Improvements and considerations for size distribution retrieval from small-angle scattering data by Monte Carlo methods

Brian R. Pauw,[a,b,c]* Jan Skov Pedersen,[d] Samuel Tardif,[e] Masaki Takata[b,e] and Bo B. Iversen[a]

[a]Centre for Materials Crystallography, Department of Chemistry and iNANO, Aarhus University, DK-8000 Aarhus, Denmark, [b]Structural Materials Science Laboratory, RIKEN SPring-8 Centre, Hyogo 679-5148, Japan, [c]International Centre for Young Scientists, National Institute of Materials Science, Tsukuba 305-0047, Japan, [d]Department of Chemistry and iNANO, Aarhus University, DK-8000 Aarhus, Denmark, and [e]Quantum Order Research Group, RIKEN SPring-8 Centre, Hyogo 679-5148, Japan. Correspondence e-mail: brian@stack.nl

Monte Carlo (MC) methods, based on random updates and the trial-and-error principle, are well suited to retrieve form-free particle size distributions from small-angle scattering patterns of non-interacting low-concentration scatterers such as particles in solution or precipitates in metals. Improvements are presented to existing MC methods, such as a non-ambiguous convergence criterion, nonlinear scaling of contributions to match their observability in a scattering measurement, and a method for estimating the minimum visibility threshold and uncertainties on the resulting size distributions.

1. Introduction

The search for generally applicable methods capable of determining structural parameters from small-angle scattering patterns for a broad range of samples has yielded several viable methods. For monodisperse systems consisting of identical particles, these methods attempt to find a form-free solution to the pair-distance distribution function, whereas for polydisperse systems the aim is to determine the particle size distribution. In both cases, relevant transformation should yield the observed scattering pattern (Krauthäuser et al., 1996).

There are indirect transform methods (ITMs) based on regularization techniques, which impose that the solution is as smooth as possible (Glatter, 1977, 1979; Moore, 1980; Svergun, 1991; Pedersen, 1994), and Bayesian and maximum entropy ITMs, which find the most likely solution using a Bayesian approach and entropy maximization, respectively (Hansen, 2000; Hansen & Pedersen, 1991). There are also methods available based on Titchmarsh transforms for determining size distributions (Botet & Cabane, 2012; Mulato & Chamboulleyron, 1996; Fedorova & Schmidt, 1978).

Another class of methods, such as the structure interference method (SIM; Krauthäuser et al., 1996) and some Monte Carlo (MC) methods (Martelli & Di Nunzio, 2002; Di Nunzio et al., 2004), assume a particular shape and do not appear to require smoothness constraints. These only have a positivity constraint and have so far been limited to size distributions of sphere-shaped scatterers. These methods can be used to extract the particle size distribution function of systems of scatterers whose shape is known or assumed, and not affected by concentration effects (Krauthäuser et al., 1996; Martelli & Di Nunzio, 2002; Di Nunzio et al., 2004). The MC variant

approaches the optimization by trial and error, whereas the SIM uses a conjugate gradient approach (Krauthauser, 1994). Both are conceptually easier than the ITMs and methods based on Titchmarsh transforms and provide stable and unique solutions (Martelli & Di Nunzio, 2002; Di Nunzio et al., 2004; Krauthäuser et al., 1996).

Upon implementation of one such method by Martelli & Di Nunzio (2002), hereafter referred to as 'the Martelli method', several noteworthy changes were made to that method, which are presented here. We briefly reiterate the working principle and highlight the differences of the presented MC method from the Martelli method. Then, a general solution for detection limits is derived for particles in a polydisperse set. This aids the MC method as it allows for improved contribution scaling during the optimization procedure and indicates detectability limits in the final result. Lastly, a convergence criterion is defined for the MC method, allowing for the calculation of uncertainties in the resulting size distribution. This method is shown to be applicable to scattering data obtained during the synthesis of AlOOH nanoparticles.

2. A brief overview of the implemented method

The MC method proposed here is essentially identical to that by Martelli & Di Nunzio (2002). The differences can be summarized as follows: (1) we apply a help variable p_c, which is used in the optimization to (partially) compensate for the effect of size on the scaling of the form-factor scattering, whose magnitude is derived through the definition of 'maximum observability'; (2) a limited fixed number of sphere contributions is used and constantly replaced, this number

being determined from the optimization speed; (3) a strict convergence criterion is defined on the basis of the estimated data uncertainty; and (4) standard deviations are calculated on the result through repetitive application of the MC method. A summary of the MC method is given below.

2.1. Step 0: determination of the feature size limits

If not explicitly defined, a reasonable estimate has to be made for the radius range of the determinable features in the scattering pattern. This estimate is based on the measured data set, in particular the minimum measured q and the spacing between subsequent data points in q. Here, q is defined as $q = (4\pi/\lambda)\sin\theta$, with λ the wavelength of the radiation and 2θ the scattering angle. The radius of the maximum detectable feature R_{max} in a small-angle scattering measurement is mostly defined by the size of the direct beam spot on the detector. As this is often wholly (but just) obscured by the beamstop, a reasonable assumption for $R_{max} \simeq \pi/q_{min}$, where q_{min} is the smallest measured value of q in the corrected data set. If, however, the beamstop is oversized compared to the actual beam size, R_{max} may be more closely related to the spacing between q data points, so that $R_{max} \simeq \pi/\Delta q$, where Δq is the smallest distance between two neighbouring q points in the data set.

Similarly, a reasonable estimate for the minimum radius of distinguishable features R_{min} can be defined. As above, this can be related to the largest available q value in the data points of the measured data set. It thus can assume the value of $R_{min} \simeq \pi/q_{max}$, where q_{max} is the largest measured value of q in the corrected data set. Any value for R_{min} smaller than about 0.3 nm, however, would define features approaching the length of chemical bonds, where the commonly used assumption of uniform electron density no longer applies. In the interest of generality in the MC method, R_{min} is set to zero, but it is of importance to consider these fundamental limitations when interpreting features in the result.

If no *a priori* information on the bounds is supplied, the radius range is thus set to $0 < R \leq R_{max}$, where R_{max} is the largest of the two aforementioned alternatives.

2.2. Step 1: preparation of the procedure

The initial guess of the total scattering cross section is calculated for a fixed number of contributions n_s (typically 200–300) from spheres whose radii R are randomly sampled from a bound uniform distribution (bound by the radius range described in the previous section). These radii can therefore assume any value within the size bounds. This initial guess is calculated using the general equation

$$I_{MC}(q) = b + A \sum_{k=1}^{n_s} |F_{sph,k}(qR_k)|^2 \left(\tfrac{4}{3}\pi\right)^2 R_k^{(6-p_c)}, \qquad (1)$$

where the sum runs over all spheres in the finite set and $F_{sph,k}(qR_k)$ is the Rayleigh form factor for sphere k, normalized to 1 for $q = 0$. R_k is the radius for sphere k. p_c is a parameter adjustable in the range $0 \leq p_c \leq 6$, biasing the volume weighting of the contributions, and will be discussed in

detail hereafter. b is a constant background term, and A is a scaling factor, which is related to the volume fraction φ of the scatterers through

$$A = \varphi\Delta\rho^2 \sum_{k=1}^{n_s} 1 \Big/ \left[\tfrac{4}{3}\pi R_k^{(3-p_c)}\right]. \qquad (2)$$

Here the sum runs over all spheres in the finite set and $\Delta\rho$ is the scattering contrast. The volume fraction of the scatterers φ is defined as

$$\varphi = V_{scatt}/V_{irr}, \qquad (3)$$

where V_{irr} is the irradiated sample volume and V_{scatt} is the total scatterer volume in V_{irr}.

Given the limited (fixed) number of spheres used for the calculation of the scattering cross section, a large sphere could easily dominate if normal volume-squared scaling is used for each sphere contribution (*i.e.* $p_c = 0$). As an example, a sphere of 10 nm would require a million spheres of 1 nm for an equivalent contribution (at $q = 0$). It is evident, therefore, that describing the scattering from polydisperse sets of spheres would require very large numbers of smaller sphere contributions to have any distinguishable impact on the scattering on top of that originating from a few large scatterers. The factor p_c in equation (1) is thus used as a computational aid to partially suppress this effect of the volume-squared sphere contribution scaling, by increasing the contribution from the smaller spheres in the set. The recommended value for p_c is 3, but in practice it can be freely varied in the range $2 \leq p_c \leq 3.3$ as shown in §2.5.

The choice of p_c only affects the speed of the calculation and does not affect the result as it is compensated for in A [equation (2)] and therefore also when visualizing the results (it can, as will be discussed later, also slightly affect the standard deviation of the result when set to extreme values, as shown in the supplementary information[1]). Setting $p_c = 3$ results in the internal calculation using volume-weighted spheres (as opposed to volume-squared-weighted spheres for $p_c = 0$), and setting $p_c = 2$ makes each sphere in the MC procedure assume an identical 'observability', which is a proposed measure for the maximum visibility of a scatterer in a polydisperse system, defined and explored in §3. Lastly, $p_c = 0$ removes this compensation altogether from the procedure; this approach is suitable only for very narrow distributions owing to the limited number of spheres used in the MC calculation.

A and b are tuned to the measured scattering through optimization using a least-squares residual minimization procedure, minimizing the reduced chi squared χ_r^2 (Pedersen, 1997):

$$\chi_r^2 = \frac{1}{N-M} \sum_{i=1}^{N} \left[\frac{I_{meas}(q_i) - I_{MC}(q_i)}{\sigma(q_i)}\right]^2, \qquad (4)$$

---

[1] Supplementary material is available from the IUCr electronic archives (Reference: CE5145). Services for accessing this material are described at the back of the journal.

where N denotes the number of data points and I_{meas} and I_{MC} the measured and calculated model scattering cross section, respectively. $\sigma(q_i)$ is the estimated error on each measured data point, whose estimation method is detailed in §4. M is the number of degrees of freedom in the fitting model, but is unfortunately ill-defined in an MC model and is thus set equal to 2 to account for the scaling parameter A and background contribution parameter b.

2.3. Step 2: optimization cycle

After random selection of the n_s sphere radii and the subsequent calculation of the initial guess, the Monte Carlo optimization cycle then begins. A random sphere is selected from the set of n_s spheres, and a change of its radius to another random value within the previously discussed radius bounds is suggested. The intensity after this suggested change is calculated by subtracting the contribution of the previous sphere radius and adding the contribution of the new sphere radius, and re-optimizing the scaling factor A and background level b. The suggested radius change is accepted if it improves the agreement between measured and MC intensity, *i.e.* if the change reduces the χ_r^2 value. Similar to the SIM and the Martelli method, a rejection–acceptance mechanism that occasionally accepts 'bad moves' was found not to be necessary (Krauthäuser *et al.*, 1996; Martelli & Di Nunzio, 2002; Di Nunzio *et al.*, 2004).

This method differs from the Martelli method in that the Martelli method continually attempts to add new sphere contributions to an ever growing set, leaving the prior established set of contributions untouched. Conversely, the adaptation presented here leaves the number of sphere contributions in the set unchanged and repeatedly tries to change the radius of a random contribution in the set. An attempt was made to adjust the MC method here to add spheres instead of changing sphere radii, more closely following the method presented by Martelli. It was found, however, that such a modification would occasionally prevent convergence.

2.4. Step 3: convergence and post-optimization procedures

The optimization is ceased once the condition $\chi_r^2 < 1$ has been reached (*cf.* §4). If convergence has not been reached within a certain number of optimization cycles (here set to one million), and fails to do so repeatedly, the pattern is considered unsuitable for fitting with this method.

After convergence has been reached, the partial volume fraction for a sphere contribution k, φ_k, is then calculated through reformulation of equation (2):

$$\varphi_k = A \tfrac{4}{3}\pi R_k^{(3-p_c)}/\Delta\rho^2, \qquad (5)$$

where A is known through least-squares fitting from step 2. If this equation is calculated for all n_s sphere contributions, the volume fraction of the scatterers φ can be calculated as $\varphi = \sum_{k=1}^{n_s} \varphi_k$.

For visualization and analysis purposes, the obtained partial volume fractions φ_k can be distributed over a number of radius bins to form a histogram or a volume-weighted size distribu-

tion (an example of which is given in §7). In this histogram, each radius bin (denoted with the subscript Rbin) is defined by its bin edges, R_n and R_{n+1}, also known as the class limits. If a particular sphere contribution radius falls within the bin, its volume fraction φ_k is added to the contents of the bin. In other words, each radius bin defined by its edges R_n and R_{n+1} will assume the value that is the sum of all partial volume fractions of contributions whose radii fall within that bin to make up the bin volume fraction $\varphi_{R\mathrm{bin}}$. If instead of the volume-weighted size distribution a number-weighted size distribution is sought, this procedure is identical except that each contribution adds $\varphi_k V_{irr}/(\tfrac{4}{3}\pi R_k^3)$ to the bin.

Accurate mean and standard deviations for the final volume-weighted size distribution are calculated from the results of several tens of independent repetitions of the whole MC procedure. This is considered to provide a reasonable check for ambiguity, and such standard deviations may aid in the comparison of the size distribution obtained with small-angle scattering with those obtained using other methods.

2.5. Determination of the number of contributions n_s and compensation parameter p_c

A suitable value must be found for the number of contributions n_s in the MC method. Too large a number of contributions will result in excessive computation time and may effect an overly smooth solution and underestimated standard deviations on the result (an effect shown in the supplementary information). Too few contributions will also lead to a long computation time as the few contributions need to assume increasingly strict radii. Furthermore, the number of required contributions is dependent on the data quality, with data of higher quality requiring more contributions to achieve convergence.

It is therefore suggested that n_s is set so that convergence is achieved in as few MC iterations as possible, *i.e.* an efficiency-maximizing approach. Plotting the number of iterations required for convergence *versus* n_s will show what value should be used for a particular data set (Fig. 1). For consis-
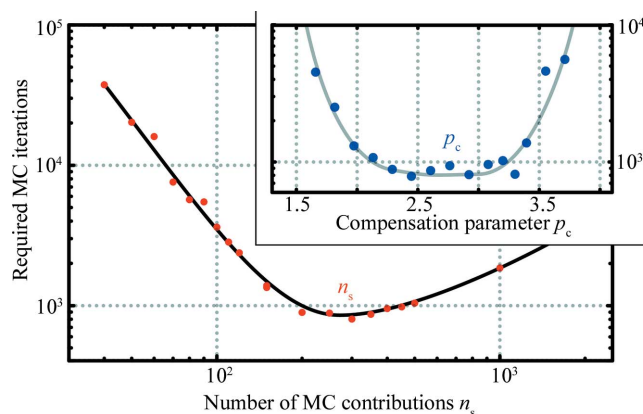
Figure 1
Optimization of the number of MC contributions n_s (main figure) and p_c (upper right, calculated for 300 contributions) for the Boehmite example data (see §6) by selecting the values requiring the least number of steps. Lines are added to guide the eye.

tency in series of measurements, such as time or temperature series, it is recommended to fix n_s to a single value suitable for the majority of the measurements.

A similar approach can be used to fine-tune p_c. As shown in Fig. 1, however, there are a wide range of values with nearly identical calculation efficiencies. This indicates that suitable choices for this example lie within the range $2.1 \leq p_c \leq 3.2$.

3. Observability of isolated spheres in a polydisperse set

As the scattering power of particles scales proportionally to their volume squared, the scattering of smaller particles in a polydisperse set is quickly drowned out by the signal of the large particles. This effect, however, is partially compensated for by the different q dependence of the scattering of the smaller particles.

To investigate how large this compensatory effect is, we define the 'maximum observability' $\mathrm{Obs}_{\mathrm{max},i}$ of a particle i in a set N as the maximum of the ratio of its individual contribution $I_i(q)$ to the total intensity $I_{\mathrm{tot}}(q)$ in the range $[q_{\mathrm{min}}, q_{\mathrm{max}}]$ [equation (6)]. The location of the maximum is denoted as q_{m}.

$$\mathrm{Obs}_{\mathrm{max},i} = \frac{I_i(q_{\mathrm{m}})}{I_{\mathrm{tot}}(q_{\mathrm{m}})} = \max_{q \in [q_{\mathrm{min}}, q_{\mathrm{max}}]} \left[\frac{I_i(q)}{I_{\mathrm{tot}}(q)} \right]. \quad (6)$$

The individual contribution $I_i(q)$ of sphere i with radius R_i is given by equation (1) with $p_c = 0$.

The maximum observability has been calculated for three separate unimodal distributions, with each number-weighted size distribution $P(R)$ sampled using 50 000 sphere radii in the range $0.01 \leq R \leq 35$ nm. There are three distributions: (1) a uniform distribution, (2) a 'trailing' triangular distribution with its mode set to 0.01 nm and (3) a 'leading' triangular distribution with its mode set to 35 nm (as shown in the inset of Fig. 2). Fig. 2 shows the maximum observability to be unaffected by the choice of distribution. From the figure, it is evident that the maximum observability scales as $\mathrm{Obs}_{\mathrm{max},i} \propto R^2$ for particles with sizes larger than $R_{\mathrm{lim}} \simeq \pi/q_{\mathrm{max}}$. Particles smaller than R_{lim} exhibit an observability scaling in line with the volume-squared intensity scaling (R^6). A similar, independently determined lower limit has been obtained through MC size distribution retrieval exercises (available in the supplementary material).

The information on the observability can be used for three purposes. First and foremost, there is a clear indication of the limits of small-angle scattering for resolving the smaller sizes, which can be linked to the maximum measured q. This is evident from Fig. 2 as observability scales with radius to the sixth power for particles with radii smaller than $R_{\mathrm{lim}} \simeq \pi/q_{\mathrm{max}}$, which renders their contribution rapidly indistinguishable. This reinforces the point that the minimum detectable feature radius R_{min} can be estimated to be π/q_{max}, as discussed in §2.

Secondly, this information directly gives us a value of 2 for p_c, which will result in an equal maximum observability for every contribution in the MC method, provided that the contribution radii fall within the bounds imposed by the angular limits of the measurement. It is still considered

preferable, however, to tune p_c for each sample (series) according to the method described in §2.5.

Thirdly, the concept of observability can be used to determine whether a particular feature in a size distribution is significant or not. To be more precise, we can calculate what level is required in any radius bin (*cf.* step 3 of §2) in order to contribute scattering that can be distinguished in the fitted scattering pattern above the scattering pattern uncertainty $\sigma(q)$. In the MC calculation, a model component k contributes scattering $I_k(q)$ which is proportional to its partial volume fraction φ_k [where $I_k(q)$ is calculated using equation (1) for a single contribution and with $b = 0$]. Its scattering contribution can be considered to become distinguishable at a minimum partial volume fraction $\varphi_{\mathrm{min},k}$ where it contributes at least as much as the data uncertainty:

$$\left[I_k(q)/\varphi_k \right] \varphi_{\mathrm{min},k} \geq \sigma(q). \quad (7)$$

This allows the calculation of $\varphi_{\mathrm{min},k}$ in two ways: (1) as the minimum of all values calculated over $q \in [q_{\mathrm{min}}, q_{\mathrm{max}}]$ or (2) as the value at q_{m}, where the contribution from k is strongest [equation (6)], which in practice turns out to be a good approximation for (1):

$$\varphi_{\mathrm{min},k} = \min_{q \in [q_{\mathrm{min}}, q_{\mathrm{max}}]} \left[\frac{\sigma(q)\varphi_k}{I_k(q)} \right] \simeq \frac{\sigma(q_{\mathrm{m}})\varphi_k}{I_k(q_{\mathrm{m}})}. \quad (8)$$

For a given radius bin in the visualization (as defined in step 3 of §2), we can then estimate the minimum visibility threshold $\varphi_{\mathrm{min},R\mathrm{bin}}$ of that bin. This can be estimated as either the average of the $\varphi_{\mathrm{min},k}$ values of the components that fall into this bin or – to be on the safe side – the largest value of $\varphi_{\mathrm{min},k}$ in that bin. Likewise, if the visualization choice was of a number-weighted distribution following the procedure described above, the minimum number required to make a measurable impact can
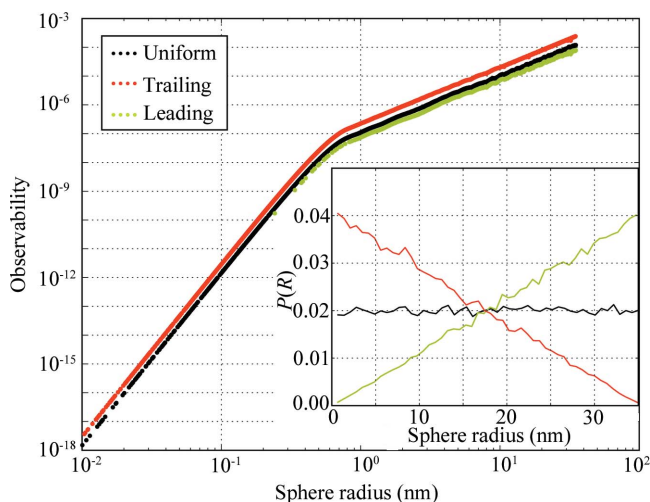
Figure 2
Observability for three unimodal number-weighted size distributions sampled using 50 000 spheres within $0.01 \leq q \leq 0.35$. The number frequency plot of the size distributions is shown in the inset, with the distributions divided over 50 radius bins and normalized to $1 = \int P(R)\,\mathrm{d}R$. A change in slope between R^6 and R^2 is observed for all three distributions at $R_{\mathrm{lim}} \simeq \pi/q_{\mathrm{max}}$.

be calculated from the minimum $\varphi_{min,k}$ through $n_{min,k} = \varphi_{min,k}V_{irr}/(\frac{4}{3}\pi R_k^3)$.

In this way, any size distribution visualization (*e.g.* number-weighted or volume-weighted size distributions) can contain a line indicating a rough estimate for the minimum detectability threshold (an example is shown in §7). Inclusion of this line can prevent the drawing of erroneous conclusions.

4. Data point weighting and convergence criterion

In most small-angle scattering measurements, the many data points I_p collected from each pixel p on the detector are reduced into a small number of q bins, I_{qbin}, before the data analysis procedures. In this reduction step, each measured data point collected between the bin edges q_n and q_{n+1} is averaged and assumed valid for the mean $\overline{q} = \langle q_n, q_{n+1}\rangle$, in a process known colloquially as 'radial integration' or 'radial averaging'. In other words, $I_{qbin}(\overline{q}) = \langle I_p(q_n \leq q \leq q_{n+1})\rangle$. Keeping track of the uncertainties during this process (and other steps in the data correction and reduction methodology) will result in the availability of uncertainties on the binned intensity data points.

These estimates of the level of uncertainty or 'errors' on each given data point are invaluable in assessing the veracity of model fitting results: to determine whether the analysis provided a solution to within the uncertainty estimate. Additionally, knowledge of the uncertainty can help unlink the model fitting result from more arbitrary parameters such as the number or spacing of the integration bin edges. By weighting of the goodness-of-fit parameter used in the least-squares minimization function by this error [*cf.* equation (4)] uncertainties on the MC solution can be established.

The counting-statistics-based Poisson error $\sigma = I^{1/2}$, where I is the number of detected counts in the raw data, gives the lower limit of the uncertainty estimate. These errors need to be appropriately propagated through the binning procedure and various data corrections such as dead time, dark current and background. Furthermore, if this estimate is exceeded by the sample standard error of the mean of the values contained in each individual q bin, the sample standard error should be the preferred error estimate for that bin as it can account for some detector irregularities. This sample standard error for each bin E_{qbin} is defined as

$$E_{qbin} = \left[\frac{1}{N_{qbin}-1}\sum_{p=1}^{N_{qbin}}(I_p - I_{qbin})^2\right]^{1/2}, \qquad (9)$$

where I_{qbin} is the mean intensity in the q bin, I_p the intensity of data point p in the bin and N_{qbin} the total number of data points in the bin.

Lastly, it is commonly challenging to get the absolute uncertainty of the measured intensity below 1%, even with the most elaborate data corrections in place, as shown by Hura *et al.* (2000), who achieved an overall uncertainty of about 2%. Thus, the uncertainty should be taken to be 1% of the measured intensity if this value exceeds the other two estimates. These errors can then be used in equation (4) to determine the goodness-of-fit parameter for an MC proposition.

The advantage of using these errors in the expression for χ_r^2 is that, if this parameter drops below one, the deviations between model and measured intensities are on average smaller than the statistical uncertainties. This thus provides a cutoff criterion for the MC method, allowing for the estimation of the mean and standard deviations of the final particle size histogram (as shown in §7). Additionally, by using these errors in the expression for the goodness-of-fit parameter, the intensities are weighted by their relative errors in the fitting procedures and thus become less sensitive to arbitrary values such as bin widths or number of data points used in the fit.

Since the MC method does not provide us with an intensity at the same level as the measured intensity, and as there often is a constant background associated with small-angle scattering patterns as described by Ruland (1971) and Koberstein *et al.* (1980), these two parameters will have to be determined separately. Thus, after every MC proposed change, but before calculation of the goodness of fit, an intermediate least-squares minimization routine is applied to optimize the model intensity scaling A and background parameter b [equation (1)]. As A does not affect the shape of the scattering pattern, the MC result remains unaffected. The addition of the background term may affect the presence of MC contributions with an effectively flat scattering profile, whose radii will be much below the radius limit imposed by the maximum measured q. If required, the least-squares minimization method may be expanded to include more terms, at the cost of speed and stability. One example of such an inclusion could be to include a power-law slope (with optional cutoff) to compensate for scattering from large structures or some interparticle scattering effects (Pedersen, 1994; Beaucage, 1995, 1996).

5. Uncertainties on the resulting distribution

One common criticism of MC methods is the potential for ambiguity in the result, *i.e.* the possibility to arrive at any number of equally valid but unrelated solutions. This MC implementation addresses that issue in two ways, firstly by limiting the number of sphere contributions through association with data quality, and secondly through determination of the variance between several tens of independent solutions, each time optimizing until $\chi_r^2 < 1$ has been achieved. As it is difficult to compare and visualize sets of radii, each resulting set is converted using a histogramming procedure to a density map or a volume-weighted size distribution. The variance between the histograms of the independent solutions will provide a mean and standard deviation for each histogram bin.

The visualization of the resulting data set in a histogram can be fine-tuned to some degree by adjustment of the number and spacing of the histogram radius bins. A greater number of radius bins (N_b) will result in more detail at the cost of larger standard deviations on the bin values. A good estimate for N_b can be found by means of the sampling theorem $r_{res} = \pi/q_{max}$, $N_b \simeq q_{max}/q_{min}$, assuming the largest measurable dimension is identical to the measurement limit (Hansen & Pedersen, 1991; Moore, 1980; Taupin & Luzzati, 1982). However, N_b also has to be dependent on the uncertainty of the underlying data set:

high-quality data with small uncertainties will result in smaller standard deviations on the visualized result, indicating that N_b can be doubled for more detail. A lot of noise or a large uncertainty in the measured data, on the other hand, will result in a large standard deviation on the histogram, indicating that N_b should be halved. While none of this affects the result of the MC optimization, the balance between detail and accuracy of the result visualization cannot be strictly defined. Lastly, the spacing between the bin edges can be adjusted in the visualization, which can be necessary if the result spans several decades in size (for which logarithmically spaced bin edges can be applied) or has sharp features [for which a natural binning has been suggested by Ilavsky & Jemian (2009)]. It is important to emphasize that none of these adjustments to the spacing or number of bins affect the MC result, and that there is an inherent link between the standard deviation and the choice of visualization parameters.

It should be noted, however, that while the standard deviations on the histogram are linked to the data uncertainty, there remains room for improvement as the standard deviation (but not the mean) on the resulting histogram can be affected slightly by the choice of p_c (especially when set to $p_c = 0$ or $p_c = 6$, as shown in the supplementary information). Despite this, the procedure does provide the user with a reasonable estimate of the standard deviation, which can be coupled with the observability information to obtain a size distribution with statistics.

6. Experimental

6.1. Synthesis

Boehmite (AlOOH) particles were synthesized *in situ* using an automated and modified version of a high-pressure high-temperature reactor (Becker *et al.*, 2010). The sapphire capillary in which the reaction takes place has an inner diameter of 1.0 mm and an outer diameter of 1.57 mm. The particles were synthesized from a solution of 0.5 M Al(NO$_3$)$_3$ precursor in water. The start of the reaction was considered to be the moment at which the pressurized solution [maintained at a pressure of 250 bar (25 MPa)] is heated to its reaction temperature of 548 K. The measurement used in this paper was obtained 1700 s from the start of the reaction. Further details and results will be presented in a forthcoming paper.

6.2. Beamline details

Synchrotron SAXS experiments were performed at the BL45XU beamline of the SPring-8 synchrotron in Japan. The beam was collimated to a 0.4 × 0.2 mm beam (horizontal by vertical, respectively), with photons with a wavelength of 0.09 nm. The sample-to-detector distance was 2.59 m. The scattering patterns were recorded on a Pilatus 300k detector whose total surface area covers 33.5 × 254 mm, consisting of 195 × 1475 pixels measuring 0.172 × 0.172 mm. Transmission values were determined using in-line ionization chambers. The polarization factor was assumed to be 0.95. The measurements were collected at a rate of 1 Hz.

6.3. Data correction

The data were corrected for background (water at 548 K and 250 bar), incoming flux, measurement time, transmission, polarization and spherical correction and calibrated to absolute units using a glassy carbon sample from series H, supplied by Dr J. Ilavsky from the Advanced Photon Source (Zhang *et al.*, 2009). Statistics were calculated according to the procedure outlined in §4, with the minimum possible error set to 1% of the measured intensity. The scattering contrast $\Delta \rho^2$ is set to 2.713×10^{30} m^{-4}.

7. Results and discussion

The collected data set, consisting of 200 data points, the errors and the MC fit are shown in Fig. 3, where the MC fit intensity is the average of 100 repetitions of the MC procedure. While not as fast in its current implementation as some of the alternative ITMs, the calculation of all 100 repetitions takes about 30 s on a single core of an Intel Core i7 processor running at a clock speed of 1.8 GHz. This performance can be improved by an estimated factor of 50 through compilation of selected parts of the code, as well as exploitation of multicore processing. Given that the speed is not yet a limiting factor in the overall methodology, no steps have been made in this direction.

While a single run also delivers a model intensity on average to within the determined error, the mean intensity is shown here as it matches the mean of the size distributions shown in Fig. 4. The error bars indicate the standard deviation of the histogrammed values for each of 100 repetitions. This figure also contains the estimate of the minimum visibility threshold $\varphi_{\min, Rbin}$.

The volume-weighted size distribution shows a distinctly triangular size distribution between 5 and 30 nm, with the distribution maximum at 10 nm. The total volume fraction of measurable precipitates is 0.88%, which is close to, but not
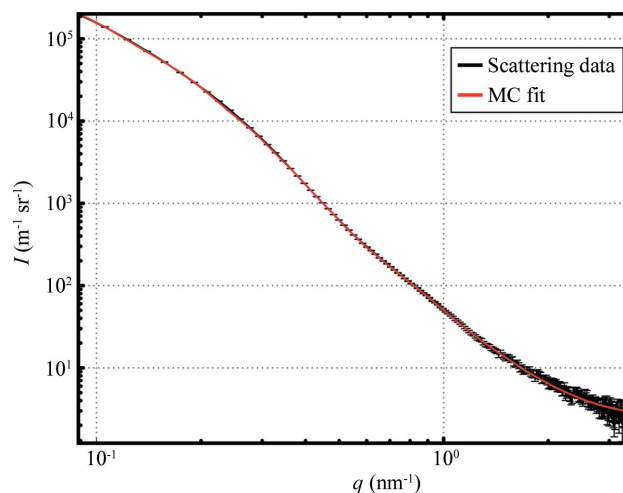
Figure 3
Data (black) and MC fit [grey (red in the electronic version of the journal), using $n_s = 300$ and $p_c = 3$] for a 1 s measurement in a time series of AlOOH nanoparticles in aqueous solution. The MC fit is at convergence. The background value resulting from the optimization is 2.4 (2) m^{-1} sr^{-1}.
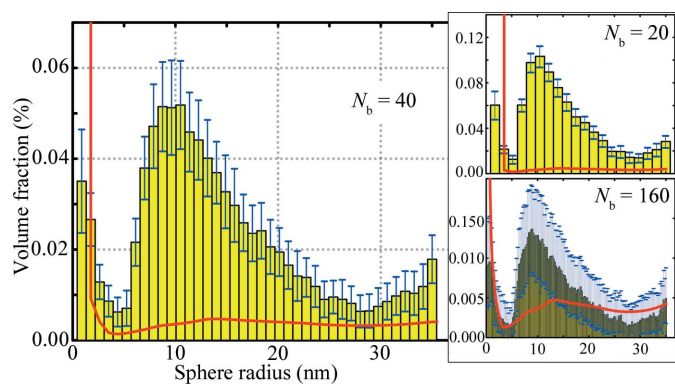
Figure 4
Volume-weighted size distribution as used for the MC fit shown in Fig. 3, for $N_b = 40$. Error bars indicate sample standard deviation over 100 repetitions. The minimum visibility threshold $\varphi_{\min,R\mathrm{bin}}$ is shown as a thick grey (red in the electronic version of the journal) line. The right-hand figures show the effect of selecting $N_b = 20$ and $N_b = 160$, with a clear effect on $\varphi_{\min,R\mathrm{bin}}$ as well as the uncertainty estimates.

quite, the 0.99% expected for full conversion of the precursor. This deviation may be attributed to large uncertainties on some of the values used in the data reduction (*e.g.* transmission factor, capillary diameter), partial precipitation of the nanoparticles at the bottom of the capillary or even precipitates growing beyond the coherence limit (which then no longer contribute to the scattering pattern).

In this example, the number of histogram bins is 40, the recommended number of divisions dictated by the sampling theorem, but one can choose more or fewer bins. The effect of this is shown in the subplots in Fig. 4 for 20 and 160 histogram bins, which clearly show the relation between the standard deviation in the bins and the minimum visibility threshold $\varphi_{\min,R\mathrm{bin}}$. If the number of histogram bins is high, the uncertainty for each value is equally large and the minimum visibility threshold increases. If the number of bins is reduced, both the uncertainty and $\varphi_{\min,R\mathrm{bin}}$ reduce, at the cost of a loss of detail.

8. A final comment

All the above results were obtained assuming that the scatterers are spherical in shape. This does not have to be the correct particle shape for the method to arrive at a solution, as the size distribution and shape cannot both be uniquely distinguished from scattering patterns alone (which has been tested for simulated isotropic scattering patterns from polydisperse sets of prolate and oblate ellipsoids, as shown in the supporting information). The solution from the MC method, then, shows the user what the size distribution would be *if* the scattering pattern originated from spherical particles.

If the shape of the scatterers is known from other investigations such as electron microscopy, and deviates from a spherical shape, this information can be used to obtain the correct size distribution for that particular shape (Pedersen *et al.*, 1996). This can be done either by adjusting the particular scattering function in the MC method [as recently shown for rod-like precipitates in MgZn alloys by Rosalie & Pauw

(2012)] or by analysis of the shape-independent correlation function $\gamma(r)$, which can be calculated from the sphere-based MC method result (Feigin & Svergun, 1987).

9. Conclusions

Discussed in this paper are modifications to the Martelli MC method, the general veracity of the result and the application of it to a SAXS measurement. It is shown that by using the methodology described in this paper a particle size distribution can be retrieved from a scattering pattern, uncertainties can be estimated for the particle size distribution and the minimum number of particles required to make a measurable impact on the scattering pattern (the minimum visibility threshold) can be indicated for each feature in the distribution.

The MC code is available for inspection, improvements and application under a Creative Commons Attribution-Share-Alike license and the latest copy will be freely supplied by the author upon request.

References

Beaucage, G. (1995). *J. Appl. Cryst.* **28**, 717–728.
Beaucage, G. (1996). *J. Appl. Cryst.* **29**, 134–146.
Becker, J., Bremholm, M., Tyrsted, C., Pauw, B., Jensen, K. M. Ø., Eltzholt, J., Christensen, M. & Iversen, B. B. (2010). *J. Appl. Cryst.* **43**, 729–736.
Botet, R. & Cabane, B. (2012). *J. Appl. Cryst.* **45**, 406–416.
Di Nunzio, P. E., Martelli, S. & Ricci Bitti, R. (2004). *J. Disper. Sci. Technol.* **25**, 491–501.
Fedorova, I. S. & Schmidt, P. W. (1978). *J. Appl. Cryst.* **11**, 405–411.
Feigin, L. A. & Svergun, D. I. (1987). *Structure Analysis by Small-Angle X-ray and Neutron Scattering*, edited by G. W. Taylor. Princeton: Plenum Press.
Glatter, O. (1977). *J. Appl. Cryst.* **10**, 415–421.
Glatter, O. (1979). *J. Appl. Cryst.* **12**, 166–175.
Hansen, S. (2000). *J. Appl. Cryst.* **33**, 1415–1421.
Hansen, S. & Pedersen, J. S. (1991). *J. Appl. Cryst.* **24**, 541–548.
Hura, G., Sorenson, J. M., Glaeser, R. M. & Head-Gordon, T. (2000). *J. Chem. Phys.* **113**, 9140–9148.
Ilavsky, J. & Jemian, P. R. (2009). *J. Appl. Cryst.* **42**, 347–353.
Koberstein, J. T., Morra, B. & Stein, R. S. (1980). *J. Appl. Cryst.* **13**, 34–45.
Krauthauser, H. G. (1994). *Physica A*, **211**, 317–326.
Krauthäuser, H. G., Lennartz, W. & Nimtz, G. (1996). *J. Appl. Cryst.* **29**, 7–15.
Martelli, S. & Di Nunzio, P. E. (2002). *Part. Part. Syst. Charact.* **19**, 247–255.
Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.
Mulato, M. & Chambouleyron, I. (1996). *J. Appl. Cryst.* **29**, 29–36.
Pedersen, J. S. (1994). *J. Appl. Cryst.* **27**, 595–608.
Pedersen, J. S. (1997). *Adv. Colloid Interface Sci.* **70**, 171–210.
Pedersen, J. S., Horsewell, A. & Eldrup, M. (1996). *J. Phys. Condens. Mater.* **8**, 8431–8455.
Rosalie, J. M. & Pauw, B. R. (2012). arXiv:1210.5366 [cond-mat.mtrl-sci].
Ruland, W. (1971). *J. Appl. Cryst.* **4**, 70–73.
Svergun, D. I. (1991). *J. Appl. Cryst.* **24**, 485–492.
Taupin, D. & Luzzati, V. (1982). *J. Appl. Cryst.* **15**, 289–300.
Zhang, F., Ilavsky, J., Long, G. G., Quintana, J. P. G., Allen, A. J. & Jemian, P. R. (2009). *Metall. Mater. Trans. A*, **41**, 1151–1158.