanning the field. Moreover, in our previous RR study, we showed that near-identical results could be obtained regardless of software choice.

Lastly, it is clear that a knowledge gap exists with the users of some software packages *vis à vis* the weighting used for the reported population values (*i.e.* means and widths). This is shown by 44% of the *SasView* users indicating that the values are volume weighted against 51% reporting that they represent number-weighted values (and a few hedging their bets and not reporting weighting at all). For *SASfit*, this is 52% and 43%, respectively. For the record, *SasView* and *SASfit* both report number-weighted population statistics. *SASfit* plots the distribution in a volume-weighted form by default [as indicated by the *y*-axis label $N(R)R^3$] but always exports the number-weighted distribution $N(R)$, adding to the confusion. This seems to be a user-interface issue, as no such confusion appears to exist with the users of *McSAS* and *Irena*, where 97% and 78%, respectively, reported the values as representing volume-weighted values.

3.3. Findings on all datasets

When the relative spread of the submitted population means are compared for each population in each dataset, it becomes apparent that the more challenging powder-based samples exhibit a much larger spread (Fig. 6). This can be attributed to the complications posed by the presence of a significant structure factor in dataset 3, and the near-fractal broadness of the distribution in combination with a structure factor underlying dataset 4. Changes in the chosen structure factor, or the structure-factor (local) volume-fraction parameter, can significantly affect the determined means. Likewise, differences in size-distribution models can equally impact the end result.

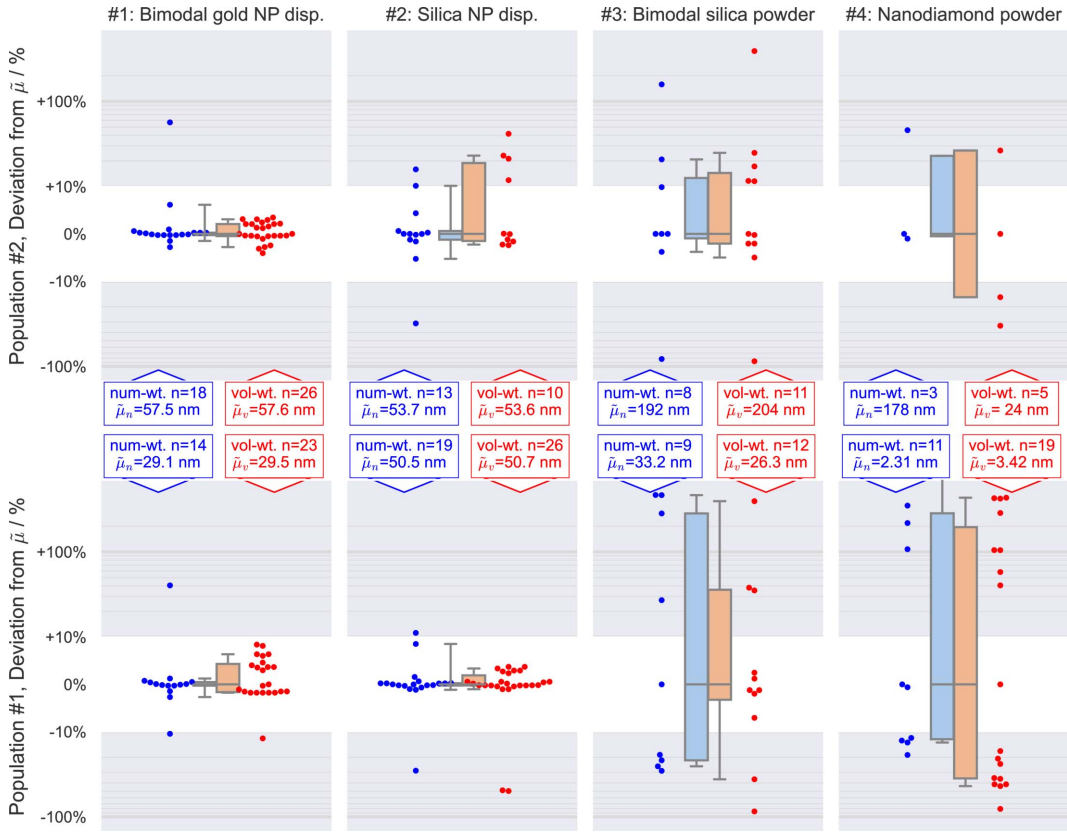


Figure 6
The precision of the means submitted to the DARR. This is represented as a deviation of the reported means from the median mean in percent for each population for each dataset, separated by number- (blue) and volume-weighting (red), each referenced to their respective median means ($\overline{\mu}_n$ and $\overline{\mu}_v$). These data are represented as boxplots (showing quartiles, with the whiskers indicating the 95% confidence interval) accompanied by the individual datapoints. The plots are shown on a symmetrical logarithmic scale with the lighter region representing a linear region covering $-10 \leq \overline{\mu}\,(\%) \leq 10$. Both populations of each dataset contain their absolute median mean and number of entries in the inset box. A larger spread of the submitted values can be observed for the more challenging powder-based samples compared with the dispersed samples.

This implies that a distinction could be made of the results between the deviations of entries of datasets 1 and 2, and of 3 and 4. Once this is done (Fig. 7), we can conclude that, for low-concentration dispersions, the 95% confidence interval of the determination of the population means can be determined within ∼10%. The widths, however, are not as consistent between the participants, probably because of the aforementioned inconsistencies in the reporting between the various packages, and as it stands can vary by more than 100%. The same analysis for the powder samples shows a remarkably broad distribution of results, indicating that blind intercomparisons between powder results of distinct laboratories may not yet be reliable. This could, perhaps, be improved by agreeing on a consistent approach for analysis of such samples (*cf.* Section 4.3).

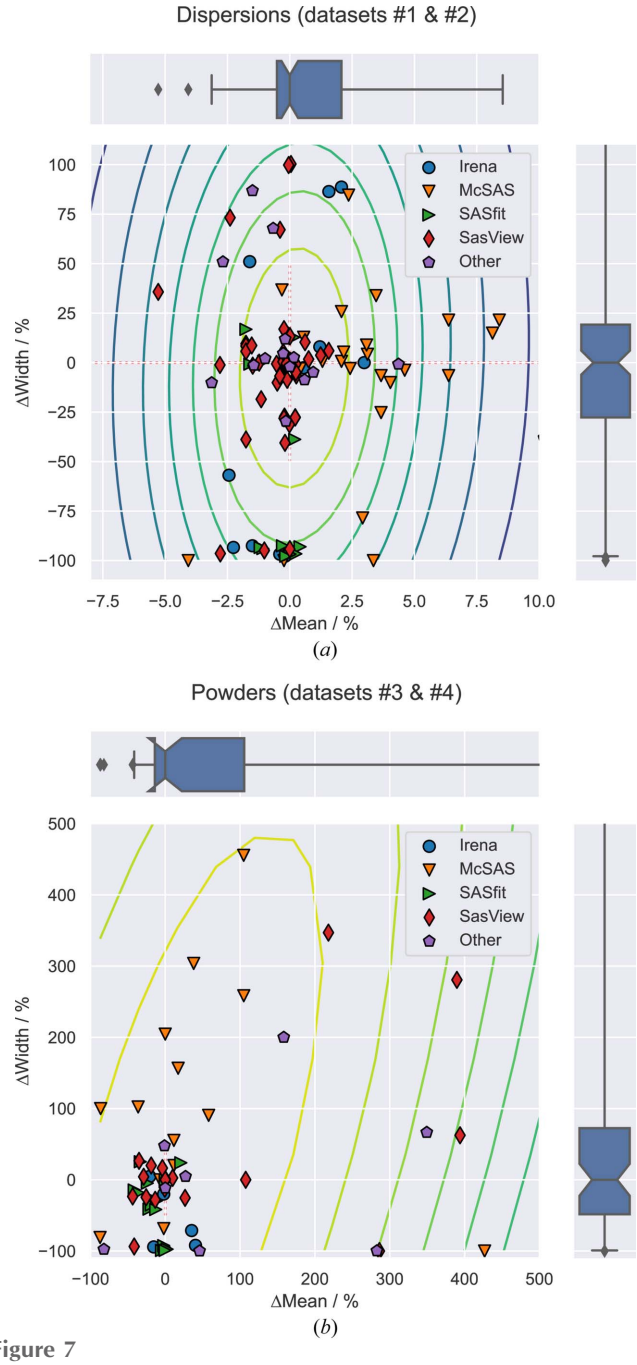


Figure 7
All entries for both populations of (*a*) datasets 1 and 2 and (*b*) datasets 3 and 4, shown as deviations from the median population mean (horizontal) and deviations from the median population width (vertical). Datapoints are grouped by software. The 2D KDE is shown as contour lines, the median is highlighted with a red dashed line cross, and the notched box plots show the quartiles for their respective mean or width values, with the whiskers indicating the 95% confidence interval.

3.4. Volume fractions

Participants were asked to enter information on the volume fractions for each population where available. To enable this determination, the data were scaled to absolute units (though, perhaps, using an incorrect thickness for the diamond powder and bimodal silica samples, as the apparent thickness of the materials was used, based on their X-ray absorption rather than the thickness of the actual container in which they resided).

While volume fractions are unambiguously defined on paper, the results show large inconsistencies in submitted values, though some clustering is present (Fig. 8). The origin of the spread, and thus the path through which it can be corrected, is not immediately evident. On the positive side, the volume fractions for the bimodal silica powder and nanodiamond powder are within a realistic order of magnitude. The sole but unsatisfactory conclusion is that there is unifying work to be done as well as cross checks on how the volume fractions are computed and presented by the various software packages.

4. Discussion

4.1. Just a moment: on number- versus volume-weighting

The various software packages typically present the key population statistics based either on a number-weighted or on a volume-weighted distribution. A mean size, for example, can thus be expressed as the mean size by number or the mean size by volume (or, more precisely, by mass).

We can express these mathematically by using the definition of weighted sample moments as a basis. This allows us to define total amount (zeroth raw moment), mean (first raw moment), variance (second central moment) and any higher (central) moments, $k > 2$ (with increasing uncertainty), using the equations in Table 2.

From these, the width σ is obtained from the variance m_2 through

$$\sigma = (m_2)^{1/2}, \tag{1}$$

and an optional adjustment for sampling bias to obtain unbiased moments can be determined through

$$m_{k,\text{unbiased}} = m_k \frac{N}{N-1}. \tag{2}$$

Maths aside, the practical difference between the two weightings is that a volume-weighted mean is always larger than a number-weighted mean, with increasing discrepancy

Table 2
Defining number- and volume-weighted moments m of populations consisting of N contributions.

The subscripts 'n' and 'v' denote number- and volume-weighting, respectively. D_i is the diameter of an object (sample) i and $v_{f,i}$ is the volume fraction of the same.

k	Meaning	Number weighted	Volume weighted
$k = 0$	Integral value	$m_{0,n} = 1$	$m_{0,v} = v_f = \sum_{i=1}^{N} v_{f,i}$
$k = 1$	Sample mean	$\mu_n = \frac{1}{N}\sum_{i=1}^{N} D_i$	$\mu_v = \frac{1}{v_f}\sum_{i=1}^{N} D_i v_{f,i}$
$k \geq 2$	Variance, skew, kurtosis *etc.*	$m_{k,n} = \frac{1}{N}\sum_{i=1}^{N}(D_i - \mu_n)^k$	$m_{k,v} = \frac{1}{v_f}\sum_{i=1}^{N}(D_i - \mu_v)^k v_{f,i}$

for broader distributions (the other moments are also non-interchangeable as they define different population distributions). While previous studies found that the information in a SAS dataset closely represents a volume-weighted distribution (Pauw *et al.*, 2013), there is nothing stopping analytical fitting methods from modelling a size distribution using number-weighted parameters, albeit with increasing uncertainty on the smaller end as the distribution broadens. As this uncertainty on the distribution is not normally shown in analytical modelling packages, a user can be led to believe that such a number-weighted distribution is determined with equal precision over the entire range. Results from this RR, for example from dataset 1 (see above), show either that a misconception exists on what the values reported by some software packages represent or that it is unclear that volume-weighted and number-weighted parameters are inherently different.

4.2. A word on self-assessed experience

Several information fields were provided that at least tenuously link to the experience of the participant. These are the working years, the percentage of SAS in their working life and a self-assessment of their level of knowledge (on a scale of 1–10). While this information offers only a very crude quantification of the scattering career of each individual (missing information, for example, includes the field of expertise of the participants, the changes in percentage of SAS in their working life over the years, experience with analysis in particular *etc.*), we can attempt to derive some insights from this. As is to be expected (Fig. 9), the self-assessed level of knowledge does correlate weakly with cumulative years of working with SAS, calculated as the product of the percentage of SAS in their working life with the years of their working life. In other words, the longer and more they work in the field, the higher their estimate of their working knowledge.

It is perhaps to be expected that some of these 'experience' measures would correlate with the proximity of their result to the median results, under the assumption that the median is proximate to the target or 'true' value for a given population.

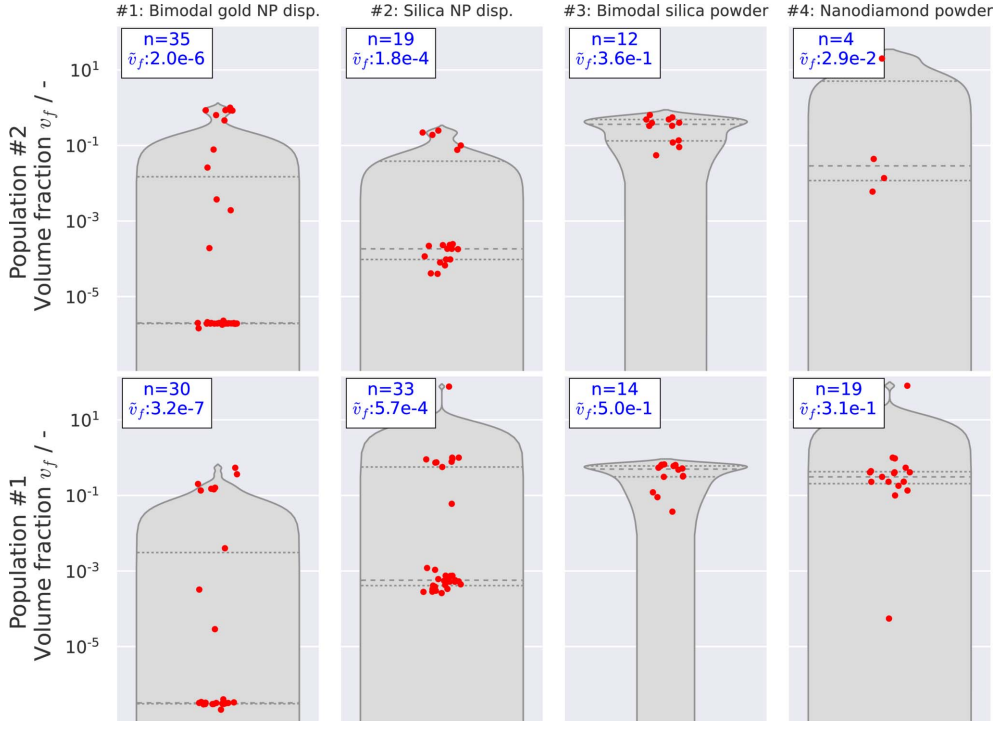


Figure 8
The volume fractions for each population of each dataset (in fraction, not percent), as reported by each participant. The overall distribution of values is visualized as a violin plot, where the width of the plot represents the number of entries with that approximate volume fraction. The violin plot also shows the quartiles as dashed lines, while the red dots are the entries. The median volume fraction is shown with the number of reported volume fractions in the text box of each plot.

Figure 9
The self-assessed degree of SAS knowledge as a function of the cumulative SAS experience, showing a weak correlation as expected.

Fig. 10 shows that such a correlation is, if present, only weakly present. This is unfortunate, as it would imply that we are not automatically getting better with more experience. One explanation could be that the true genius of the participants is being held back by both the limitations in the reporting by the software and the mentally taxing needless dichotomies found in the field (*cf.* Section 4.3). It seems, then, that in order to improve as a community, we need to do more than merely get older.

4.3. Potential steps for immediate improvement

Apart from the population means for the dispersions, the large spread of the remaining population parameters found in this work highlights that the human factor has the potential to introduce a significant uncertainty into the overall SAS data-interpretation process. This uncertainty, as estimated in this study, is much larger in magnitude than those arising from data collection and corrections alone (Pauw *et al.*, 2017*a*; Smales & Pauw, 2021; Schavkan *et al.*, 2019). Some of the difficulties associated with the interpretation of scattering data start when researchers are faced with a barrage of possible units, non-standardized data formats and poorly specified data practices even before analysis can begin. Expecting unfamiliar users to gain an in-depth understanding of the various redundant units in circulation, in addition to the strengths and pitfalls of each analysis method, and to understand the differences in their implementations in respective fitting software forms a high barrier of entry. A further problem is that this barrier of entry appears invisible to many within the community (or worse: is considered a rite of passage), all of which can easily lead to unsatisfactory interpretations.

To alleviate this, our community should refrain from actively confusing users through a lack of constraint and definition. In other words, instrument responsibles in collaboration with software developers have to agree on – and themselves adhere to – a consistent set of units and definitions. Secondly, universal guides should be established (perhaps by a CanSAS or IUCr working group) on how to approach data-analysis challenges of common sample types, using a range of tools, rather than relying on local knowledge transfer alone. Lastly, users of software packages should take some time to read software documentation and understand the values the software is presenting. Conversely, software documentation can be written to contain easy-to-understand sections, but, while some useful tutorials exist (https://www.sasview.org/docs/user/tutorial.html; SASfitScience, 2023; Pauw & Bressler, 2023), much work remains to be done (Wuttke *et al.*, 2022).

Thus, immediate improvements in interlaboratory result consistency may be obtained through:

(1) Gradually aligning the information reported by the various software packages, *e.g.* presenting universal population information in the form of distribution moments (total value, mean, variance, skew and kurtosis) for each population.

(2) Providing user guides for approaching standard scattering analysis problems, giving robust model suggestions and adaptation approaches for dilute as well as dense systems. Additional methods for rough-estimate cross checks and result validation should be provided as well, *i.e.* making sure the dimensions are commensurate with the Q range *etc.*

(3) Introducing and using practically reasonable data-uncertainty estimates [*e.g.* employing methods such as those used by Smales & Pauw (2021)] in fits, so that reliable datapoints weigh more heavily than unreliable datapoints. This will result in better fits as well as better uncertainty estimates on the morphological parameters resulting from the fits.

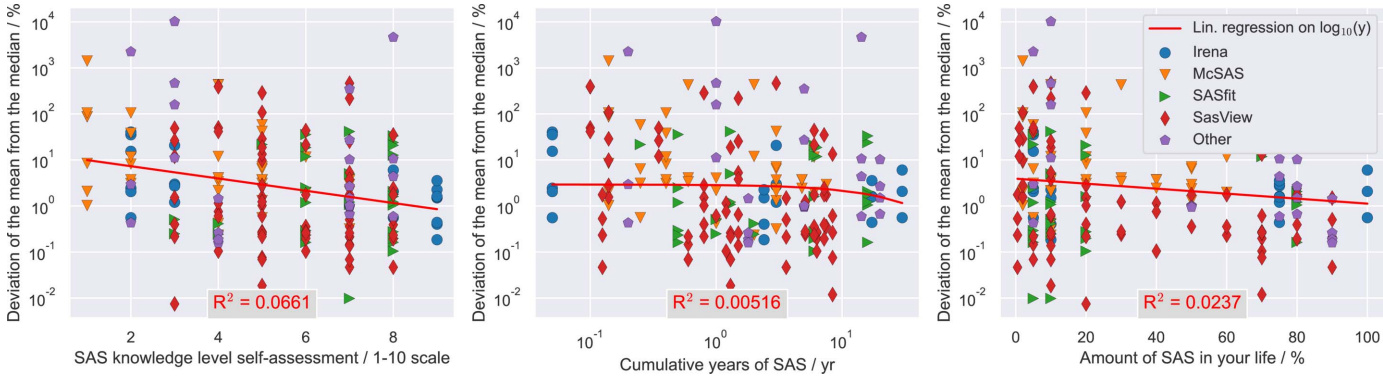


Figure 10
Weak evidence of correlations between several experience measures (SAS knowledge, cumulative years of SAS and the percentage of SAS in life) and the closeness to the median result. The results unfortunately show little to no correlation.

(4) The provision of consistent and comparable goodness-of-fit measures to qualify a fit. When good uncertainty estimates are provided on the data, these goodness-of-fit measures will also have meaning.

(5) Removing trite, time-consuming, yet unnecessarily confusing dichotomies and the risks of errors in the therefore required unit conversions (although the mechanism by which any community may agree on one of two options is itself a veritable wasp's nest of conflict). A non-exhaustive list could be:

(i) Default units of (Q) should be defined (*e.g.* nm^{-1}).

(ii) Default units of I should be defined [*e.g.* $(m\ sr)^{-1}$].

(iii) Size should consistently refer to the full length (diameter) of objects instead of occasionally referring to the half length (radius) for select shapes. This way, mistakes in factor-of-two shifts are avoided when moving to other scatterer shapes.

(iv) Population information should be either volume weighted, for a closer reflection of the information content of a scattering pattern, or number weighted, but whichever it is, it has to be clearly and repeatedly indicated.

4.4. Tips for future round-robin experiment designs

No experiment is perfect, and this RR is no exception. Future iterations may include the following improvement suggestions.

One way to bring together a larger community and gain insight into the progression (or regression) of the agreement is to stage regular smaller RR studies. This could be as straightforward as providing one dataset per semester. This has the added advantage of building up a library of data and fit examples.

Further separation of the effects of the user versus that of the software will help to identify the main source of uncertainty. To that end, some RR studies could dictate the use of a particular software package or a particular model.

Cross evaluation of the quality of fits might allow an assessment on the level of agreement on what constitutes a 'good fit'. This can also lead to the identification of the target or best fit to compare against. Following on from this, all necessary metadata required to reproduce a fit should be preserved by each participant, so that sources of disagreement can be better identified once a consensus fit has been established.

5. Conclusions

A round-robin study has been carried out to study the effect of individual researchers on the numerical results of a small-angle scattering pattern analysis. Before analysis, several results required corrections to compensate for field-specific dichotomies resulting from omitted or incorrectly applied unit conversions in ingestion as well as reporting.

The results highlight a narrow spread in determined population means for samples consisting of low-concentration dispersions of globular scatterers, with half of the entries falling within 1.5% of the median mean. For more challenging scattering patterns of concentrated powders, the spread is considerable to excessive, with half of the entries within 44% of the median mean. This is probably due to the results being additionally affected by the choice of structure-factor model and volume fraction. The determined population widths for both types vary wildly and are ostensibly incomparable due to the differences in parameters that are reported by the various software packages (this, despite the nominal answer format specifying specifically a width in the form of a standard deviation). Lastly, considerable confusion exists on whether some software packages report fitting parameters as volume- or number-weighted values.

Additionally, while participants do estimate their knowledge to be higher the longer they work with the method, this does not strongly correlate to a closer proximity to the median means. Therefore, alternative suggestions (*i.e.* besides acquiring years of professional experience) are provided in Section 4.3 that could help improve the intercomparability of obtained results, in particular for widths and volume fractions. The implementation of a subset of these is bound to have a positive effect on the comparability of scientific results obtained with small-angle scattering.

6. Data availability

The data as well as the *Jupyter Notebook* used for data sanitizing, analysis and graphing are available on the Zenodo open-access data repository (https://zenodo.org/records/7509710). Further analysis and extension of these data are strongly encouraged.

Acknowledgements

Certain commercial equipment, instruments, materials or software are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose. Open access funding enabled and organized by Projekt DEAL.

Funding information

KT acknowledges the NIST–NRC postdoctoral fellowship program for support. This work was partially funded through the European Metrology Programme for Innovation and Research (EMPIR) project No. 17NRM04.

References

Bartczak, D. & Hodoroaba, V.-D. (2022). *Report on the Development and Validation of the Reference Material Candidates with Non-spherical Shape, Non-monodisperse Size Distributions and Accurate Nanoparticle Concentrations (Deliverable D3)*, https://zenodo.org/record/7016466.

Basham, M., Filik, J., Wharmby, M. T., Chang, P. C. Y., El Kassaby, B., Gerring, M., Aishima, J., Levik, K., Pulford, B. C. A., Sikharulidze, I., Sneddon, D., Webber, M., Dhesi, S. S., Maccherozzi, F., Svensson,

O., Brockhauser, S., Náray, G. & Ashton, A. W. (2015). *J. Synchrotron Rad.* **22**, 853–858.
Bender, P., Balceris, C., Ludwig, F., Posth, O., Bogart, L. K., Szczerba, W., Castro, A., Nilsson, L., Costo, R., Gavilán, H., González-Alonso, D., Pedro, I., Barquín, L. F. & Johansson, C. (2017). *New J. Phys.* **19**, 073012.
Bressler, I., Pauw, B. R. & Thünemann, A. F. (2015). *J. Appl. Cryst.* **48**, 962–969.
Deumer, J. & Gollwitzer, C. (2022). *npSize_SAXS_data_PTB*, https://doi.org/10.5281/zenodo.5886834.
Dong, Y., Etienne, A., Frolov, A., Fedotova, S., Fujii, K., Fukuya, K., Hatzoglou, C., Kuleshova, E., Lindgren, K., London, A., Lopez, A., Lozano-Perez, S., Miyahara, Y., Nagai, Y., Nishida, K., Radiguet, B., Schreiber, D. K., Soneda, N., Thuvander, M., Toyama, T., Wang, J., Sefta, F., Chou, P. & Marquis, E. A. (2019). *Microsc. Microanal.* **25**, 356–366.
Hopkins, J. B., Gillilan, R. E. & Skou, S. (2017). *J. Appl. Cryst.* **50**, 1545–1553.
Ilavsky, J. & Jemian, P. R. (2009). *J. Appl. Cryst.* **42**, 347–353.
Kohlbrecher, J. & Breßler, I. (2022). *J. Appl. Cryst.* **55**, 1677–1688.
Krumrey, M. & Ulm, G. (2001). *Nucl. Instrum. Methods Phys. Res. A*, **467–468**, 1175–1178.
Madsen, I. C., Scarlett, N. V. Y., Cranswick, L. M. D. & Lwin, T. (2001). *J. Appl. Cryst.* **34**, 409–426.
Osterrieth, J., Rampersad, J., Madden, D. G., Rampal, N., Skoric, L., Connolly, B., Allendorf, M., Stavila, V., Snider, J., Ameloot, R. *et al.* (2022). *ChemRxiv*. Cambridge Open Engage.
Pauw, B. R. & Bressler, I. (2022). *McSAS3*, https://github.com/BAMresearch/McSAS3.
Pauw, B. R. & Bressler, I. (2023). *McSAS Quick Usage Guide*, https://mcsas.readthedocs.io/en/latest/quickstart.html.
Pauw, B. R., Kästner, C. & Thünemann, A. F. (2017*a*). *J. Appl. Cryst.* **50**, 1280–1288.
Pauw, B. R., Pedersen, J. S., Tardif, S., Takata, M. & Iversen, B. B. (2013). *J. Appl. Cryst.* **46**, 365–371.
Pauw, B. R., Smith, A. J., Snow, T., Terrill, N. J. & Thünemann, A. F. (2017*b*). *J. Appl. Cryst.* **50**, 1800–1811.
Pollen Metrology (2021). *nPSize CEA Images as 2D Arrays*, https://zenodo.org/record/5578680.
Rennie, A. R., Hellsing, M. S., Wood, K., Gilbert, E. P., Porcar, L., Schweins, R., Dewhurst, C. D., Lindner, P., Heenan, R. K., Rogers, S. E., Butler, P. D., Krzywon, J. R., Ghosh, R. E., Jackson, A. J. & Malfois, M. (2013). *J. Appl. Cryst.* **46**, 1289–1297.
SASfitScience (2023). *SASfit Tutorials*, https://www.youtube.com/@SASfitScience.
Scarlett, N. V. Y., Madsen, I. C., Cranswick, L. M. D., Lwin, T., Groleau, E., Stephenson, G., Aylmore, M. & Agron-Olshina, N. (2002). *J. Appl. Cryst.* **35**, 383–400.
Schavkan, A., Gollwitzer, C., Garcia-Diez, R., Krumrey, M., Minelli, C., Bartczak, D., Cuello-Nunez, S., Goenaga-Infante, H., Rissler, J., Sjostrom, E., Baur, G. B., Vasilatou, K. & Shard, A. G. (2019). *Nanomaterials*, **9**, 502.
Smales, G. J. & Pauw, B. R. (2021). *J. Instrum.* **16**, P06034.
Taché, O., Spalla, O., Thill, A., Carriere, D., Testard, F. & Sen, D. (2017). *pySAXS, an Open Source Python Package and GUI for SAXS Data Treatment*, https://iramis.cea.fr/en/Phocea/Vie_des_labos/Ast/ast_sstechnique.php?id_ast=1799.
Trewhella, J., Vachette, P., Bierma, J., Blanchet, C., Brookes, E., Chakravarthy, S., Chatzimagas, L., Cleveland, T. E., Cowieson, N., Crossett, B., Duff, A. P., Franke, D., Gabel, F., Gillilan, R. E., Graewert, M., Grishaev, A., Guss, J. M., Hammel, M., Hopkins, J., Huang, Q., Hub, J. S., Hura, G. L., Irving, T. C., Jeffries, C. M., Jeong, C., Kirby, N., Krueger, S., Martel, A., Matsui, T., Li, N., Pérez, J., Porcar, L., Prangé, T., Rajkovic, I., Rocco, M., Rosenberg, D. J., Ryan, T. M., Seifert, S., Sekiguchi, H., Svergun, D., Teixeira, S., Thureau, A., Weiss, T. M., Whitten, A. E., Wood, K. & Zuo, X. (2022). *Acta Cryst.* D**78**, 1315–1336.
Wernecke, J., Gollwitzer, C., Müller, P. & Krumrey, M. (2014). *J. Synchrotron Rad.* **21**, 529–536.
Whitfield, P. S. (2016). *Powder Diffr.* **31**, 192–197.
Wuttke, J., Cottrell, S., Gonzalez, M. A., Kaestner, A., Markvardsen, A., Rod, T. H., Rozyczko, P. & Vardanyan, G. (2022). *J. Neutron Res.* **24**, 33–72.
Xenocs (2022). *Xenocs XSACT Software*, https://www.xenocs.com/saxs-products/xsact-software/.