Scattering techniques for nanoparticle analysis: classical curve fitting and Monte Carlo methods

vorgelegt von

Diplom-Ingenieur Ingo Breßler

geboren in Ludwigsfelde

Von der Fakultät II - Mathematik und Naturwissenschaften

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Ingenieurwissenschaften

Dr.-Ing.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Thorsten Ressler (Technische Universität Berlin)
1. Gutachter:	Prof. Dr. Michael Gradzielski (Technische Universität Berlin)
2. Gutachter:	Prof. Dr. Andreas F. Thünemann (Bundesanstalt für Materialforschung und -prüfung)

Tag der wissenschaftlichen Aussprache: 7. Dezember 2017

Berlin 2017

List of publications

Please note that parts of this thesis were already published in open-access articles which permit reuse in accordance to the Creative Commons Attribution (CC-BY) license[1].

Papers

[1] I. Breßler, B. R. Pauw, A. F. Thünemann, 'McDLS: Software for Modelling Dynamic Light Scattering Data', *(in preparation)* **2017**.

[2] I. Bressler, B. R. Pauw, A. F. Thünemann, 'McSAS: software for the retrieval of model parameter distributions from scattering patterns', *Journal of Applied Crystallography* **2015**, *48*, 962, DOI 10.1107/S1600576715007347.

[3] I. Breßler, J. Kohlbrecher, A. F. Thünemann, 'SASfit: a tool for small-angle scattering data analysis using a library of analytical expressions', *Journal of Applied Crystallography* **2015**, *48*, 1587, DOI 10.1107/S1600576715016544.

Talks

[1] I. Breßler, 'SASfit Plug-in Tutorial', Workshop, ESRF, Grenoble, **Sept. 7, 2016**.

[2] I. Breßler, 'Software for small-angle scattering analysis: SASfit and McSAS – specialized solutions for particular problems', PSCM Soft Matter Colloquium, ESRF, Grenoble, **Sept. 6, 2016**.

[3] I. Breßler, 'SASfit: A comprehensive tool for small-angle scattering data analysis', Poster, 16th International Conference on Small Angle Scattering, Berlin, **2015**.

[4] I. Breßler, 'Customized SAS analysis – Writing plug-ins for SASfit', Workshop, TU Berlin, **May 28, 2014**.

[1] http://creativecommons.org/licenses/by/2.0/uk/legalcode

Table of Contents

Abstract

This work presents three program packages implementing different methods for nanoparticle analysis, focusing on the quantitative determination of particle size distributions required in standardization of nanoparticle reference materials. In part one, classical curve fitting is discussed. It finds optimal parameters of a scattering model composed of analytical expressions. The classical curve fitting toolbox is represented by *SASfit*: it is one of the mature programs for small-angle scattering data analysis and has been available for many years. Improvements to the basic data processing and analysis workflow are developed. They include (i) advanced algorithms for reduction of oversampled data sets, (ii) improved confidence assessment in the optimised model parameters and (iii) a flexible plug-in system for custom user-provided models. A scattering function of a mass fractal model of branched polymers in solution is provided as an example for implementing a plug-in. Alternatively to classical curve fitting, part two develops a user-friendly open-source Monte Carlo regression package (*McSAS*). It structures the analysis of small-angle scattering (SAS) using uncorrelated scattering contributions without any assumptions on the expected nanoparticle parameter distribution. Most importantly, the form-free Monte Carlo nature of *McSAS* means, it is not necessary to provide further restrictions on the mathematical form of the parameter distribution; without prior knowledge, *McSAS* is able to extract complex multimodal or odd-shaped parameter distributions from SAS data. Finally, part three adapts the Monte Carlo regression method for analysing multiangle dynamic light scattering (DLS) data and develops the program package *McDLS* to overcome limitations of existing methods at reliably determining multimodal size distributions. The reliability of the method is tested on simulated and experimentally measured DLS data of monomodal and multimodal particle ensembles.

Zusammenfassung

Diese Arbeit diskutiert drei Programmpakete, die verschiedene Verfahren zur Analyse von Nanopartikel-Ensembles implementieren, mit besonderem Fokus auf die quantitative Bestimmung von Größenverteilungen von Nanopartikeln, die besonders bei der Standardisierung von Referenzmaterialien von Interesse sind. Teil eines diskutiert die klassische Kurvenanpassung, welche die optimalen Parameter eines Streumodells bestimmt, das aus analytischen Ausdrücken zusammengesetzt ist. Die klassische Toolbox zur Kurvenanpassung wird durch *SASfit* repräsentiert: Es ist eines der ausgereiften Programme zur Analyse von Messdaten der Kleinwinkelstreuung und seit mehreren Jahren verfügbar. Verbesserungen in der grundlegenden Datenverarbeitung und im Arbeitsablauf der Datenanalyse wurden entwickelt. Sie umfassen (i) fortgeschrittene Algorithmen zur Reduktion über-abgetasteter Datensätze, (ii) eine verbesserte Beurteilung von Vertrauensbereichen optimierter Modellparameter und (iii) ein flexibles Plug-in-System für benutzerdefinierte Modelle. Eine Streufunktion eines Massenfraktal-Modells für verzweigte Polymere in Lösung wird als Beispiel für die Implementierung eines Plug-ins vorgestellt. Alternativ zur klassischen Kurvenanpassung, entwickelt Teil zwei ein benutzerfreundliches, quelloffenes Programmpaket für die Monte-Carlo-Regression (*McSAS*). Es strukturiert die Analyse von Kleinwinkelstreudaten unter Verwendung unkorrelierter Streubeiträge ohne weitere Annahmen über die erwartete Nanopartikel-Parameterverteilung. Insbesondere das formfreie Monte-Carlo-Verfahren von *McSAS* bedeutet, dass es nicht notwendig ist, weitere Bedingungen an die mathematische Form der Parameterverteilung zu stellen; ohne weitere Vorinformationen ist *McSAS* in der Lage, komplexe multimodale oder seltsam geformte Parameterverteilungen aus den SAS-Daten abzuleiten. Teil drei entwickelt schließlich eine Monte-Carlo-Regressions-Methode zur Analyse von mehrwinkel-Messungen mit dynamischer Lichtstreuung (DLS) und zielt mit dem Programmpaket *McDLS* darauf ab, die Einschränkungen existierender Verfahren bei der Bestimmung multimodaler Parameterverteilungen zu überwinden. Die Zuverlässigkeit des Verfahrens wird mit simulierten und experimentell gemessenen Daten monomodaler sowie multimodaler Partikel-Ensembles überprüft.

1

Introduction

1.1. Motivation

Quantification of nanoscale structures is set to become a requirement in industrial preparation of materials (Potočnik, 2011). In order to safeguard consumers and define industry standards it is still a challenge how to determine sizes of nanoparticles precisely, accurately and reliably while maintaining moderate costs. It is an open question for the legislator, crossing multiple disciplines, such as physical chemistry, analytical chemistry and nanotechnology. Therefore, a toolset is desired to obtain quantitative morphological parameter distributions of (size-)disperse nanoparticle mixtures with minimal effort, high flexibility, accuracy and high reliability.

There are many new methods as well as established methods which provide values for nanoparticle sizes but most of them have to be taken as nominal values (Tiede et al., 2008), because the specific numbers in nanometers are highly specific to the chosen methods and the devices used. For nanoparticle samples consisting of a very narrow size distribution approaching the Dirac delta function, various analysis methods often provide similar results (Borchert et al., 2005). Real-world nanoparticles in products and materials, however, often consist of a broad size distribution or even a multi-modal distribution (Mahl et al., 2011; Tsuji et al., 2006). With increasing size distribution width, the obtained morphological parameters are highly method-dependent. Especially determining size distributions of nanoparticles with sizes below 10nm represent a major challenge (Dieckmann et al., 2009).

That non-uniform sizes of nanoparticles pose a serious challenge for analytics was already formulated by Bragg in the 1960s (Bragg et al., 1960). Likewise, it was shown rather early as well, that uniform nanoparticles can be characterized with good consistency: For example, Ludox, a colloidal amorphous silicone dioxide, produced on a large scale with a narrow size distribution was measured consistently by electron microscopy and light scattering in the 1960s (Deželić et al., 1960).

It may also serve as an example to demonstrate the analytical problem of significantly diverging mean values for various mean values obtained by different methods. Because of its high availability and narrow size distribution, this particle was released in 2012 as

reference material ERM-FD304 ('Silica Nanoparticles in Aqueous Solution') by the European Commission Joint Research Centre Institute for Reference Materials and Measurement (IRMM) (Franks et al., 2012). Its certification report specifies the 'equivalent spherical diameter of the suspended particles' determined by dynamic light scattering (DLS): The modal Stokes diameter is given by 33.0(15) nm and the harmonic intensity-weighted arithmetic average particle diameter, as determined using the method of cumulants is given by 42.1(3) nm (Franks et al., 2012). Other techniques, such as electron microscopy (TEM/SEM), however, placed the number-based modal diameter at 27.80(75) nm.

Transmission electron microscopy (TEM) is one of the most commonly used technique for nanostructural quantification. It is essential in determining the overall morphology of the nanostructural features and can often be used to coarsely quantify their parameters. Obtaining a statistically representative quantification of the nanostructure, however, is reliant on the probing of large numbers of objects. To improve its representation of the ensemble of the sample, it should preferably be performed through sampling from multiple locations throughout a bulk-scale sample containing several thousand particles (Klein et al., 2011; Meli et al., 2012).

As TEM has remained largely resilient to automation efforts, this continues to be a tedious and labour-intensive task. Therefore, it might be beneficial to combine the localized superior spacial resolution of microscopy with another technique more suited for ensemble averaging nanostructural quantification such as small-angle scattering (SAS) (ISO, 2014; Pauw, Pedersen, et al., 2013) or dynamic light scattering (DLS) (ISO, 1996).

1.2. Scattering techniques

Small-angle scattering

Small-angle scattering offers one reliable route to quantification of nanomaterials: it can characterize the nanostructure of large amounts of material with a minimum of tedium, for example, easily extracting size distributions and volume fractions. There are a variety of approaches to obtain such structural parameters from the data, including model-free analysis, model fitting and inversion methods, several of which have been implemented in available software. A few well-established programs for model-based SAS data analysis are actively maintained. *IRENA*, for example, is suitable for a wide range of sample types (Ilavsky and Jemian, 2009). *Scatter* is a program geared towards the analysis of two-dimensional data from nano- and mesoscale oriented structures (Förster et al., 2010). *SasView* evolved from a focus on small-angle neutron scattering (SANS) to a solution for small-angle scattering (SAS) data analysis at large scattering facilities (Doucet et al., 2017). Furthermore, there is the *ATSAS* project, consisting of a comprehensive set of sophisticated

tools primarily intended for biological systems encountered in protein analytics (Petoukhov et al., 2012) as well as the *ScÅtter* program (Rambo, 2017) with a similar focus. For this special topic around protein analytics, please refer to other recent work by Trewhella (2016) or Svergun, Koch, et al. (2013). Independent of the field of application, every data analysis tool relies on the quality of the input data provided.

One of the biggest stumbling blocks in the application of SAS has been the data correction and analysis. Although the discussion of data corrections is beyond the scope of this work [see Jacques et al. (2012), Pauw, Pedersen, et al. (2013) and Kieffer and Karkoulis (2013) for such discussions], it has to be stressed that correct analysis of data is reliant on the quality thereof. There can be no good results without proper data which, in turn, cannot be considered complete without reasonable uncertainty estimates on the data values.

Dynamic light scattering

Dynamic light scattering (DLS), also named photon correlation spectroscopy (PCS) or quasi-elastic light scattering (QELS), is a mature analytical technique for the determination of hydrodynamic radii of nanoparticles (Schärtl, 2007), micro emulsions (Khan et al., 2016), proteins (Gun'ko et al., 2003), polymers (W. Brown, 1993), etc.. The DLS method allows for a statistically representative quantification of nanoparticles while offering a number of advantages: short measurement times of less than a minute, easy-to-use instruments and highly automated data evaluation software accompanying commercial instruments and typically, the laser which illuminates the sample, is not accessible by the operator which makes it safe for use by untrained staff. A recent short overview of the method is given by Hassan et al. (2015) and a deeper insight is found in the classical work of Finsy (1994). Another recent review focusing on DLS basics and applications in nano pharmaceutics if given by Bhattacharjee (2016). Possible pitfalls of the method were explained by Fischer and M. Schmidt (2016). In short, the analysis of particle sizes from DLS is mathematically an ill-posed problem for which typically regularization techniques are applied to get meaningful results at all (Nyeo and Ansari, 2015). The underlying problem is generally found for exponential analysis of physical phenomena (Istratov and Vyvenko, 1999).

Nevertheless, various methods have been established for DLS data analysis over the years. Of utmost importance is the classical method of cumulants as derived by Koppel (1972) which is standardized for small particles of a narrow size distribution (ISO, 1996). It requires that the so-called polydispersity index, a measure for the broadness of the size distribution, has to be smaller than 0.1 which is rarely fulfilled. Therefore, the standard is often applied beyond its scope. For determination of broader and multimodal distributions, the CONTIN program is most widely spread (Provencher, 1982b). CONTIN is based

on an inverse Laplace transformation and uses a regularized non-negative least squares (NNLS) algorithm (Provencher, 1982a). Because the smoothness constraint in CONTIN is problematic with multimodal and narrow distributions, an alternative NNLS algorithm was proposed, employing a multiple-pass analysis of the same measurement (Morrison et al., 1985). Further, solutions for the determination of particle size distributions are based on the maximum entropy method (R. K. Bryan, 1990; Langowski and R. Bryan, 1991; Bryant and J. C. Thomas, 1995), singular value decomposition (Finsy, Groen, et al., 1989), stochastic methods, such as neural networks (Gugliotta et al., 2009) or particle swarm optimization (Zhu et al., 2011).

The main common drawbacks of the above stated methods are (a) a high sensitivity to noise, (b) difficulties in identifying correct peak positions for multimodal distributions with different intensities contributions of each particle class and (c) a lack of reasonable estimates of the associated uncertainties of the results. Furthermore, no method considers the variance from multiple measurements as given by *a priori* information.

1.3. Analysis methods

Classical curve fitting

After suitably corrected SAS data have been obtained, analysis thereof can be performed through a classical approach used in *SASfit*: using a least-squares optimization to match the measured data to a synthetic scattering pattern generated by a composition of models for the form factor, size or parameter distribution and an optional structure factor. All of which are defined by a handful of parameters (Pedersen, 1997).

Monte Carlo based curve fitting

The assumptions on both the scatterer shape and the mathematical form of the parameter distribution(s) in model functions of the classical approach are often insufficiently flexible to describe the morphology of many samples. Good agreement between the model function and the measured data will then not be achieved, in particular for samples where the actual dispersity does not adhere to the inherently assumed model parameter distribution form (such as lognormal, Gaussian or Schultz-Zimm), or where such a distribution form is not known or can not be assumed *a priori*.

Modern analysis methods are available for this class of samples which allow for the retrieval of model parameter distributions without assumptions on the form of the distribution. While the general shape of the scatterer still has to be defined in order to arrive at a unique solution (see, for example, Rosalie and Pauw (2014)), the methods are no longer restricted to a limited set of model parameter distribution forms. Such modern methods include Titchmarsh (Fedorova and P. W. Schmidt, 1978; Botet and Cabane,

2012) or indirect Fourier transforms, based either on smoothness criteria (Glatter, 1977; Svergun, 1991) maximum entropy optimization (Hansen and Pedersen, 1991) or Bayesian hyperparameter estimation (Hansen, 2000). While these carry a certain mathematical elegance, they can be challenging to implement, understand and apply. This mathematical obscurity furthermore hinders thorough understanding of the failure modes, which can lead to crucial errors in their application.

For determining model parameter distributions from small-angle scattering patterns a conceptually straightforward Monte Carlo-based method was presented by Pauw, Pedersen, et al. (2013). It has since been applied to explore the size distributions of a variety of samples including metal alloys (Oba et al., 2015; Rosalie and Pauw, 2014), novel oxygen reduction reaction catalysts (Schnepp et al., 2013), polymer fibres (Pauw, Ohnuma, et al., 2013), plasmoids (Meir et al., 2013), iron oxide nanoparticles (Lak et al., 2015) and quantum dots (Schindler et al., 2015; Abécassis et al., 2015). While these results have been encouraging, the lack of user friendliness of the method has hindered its adoption by a broader audience.

Through a multinational collaborative effort spanning several years, a drastic improvement on the software usability was effected. It became useable for non-programmers conducting scientific analysis on a multitude of measurements on a daily basis. Additional functionality was added to cope with various levels of data quality as well. These include basic data reduction mechanisms to crop and bin the data (logarithmically) ahead of the analysis as well as providing comprehensive information on the parameter distributions determined: number-, volume-, surface- or intensity-weighted histograms on user-defined parameter ranges, to support a wide range of applications. Last but not least, a structured approach was developed to store the results in a common portable scientific data format (*HDF5*), together with their source data and the program configuration leading to them, which is important to assure the quality and reproducibility of scientific work. This was mostly accomplished through a comprehensive rewrite of the implementation by the author of this work: following modern coding standards and conventions by harnessing the potential of proven and mature program libraries such as *NumPy*, *SciPy* and *Matplotlib*, utilizing a current framework for a productive and self-descriptive user interface (*Qt*), employing a distributed version control system (*Git*) and establishing a framework for regression testing of the integral parts to maintain a consistent software quality (*nose* testing).

Consequently, it is of interest to see if this Monte Carlo method can analyse DLS data equally well. Repeated DLS measurements provide uncertainties which can be used for determining form-free size distributions accompanied with uncertainties for individual sizes. This would enhance the information value offered by DLS measurements, as that measurement uncertainties of DLS data are not yet commonly found in the literature.

The results of this work are structured in three parts: Chapter two presents the classical curve fitting toolbox *SASfit*. The Monte Carlo-based approach implemented by the *McSAS* program is detailed in chapter three while chapter four describes its transfer to DLS analysis implemented in the program *McDLS*.

2

SASfit: Classical curve fitting of SAS data

SASfit is a program primarily for model-based analysis of SAS data, which contains over 200 models for fitting and can be extended by the user to include more. The fitting interface also allows for the construction of compound models which can then be used to fit one or more data sets. Examples of the application of this program include the traceable size determination of gold nanoparticles (Meli et al., 2012), polymeric nanoparticles (Gleber et al., 2010) and vesicles (Varga et al., 2014) as well as inter-laboratory comparison of nanoparticle size distribution quantification (Pauw, Kästner, et al., 2017). Furthermore, it was used for studying the mechanism of gold nanotriangle formation (Liebig et al., 2016), characterizing ultra-small core-shell silver nanoparticles (Kästner and Thünemann, 2016), examining the asphaltene nanostructure using complementary SANS and SAXS scattering (Eyssautier et al., 2011), analysing superparamagnetic iron oxide nanoparticles (SPION) (Szczerba et al., 2017) and their stabilization by polyelectrolytes (Woźniak et al., 2017).

Moreover, the program was employed to characterise polymer blends (Kogikoski et al., 2017), comb-shaped copolymers (Gelardi et al., 2017), star block copolymers (Herfurth et al., 2016), the behaviour of dynamic covalent polymers (Fuhrmann et al., 2016) and copolymer based micellar solutions (Nguyen-Kim et al., 2016).

In colloid and interface science, it is often utilized for studying micellar structures and their behaviour, such as the self-assembly of lipopeptide in water (Soares et al., 2017), self-assembled peptide nanotubes (Hamley et al., 2017; Castelletto et al., 2017), sodium hyaluronate in oppositely charged surfactant solution (Buchold et al., 2017), spherical micelles (Kirkham et al., 2016; Jin et al., 2016) and bicelles (Isabettini et al., 2016) as well as analysing synthetic virions (Noble et al., 2016) and the assembly of nanoparticles-polyelectrolyte complexes in nanofiber cellulose (Garusinghe et al., 2017).

Likewise, *SASfit* is applied to investigate the properties of emulsions, such as 'the effect of hydrostatic pressure on the structure of a bicontinuous microemulsion in the presence of a solid interface' (Berghaus et al., 2016) or for studying the flocculation process in emulsions (Huck-Iriart et al., 2016). Also solids are regularly analysed, for example, in a 'multiscale description of shale pore systems' (Leu et al., 2016) and for developing Co-Re alloys for ultra-high-temperature applications (Gilles et al., 2016).

More exotic applications include the study of relaxation mechanisms in magnetic colloids by stroboscopic spin-polarized SANS (Wiedenmann et al., 2011), detection of toroidal supramolecular polymers using simultaneous SAXS and SANS analysis (Hollamby et al., 2016), simultaneous fitting of up to 70 measurements from contrast variation experiments (Kohlbrecher, Buitenhuis, et al., 2006; Vavrin et al., 2009) and analysing the (model-free) integral structural parameters of 400 scattering curves for creating local contrast in SANS by dynamic nuclear polarization (Brandt et al., 2007).

2.1. Data preparation

Before data fitting, it is wise to investigate the data plot and its uncertainty estimates ('errors'). At high q, the intensity tends to be low with a high degree of uncertainty. Accordingly, high-q data of radially averaged one-dimensional data tend to contribute little to the overall goodness of fit, defined by the reduced chi-squared (χ^2_r) value (Bevington and Robinson, 2003):

$$\chi^2_\mathrm{r} = \frac{1}{(N-M)} \sum_{i=1}^{N} \left[\frac{I_\mathrm{exp}(q_i) - I_\mathrm{mod}(q_i)}{\sigma_\mathrm{exp}(q_i)} \right]^2 \qquad [2.1]$$

When observing high uncertainties on a large number of data points, a data averaging step should be performed on these data points to prevent instabilities and large calculation times. For quick evaluation, however, it is an option to initially ignore a section of points in the high-q region. Furthermore, it might be necessary to skip invalid data points, which may originate from, for example, improper masking of the beam stop, edge effects in radial integration or parasitic scattering contributions.

For these reasons, additional options are provided during data import, which allow the user to specify the q range for each data file individually when merging data files. For final data analysis it is recommended, however, to bin the data properly to reduce the number of data points and to improve the data and the data statistics. The adaptive averaging method presented here can be used for this purpose.

An important step in preparing data for analysis is to reduce the number of data points of oversampled data sets. A reduced number of data points will increase fitting speed proportionally and improve fitting stability.

SASfit supports three methods to reduce the number of data points of oversampled data sets of which the author implemented the first and second method in the course of this work:

Method one is a simple and straightforward thinning out of data points according to

$$I_{\mathrm{red},i} = I_{\mathrm{raw},k(i)} \text{ with } k(i) = \mathrm{round}(i/\mathrm{ratio}). \qquad [2.2]$$

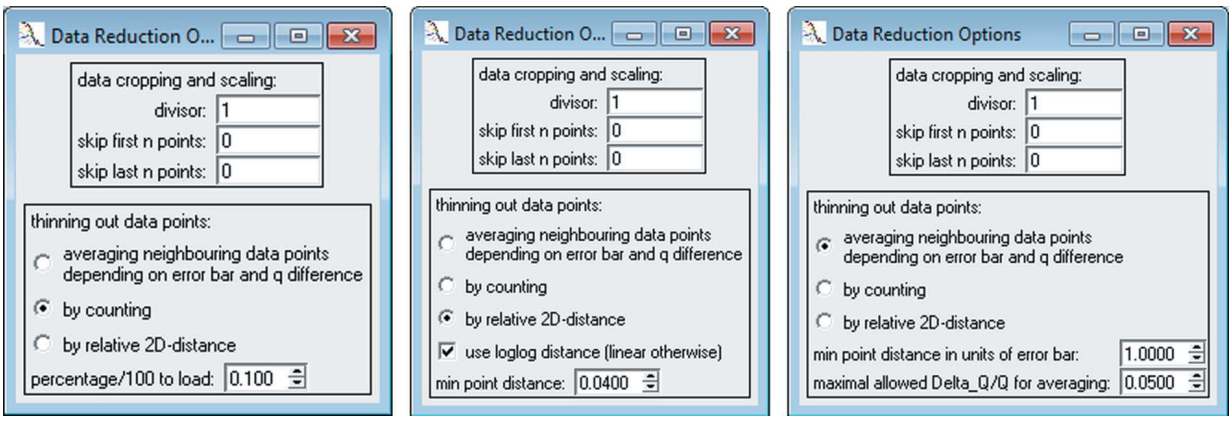

Figure 2.1: Data reduction window, providing three different methods for reducing the number of data points. Method one (leftmost figure) skips data points by a count ratio (see eqn. [2.2]). For fast data fitting of 10^3 data points, a typical fraction of 0.1 is recommended. Method two (middle figure) is a fast method for data with distinctive curve features (see eqn. [2.3]). Method three (right-hand figure) averages data points adaptively according to intensity and q spread (see eqn. [2.6]).

The user specifies a ratio of the original data points to keep. For example, 90% of the points can be omitted in the fit by specifying 0.1, as shown in the left-hand panel of Fig. 2.1. This rough method is suitable when working with high-density SAXS data of several thousand data points and when very fast fitting results are desired to get an initial overview, for example during an experiment.

The second method preserves scattering curve characteristics better than the first. It maintains a user-defined distance $\varphi_{\min}$ between data points by utilizing Pythagoras' theorem in linear or logarithmic two-dimensional space and skips those points which are less far away in accordance with

$$I_{\mathrm{red},i} = I_{\mathrm{raw},k(i)} \text{ with } \left[\delta_I(i)^2 + \delta_Q(i)^2\right]^{1/2} > \delta_{\min}. \qquad [2.3]$$

In linear context, $\delta_I(i)$ and $\delta_Q(i)$ are calculated by

$$\begin{aligned} \delta_I(i) &= I_{\mathrm{raw},k(i)} - I_{\mathrm{raw},k(i-1)}, \\ \delta_Q(i) &= q_{\mathrm{raw},k(i)} - q_{\mathrm{raw},k(i-1)}, \end{aligned} \qquad [2.4]$$

and in logarithmic context by

$$\begin{aligned} \delta_I(i) &= log\left|I_{\mathrm{raw},k(i)}/I_{\mathrm{raw},k(i-1)}\right|, \\ \delta_Q(i) &= log\left|q_{\mathrm{raw},k(i)}/q_{\mathrm{raw},k(i-1)}\right|. \end{aligned} \qquad [2.5]$$

The third, and most recommended, method for data reduction averages neighbouring data points locally, on the basis of the difference in intensity and width in q space (see

right-hand panel of Fig. 2.1). Each local interval $(k; l]$ is determined adaptively so that it contains all points n which fulfil the following condition:

$$\forall n \in (k, l] : \frac{|I_k - I_n|}{\delta I_k - \delta I_n} < D_{\min} \wedge \frac{|q_k - q_n|}{\overline{q}} < \delta q_{\max}. \quad [2.6]$$

The parameter $D_{\min}$ restricts the intensity difference within an interval proportional to the associated uncertainties. Additionally, the maximum width of an interval relative to its position in q space is scaled by the second parameter, $\delta q_{\max}$. This results in a narrow spacing between data points at low q and a wide spacing at high q. Both conditions have to be fulfilled by neighbouring data points to fall within an interval and thus to allow calculation of an average. This last method retains information on sharp features, while averaging data points. It is important to note that the original data are always stored alongside the modified data, and are also stored for traceability in *SASfit* project files along with the reduced data. Adjusting the data affects the copy of the data used for numerical analysis only. The data selected for analysis can, therefore, be changed at any time.

2.2. Model fitting

The main purpose of *SASfit* is to fit a model described by idealized scattering functions to one or more data sets. By minimizing the goodness of fit criterion χ^2 (see eqn. [2.1]) through adjustment of the model parameters, the model intensity $I_{\text{mod}}(q)$ is matched with the measured intensity $I_{\text{exp}}(q)$. *SASfit* is designed to let the user configure every aspect of $I_{\text{mod}}(q)$ defined as

$$I_{\text{mod}}(q) = \sum_{c}^{N} \left\{ \left[\int_{a}^{b} P(q, x) f(x) \, \mathrm{d}x \right] S(q) \right\}. \quad [2.7]$$

The model intensity is based on the sum of the scattering intensity contributed by different scatterers in the analysed sample. Each scattering contribution c consists of a form factor $P(q, x)$ determining the shape of a scatterer. Disperse aspects of shape-similar contributions can be reflected by applying a parameter distribution $f(x)$, in the range $[a, b]$, to an arbitrary form factor parameter x. Furthermore, *SASfit* allows defining a structure factor $S(q)$ for each contribution, reflecting attraction and repulsion of scatterers in the sample. This model composed of several scattering contributions can be managed in the graphical user interface of the program.

There are over 200 form factors available in *SASfit*. This list includes the commonly used 'Sphere' and associated form factors, but also includes models that do not strictly adhere to the form factor definition, such as the Beaucage unified fit model (Sztucki et al., 2007) and several models for disordered structures. One typical starting model for fitting

data would be the 'Sphere' model, coupled with a parameter distribution over the sphere radius R. By starting with such a simple model, alternative and typically more complex models can be assessed for their significance against the fit quality of the 'Sphere' model. This example is given in Fig. 2.2 by using a Gaussian distribution.

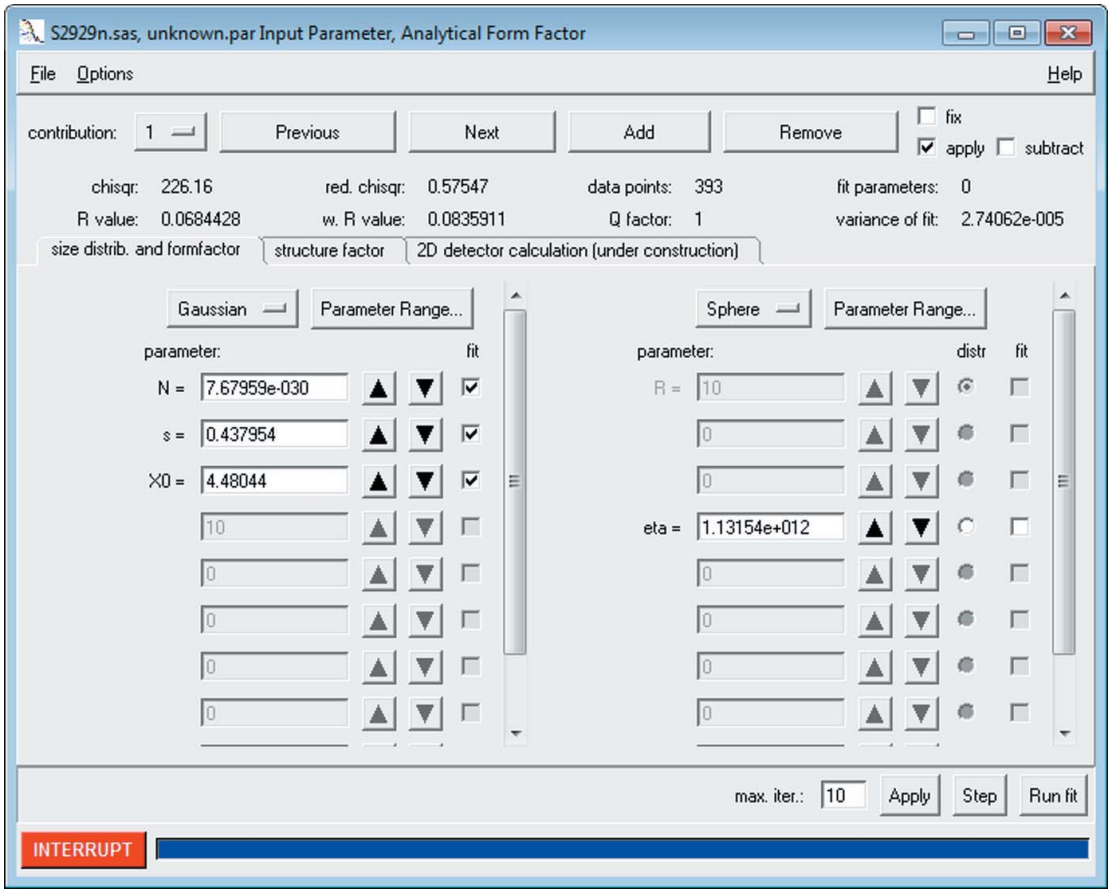

Figure 2.2: Basic model configuration, consisting of a form factor on the right (in this panel it is a 'Sphere' with the scattering contrast of gold nanoparticles 'eta' $= 1.13 \times 10^{12}$) and a distribution of one selected parameter ('distr' column) on the left (here a 'Gaussian' distribution of the radius which has a concentration parameter of 'N' $= 7.6 \times 10^{-30}$, a width parameter 's' $= 0.43$ and the mean radius parameter 'X0' $= 4.48\,\mathrm{nm}$. A structure factor can be configured on the second tab. Different scattering contributions can be managed by the top row of buttons.

Depending on the scientific field, either the Gaussian, lognormal or Schulz-Zimm distributions are typical choices for describing the (poly)dispersity of monomodal shape parameter distributions.

Each distribution consists of at least one parameter controlling the position of its maximum and one parameter controlling its width or FWHM, which defines the degree of polydispersity. The monodisperse distribution, being the exception to this, defines only a single parameter value. Nontrivial parameter distributions are integrated by linear subdivision of the integration range, whose extent and granularity are determined adaptively on the basis of the user-provided distribution parameters. Finding a suitable integration range depends on the distribution functions and cannot be handled in an

efficient and numerically stable way for all possible functions. Therefore, a fixed set of available parameter distributions is defined in *SASfit*, which cannot (yet) be extended by the user.

Besides form factors and distribution functions, *SASfit* allows consideration of interparticle scattering effects. There are several approximations implemented to allow calculation of such a structure factor: (1) monodisperse approximation (see eqn. [2.7]), (2) decoupling approach, (3) local monodisperse approach, (4) partial structure factor and (5) scaling approximation of partial structure factor. It should be noted that some of these structure factor approximations [specifically, numbers (2) and (5)] require knowledge of the scattering amplitude of the form factor, as opposed to 'just' the scattering intensity. *SASfit* might return an error if such a structure factor approximation is attempted in combination with a form factor for which the scattering amplitude is not known. Details on the exact formulae are given in the *SASfit* manual (Kohlbrecher and Breßler, 2016).

Structure factors affect a scattering curve at low q values owing to the larger lengths they inherently cover. Here, the residuum of a fit would show oscillations if disregarded particle interactions are significant. In order to assess the influence of a structure factor on the fit of the model configuration and the experimental data, it can be selected and configured in the second tab, 'structure factor', of the model configuration window shown in Fig. 2.2. In many cases, a simple 'Hard Sphere' structure factor would suffice. In its basic configuration, the repulsion radius is set slightly larger than the particle radius but maintains the same order of magnitude, whereas its volume fraction is set to small values such as 0.05 at the start of the fitting procedure.

Before moving on to the least-squares fitting procedure, it is recommended to constrain the fit parameters algorithm to a physically feasible range of parameter values. They can be defined by the 'Parameter Range' menu next to each model function for single data set fitting. If no constraints are applied, the fitting procedure may result in no solution or an unrealistic local minimum. Each parameter contains a brief explanation of its meaning, which is shown in a tooltip and at the bottom of the window upon hovering the cursor over the parameter.

2.3. The usual curve fitting workflow

It is recommended to adjust the fit parameters manually before starting the iterative least-squares optimization, in order for the model intensity to approach or intersect with the data. This helps to prevent instability during the initial iterations. Owing to the minimization exhibiting many local minima it is strongly recommended to optimize the model parameters stepwise. Otherwise, a minimization of all parameters in one step automatically either will not reach the intended best fit or may provide a physically meaningless solution. Therefore,

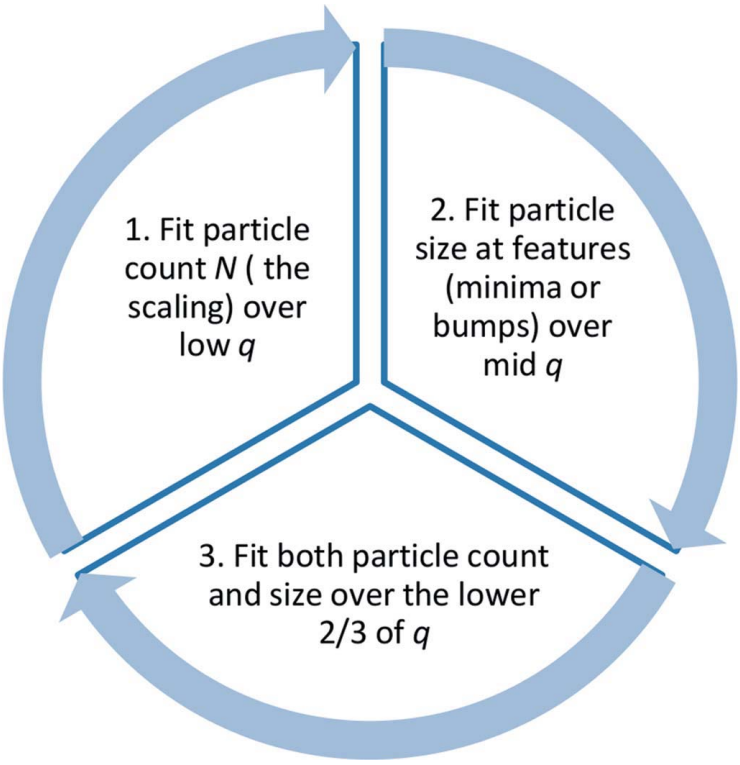

Figure 2.3: The recommended curve fitting workflow of *SASfit* consists of a three-step cycle.

the basic workflow for fitting small-angle scattering data consists of the following steps (as illustrated in Fig. 2.3):

1. The first step of a fit procedure is to match the order of magnitude of the model curve (red) and data intensity (black dots with blue error bars). This can be accomplished by initially fitting the scaling parameter at the beginning of the curve only. The distribution parameter N should be used for this purpose, which denotes the number of scattering objects involved in the measurement. Applying 'Run fit' finds the optimal value of N.

2. The size of scattering objects is optimized in the second step of curve fitting. The best results are obtained by limiting the fit to the central part of the data where, for example, a first local minimum of the curve can be observed.

3. Fitting both the scaling parameter and the size parameter at the same time over the first two-thirds of the data further improves the overall quality of the fit.

In a final step, some slight mismatch in the central part of the scattering curve can be optimized by fitting the particle radius together with the distribution width parameter 's' (width of the Gaussian distribution in the example). Steps 1-3 can be repeated, if necessary, until a good overlap of the model curve and the data is obtained.

With increasing value of the size distribution width parameter, the model curve becomes smoother since it represents a broader size range of scattering objects. Because

of this smoothing, evidence supporting a particular shape reduces, making it increasingly challenging to distinguish between different shapes coupled with large polydispersities (as many will fit the data to a similar degree). Therefore, care has to be taken in interpretation of broad size distributions which are larger than about 20% of the mean value at σ; external supporting evidence for a particular shape assumption should be provided.

2.4. Example 1: disperse gold nanoparticles

To aid in the discussion of the remaining *SASfit* aspects, such as fit interpretation and reliability, an application example will be shown first. By means of this example, necessary considerations for retrieving reliable values for scatterer population from absolute measurements and advice on the interpretation of uncertainties will be provided. In order to demonstrate the reliability of SAS data analysis with *SASfit* a dispersion of gold nanoparticles of the NIST reference material RM-8011 (De Temmerman et al., 2014; Small and Watters, 2015) was chosen.

A straightforward example in which the presented fit procedure is applied is the determination of the mean radius of spherical particles in solution, the width of their radius distribution and the particle number concentration. RM-8011 was measured for 30 min with a SAXSess instrument from Anton Paar and the data was scaled to absolute intensity using water as a primary standard as described by Orthaber et al. (2000). The resulting data have been fitted to a model composed of spheres with a Gaussian size distribution. The result is shown in Fig. 2.4. The obtained data set and a preconfigured *SASfit* project file (`GoldS2843.pdh` and `GoldS2843.sas`) is provided in the supporting information of the original work (Breßler et al., 2015).

The fit parameters in this example are values for the particle concentration N, the mean particle radius X_0 of the assumed Gaussian size distribution and the width of the size distribution s. Note that numerous size distributions are provided, including the frequently used Schulz-Zimm (Flory) and lognormal distributions. It is the users' choice to select the most appropriate one. It is recommended to commence with a Gaussian size distribution if no evidence is available to support an alternative size distribution, for example from other methods like electron microscopy or from physical considerations.

The best fit value for the mean radius in this example is $X_0 = 4.48(5)$ nm and the width of the size distribution is $s = 0.44(5)$ nm. Here, the uncertainties merely denote the standard errors as determined from the least-squares optimization method. These uncertainties can be utilized to determine the combined standard uncertainties from all input quantities (Meli et al., 2012); however, this can be a tedious procedure which is beyond the scope of this example. As a rule of thumb, the uncertainty of the size parameters from *SASfit* is of the same order of magnitude as the combined standard uncertainties.

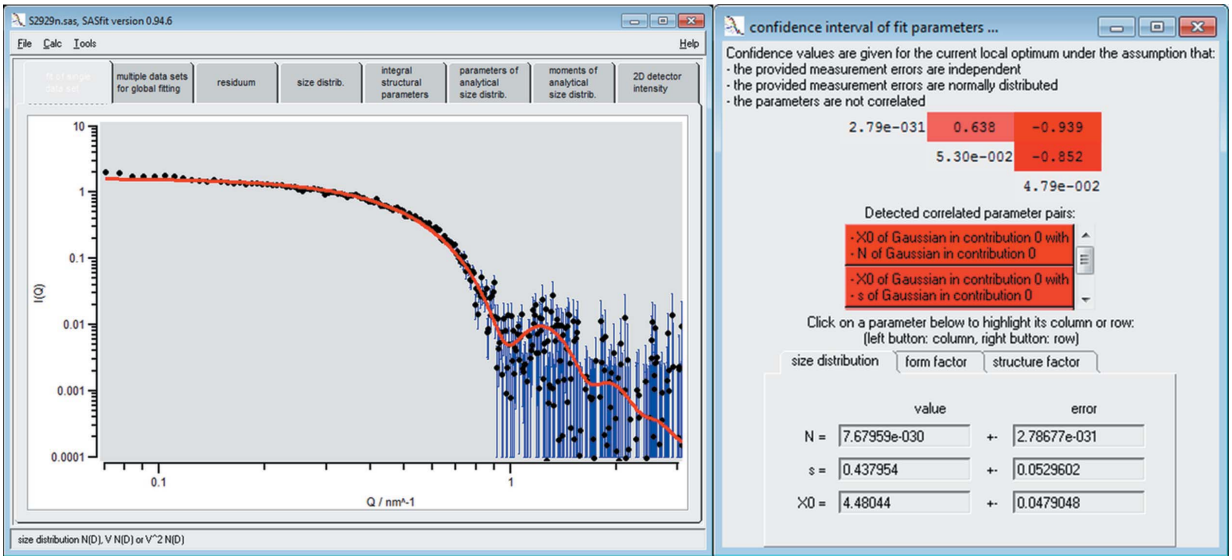

Figure 2.4: Curve fit (red solid curve) of a single data set (black dots) of gold nanoparticles of the NIST reference material RM-8011 (De Temmerman et al., 2014; Small and Watters, 2015) using a model of spheres with Gaussian size distribution and the initial parameters shown in Fig. 2.2. The uncertainties of the measured intensity values are displayed as vertical blue lines. The mean radius is $X_0 = 4.48(5)\,\mathrm{nm}$ and the width of the size distribution is $s = 0.44(5)\,\mathrm{nm}$. Right-hand figure: covariance matrix and confidence intervals of fit parameters to assess (inter-)correlation of parameters.

Another common calculation, possible when the data have been scaled to absolute units, is the determination of the concentration of scatterers in solution.

In order to convert the N value to a particle concentration (in number of particles per cm^3), the units used for absolute intensity, scattering vector and scattering length density should be considered. Here, this conversion factor is 10^{42} as the corresponding units used were cm^{-1}, nm^{-1} and cm^{-2}:

$$[N] = \frac{[I_{\mathrm{exp}}]}{[I_{\mathrm{mod}}]} = \frac{[I_{\mathrm{exp}}]}{[q]^{-6}[\Delta\eta]^2} = \frac{\mathrm{cm}^{-1}}{\mathrm{nm}^6\mathrm{cm}^{-4}} = \frac{\mathrm{cm}^{-1}}{10^{-42}\mathrm{cm}^6\mathrm{cm}^{-4}} = 10^{42}\mathrm{cm}^{-3} \qquad [2.8]$$

Therefore, in this example, the N value of $7.68(28) \times 10^{-30}$ corresponds to a particle number concentration of $7.68(28) \times 10^{12}\,\mathrm{cm}^{-3}$ or a molar concentration of $1.28(5) \times 10^{-7}\,\mathrm{mol\,l}^{-1}$. To double-check this value, it is recommended to convert the particle number concentration to the mass (and volume) fraction φ_{m} (and φ_{v}, respectively).

For a Gaussian size distribution the mass fraction is $\varphi_{\mathrm{m}} = N\varrho\langle V\rangle = N\varrho\frac{4}{3}\pi X_0^3\left[1 + 3\,(s/X_0)^3\right]$, where ϱ is the density and $\langle V\rangle$ the mean particle volume. Assuming that the gold particles in this example have the same density as the bulk material ($19.30\,\mathrm{g\,cm}^{-3}$), a mass fraction of $57.37(209)\,\mu\mathrm{g\,g}^{-1}$ was calculated. This value is in reasonable agreement with the value of $51.56(12)\,\mu\mathrm{g\,g}^{-1}$ (2.67×10^{-4} vol.%) provided by NIST, as determined by inductively coupled plasma optical emission spectrometry (Small and Watters, 2015). It should be noted that the uncertainty of the intensity measurements is of the order of 5%, which means that the

uncertainty of N must be of the same order of magnitude, in other words larger than the least-squares derived uncertainty (Orthaber et al., 2000). *SASfit* furthermore provides several measures which can be used to assess the quality of the (local) optimization minima found, which will be discussed in the next section.

2.5. Fit quality

The quality of a fit is mainly described by the relevance of the data with respect to the model and its parameters: Would the combination of model and parameters provide the same fit quality for another data set, possibly random data? To answer this question, *SASfit* provides measures that serve as indicators of the fit quality (see Fig. 2.2). These are the 'reduced chisqr', the 'Q factor' and the 'R value'.

Following the 'R factor' from crystallography (Hamilton, 1965) during the early stages of this work, the author implemented a fit quality indicator to help users of the program assessing the fit quality during daily analysis work when only the optimization criterion χ^2_{red} was available:

$$R_{\mathrm{rel}} = 100\frac{\sum_{i=1}^{N} |I_{\mathrm{mod}}(q_i) - I_{\mathrm{exp}}(q_i)|}{\sum_{i=1}^{N} |I_{\mathrm{exp}}(q_i)|} \qquad [2.9]$$

$$R_{\mathrm{relerr}} = \frac{100}{N}\sum_{i=1}^{N} \frac{|I_{\mathrm{mod}}(q_i) - I_{\mathrm{exp}}(q_i)|}{|I_{\mathrm{exp}}(q_i)| - |\sigma_{\mathrm{exp}}(q_i)|} \qquad [2.10]$$

It provides a value for the relative observable similarity and deviation of the model curve against the data points which often has to be judged by the human eye due to rare circumstance of a perfect fit. In favour of the more widely known and accepted indicators those values were superseded in the course of further development.

The 'reduced chisqr' value $\chi^2_{\mathrm{red}} = \chi^2/(N - M)$ (see eqn. [2.1]) provides a measure of fit quality across data sets and model configurations, with N the number of data points used for the fit and M the number of parameters of the fit model. χ^2_{red} serves also as the optimization parameter in the least-squares optimization procedure. This value depends heavily on the quality of the data and how well the associated uncertainties were estimated. When representative uncertainty estimates are provided, a value of $\chi^2_{\mathrm{red}} = 1$ indicates that the data are described on average to within the uncertainties. However, if the uncertainties are excessively small (underestimated), or excessively large (overestimated), this condition no longer holds true. The value of the χ^2_{red} measure, therefore, is dependent on the quality of the uncertainty estimates.

The 'Q factor', defined as

$$Q_{\text{factor}} = Q\left(\frac{N-M}{2}, \frac{\chi^2}{2}\right) = \Gamma\left(\frac{N-M}{2}, \frac{\chi^2}{2}\right) \Big/ \Gamma\left(\frac{N-M}{2}\right)$$

$$with \qquad \Gamma(a,x) = \int_x^\infty t^{a-1} exp(-t)\,\mathrm{d}t, \qquad [2.11]$$

provides a second, independent measure of fit quality. It is the probability that a random set of N data points using the same model parameters would produce a χ^2 value equal to or higher than that obtained when using the real data set. For a fit of good quality, Q_{factor} should be in the range of 0.01-0.5 (the smaller the better), with a χ^2_{red} value of approximately 1.

In analogy to the 'R factor' in crystallography (Hamilton, 1965), *SASfit* also provides an 'R value' as quality criterion of a model in data analysis results:

$$R = \frac{\sum_{i=1}^{N} \left| |I_{\text{exp}}(q_i)| - |I_{\text{mod}}(q_i)| \right|}{\sum_{i=1}^{N} |I_{\text{exp}}(q_i)|}. \qquad [2.12]$$

A value of R ranging between 0 and about 0.1 indicates a good to acceptable fit, whereas large values (up to infinity) denote a poor fit. It is especially important to realize that R is only a measure of precision and that it is not able to measure accuracy. Cases of data situations and model combinations that would be reported as false positives or negatives by the value of R are conceivable. Since the function being minimized is weighted by the uncertainties of the measured data (as can be seen in eqn. [2.13]), there is a weighted 'R value', 'wR value', provided (Hamilton, 1965), which takes those uncertainty estimates into account by

$$R_{\text{w}} = \left\{ \sum_{i=1}^{N} \left[\frac{|I_{\text{exp}}(q_i)| - |I_{\text{mod}}(q_i)|}{\sigma_{\text{exp}}(q_i)} \right]^2 \Big/ \sum_{i=1}^{N} \frac{I^2_{\text{exp}}(q_i)}{\sigma^2_{\text{exp}}(q_i)} \right\}^{1/2}. \qquad [2.13]$$

By providing several fit quality scores such as the Q factor and the R value next to the reduced chisqr value, *SASfit* assists the user in assessing the quality of each fit. Assessment of the quality of a fit using these measures is dependent on the provision of good uncertainty estimates. Recently, however, new goodness of fit tests were published to provide help for defining SAS fit quality when uncertainty estimates are not available: noteworthy are the 'correlation map' approach by Franke et al. (2015) and more recently the 'aGoFs' indicator by Henn (2016).

2.6. Confidence in fitted parameter values

In addition to the aforementioned fit quality estimators, *SASfit* provides confidence intervals for the fitted parameters and outputs the internal covariance matrix to enable

the identification of highly dependent parameters, as shown in Fig. 2.4. (The respective menu 'confidence intervals of fit parameters' is accessible via the options menu of the model configuration window.) In order to find optimal model parameters, *SASfit* uses the Levenberg-Marquardt algorithm (Levenberg, 1944) to minimize the χ^2 function [see equation (1)]. Internally, this algorithm approximates the Hessian matrix of the function subject to optimization.

The Hessian matrix provides a means to determine correlation coefficients of all pairs of fit parameters. The inverse of the Hessian matrix is the approximated formal covariance matrix C for the fit. The square root of diagonal elements $C_{jj}^{1/2}$ gives the standard deviation $\sigma = \delta p_j$ of the best-fit parameter p_j, which holds only under the assumption that measurement errors are independent and normally distributed, and that the parameters are not correlated to each other. Recent work indicates that these assumptions are reasonably fulfilled (Franke et al., 2015). Note that *SASfit* provides the standard deviation of the fit parameters, from which confidence intervals can be derived according, for example, to the *Guide to the Expression of Uncertainty in Measurements* (JCGM, 2008).

To assist the user in assessing the accuracy of the result, the author of this work implemented a display of the correlation of fit parameters: the correlation coefficient r_{jk} of every pair of fit parameters is shown by *SASfit* (see Fig. 2.4) in the upper triangular matrix in shades of red depending on their degree of correlation. For two parameters p_j and p_k being optimized, the correlation coefficient r_{jk} is given by

$$r_{jk} = C_{jk} \big/ (C_{jj} C_{kk})^{1/2} \qquad [2.14]$$

For uncorrelated parameters, r_{jk} is expected to have a value close to zero, whereas for strongly correlated parameters $|r_{jk}|$ approaches one. Within the matrix, there is one row and one column associated with each parameter being optimized. They can be highlighted by clicking on a parameter entry in the lower half of the window. By selecting the row and column of two different fit parameters, their correlation coefficient r_{jk} at the position of their common matrix element is highlighted.

When two parameters are strongly correlated it can happen that they both converge to unphysically large or small values during optimization. In this case, one has either to rewrite the form factor with fewer parameters or to fix one parameter to a value ideally determined using another technique. Another strategy would be to enhance the information content obtained from a SAS experiment, for example through contrast variation, potentially allowing for decoupling of the two strongly correlated parameters.

2.7. Batch processing

In many scientific applications series of measurements are performed on a large variety of samples. In the early stages of analysis erroneous ones have to be identified and filtered out. This can evolve into a very time consuming task given the constantly growing amount of data generated at laboratory and large facility instruments.

To increase the evaluation throughput, the author developed an automated routine for fitting many similar data sets: Once a model has been configured, it can be used for processing a batch of data files under 'Options' → 'run batch', as shown in Fig. 2.5. Hovering over the pattern input field reveals a short pop-up help text field on the pattern syntax for file selection. *SASfit* allows filtering of data file names from a user-defined input directory for model-dependent analysis.

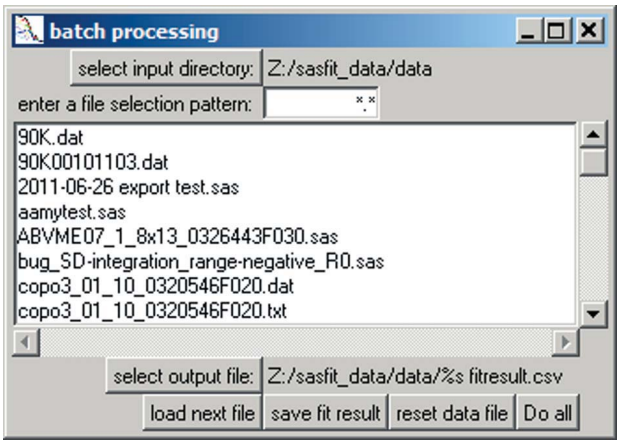

Figure 2.5: Panel for selection of data files for batch analysis and individual output file.

2.8. Custom model functions as plug-ins

In addition to the large library of existing model functions for form factors, structure factors and size distributions, the author developed a flexible plug-in system in the course of his work to allow for custom model functions: it provides everything to enable users to write their own custom form factor and structure factor functions in the C programming language. When users hit the point in their scientific work when none of the existing models sufficiently represent their hypotheses or do not answer the questions at hand, the plug-in system is a good starting point for developing custom models on top of a powerful and proven classical curve fitting toolbox.

Plug-in concept

In *SASfit*, a 'plug-in' is a container for model functions. It may contain an arbitrary number of form factors and structure factors. Both types are supported within a single

plug-in at the same time, but it is recommended to use a plug-in for grouping model functions of a similar kind. In this way, a common set of internal helper routines not accessible publicly can be created and used for all model functions of a plug-in. *SASfit* plug-ins can be exchanged freely between different *SASfit* installations even in binary form, provided the PC platforms and architectures are compatible. To create new customized plug-ins, it is strongly recommended to build *SASfit* from its source code first. In this way, the build environment is verified to work correctly and the plug-in system compatibility is thus assured.

Retrieving the source code

The latest source code of *SASfit*, including a history of all changes, can be obtained on the code-hosting page of the project (`http://sasfit.org`). There are two options to get the most recent source code: the recommended way is to use the distributed version control system (DVCS) Git (`https://git-scm.com`) to 'clone' the project repository locally. This requires a third-party client to be installed but it simplifies the effort of updating to a new version. Compatible GUI (graphical user interface) Clients for Git are listed at `https://git-scm.com/downloads/guis`. Alternatively, the complete source code of a given version or 'snapshot' can be downloaded. The technical details on the required build environment and the instructions for building the *SASfit* program on a specific platform can be found in the documentation (`http://docs.sasfit.org`).

Creating a new plug-in

The first step in creating a new plug-in involves generating a new empty plug-in template containing a directory structure of source code skeleton files. For this, *SASfit* has to be run directly from the source code directory from which it was built and the plug-in guide shown in Fig. 2.6 has to be started. It can be found under the main menu 'Tools' → 'create new plug-in' and lets the user define the setup of a new plug-in function.

The user is required to define a unique plug-in name, and at least one function has to be configured, including a descriptive name under which it can eventually be found in the model selection menu. Additionally, the plug-in guide expects the required parameters of each function to be defined. It is important to know the numerical implementation of the desired model function beforehand to determine the specific parameters needed. As existing model functions cannot be easily modified by the user, it is recommended to replicate a plug-in when modifications are required.

When created, new plug-in templates already contain the configured model functions, but lack any functionality and evaluate to zero. This ensures that the plug-in can be built right from the beginning by issuing the previously used build commands again. This will build only those source code files which are new or changed since the last run. In this

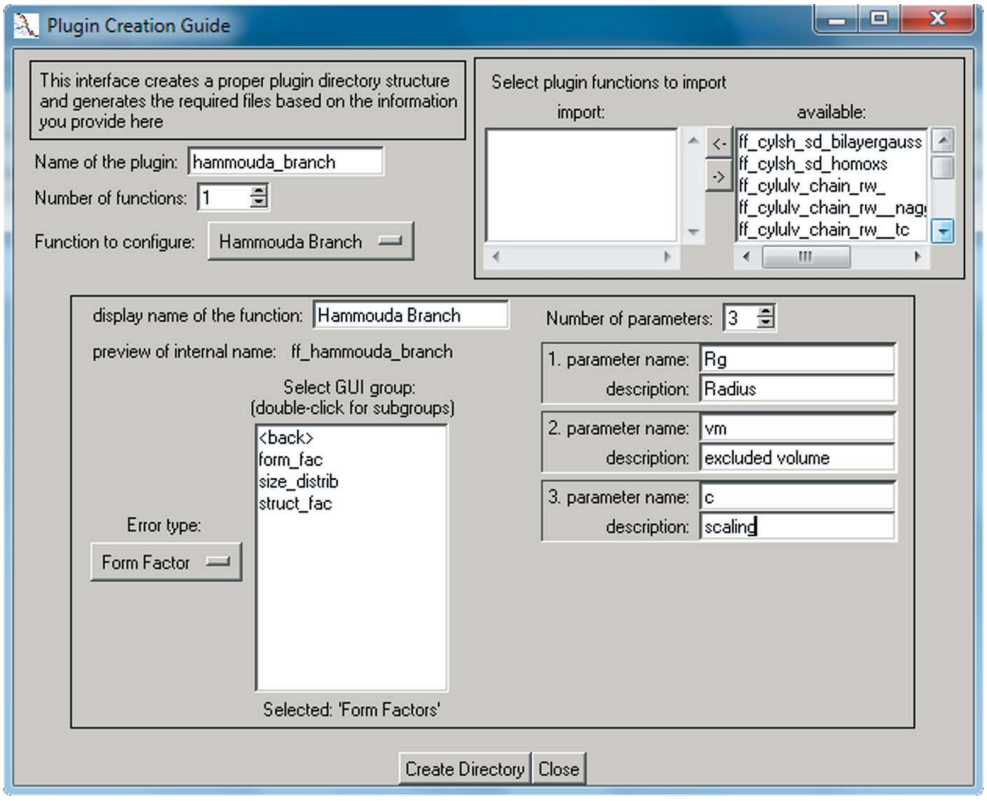

Figure 2.6: User interface for creating a new plug-in template, consisting of user-defined model functions, filled out according to the branched polymer example plug-in.

case it will build the newly created empty plug-in only and add its binary files to the appropriate location automatically. To verify that the plug-in was built correctly *SASfit* has to be restarted, after which the new plug-in will be listed in the appropriate model selection list under 'by plug-ins'.

Branched polymer plug-in function example

Once the initial build of the new plug-in has been successful, it can be populated with the desired model implementation. The following example implements a single-polymer form factor for branched polymers formulated by Hammouda (2012). By using a (mass) fractal model for the minimum path corresponding to the main chain backbone of the polymer, the form factor is described by

$$P_{\mathrm{B}}(Q) = \frac{1}{\mathrm{Norm}} 2 \int_0^1 \mathrm{d}x (1-x) x^{c-1} exp(-U_{\mathrm{B}} x^{2v}), \qquad [2.15]$$

with the normalization factor being defined by

$$\mathrm{Norm} = 2 \int_0^1 \mathrm{d}x (1-x) x^{c-1} = \frac{2}{c(c+1)}, \qquad [2.16]$$

and the scattering variable U_B is expressed in terms of the radii of gyration R_g:

$$U_\mathrm{B} = Q^2 R_\mathrm{g}^2 (2v + c)(2v + c + 1)/6. \qquad [2.17]$$

With a change of variable in $t = U_\mathrm{B} x^{2v}$ and $\mathrm{d}t = 2vU_\mathrm{B} x^{2v-1}\,\mathrm{d}x$ the integral $P_\mathrm{B}(Q)$ evaluates to

$$P_\mathrm{B}(Q) = \frac{1}{\mathrm{Norm}} \left\{ \frac{1}{vU_\mathrm{B}^{c/2v}} \gamma\left(\frac{c}{2v}, U_\mathrm{B}\right) - \frac{1}{vU_\mathrm{B}^{(c+1)/2v}} \gamma\left[\frac{(c+1)}{2v}, U_\mathrm{B}\right] \right\}. \qquad [2.18]$$

The remaining variables v for the excluded volume and c for the scaling factor become parameters of the model function next to the radii of gyration R_g. This formulation of the form factor can be translated into source code of the respective model function in a *SASfit* plug-in.

```
01 scalar sasfit_ff_hammouda_branch(scalar q, sasfit_param * param)
02 {
03     scalar ub, norm_inv;
04     scalar (*gamma) (scalar, scalar);
05
06     SASFIT_ASSERT_PTR(param); // assert pointer param is valid
07
08     // modify conditions to your needs
09     SASFIT_CHECK_COND1((q < 0.0), param, "q(%lg) < 0", q);
10     SASFIT_CHECK_COND1((RG < 0.0), param, "Rg(%lg) < 0", RG);
11     SASFIT_CHECK_COND1((VM < 0.0), param, "vm(%lg) < 0", VM);
12     SASFIT_CHECK_COND1((C < 0.0), param, "c(%lg) < 0", C);
13
14     // insert your code here
15     ub = q*q * RG*RG * (2.*VM + C) * (2.*VM + C + 1.) / 6.;
16     norm_inv = .5 * (C*C + C);
17     gamma = gsl_sf_gamma_inc_P;
18     return ( gamma(.5* C    /VM, ub) / (VM*pow(ub, .5* C    /VM))
19            - gamma(.5*(C+1.)/VM, ub) / (VM*pow(ub, .5*(C+1.)/VM))
20             ) * norm_inv;
21 }
```

Figure 2.7: Source code of the branched copolymer form factor function of a custom *SASfit* plug-in.

The updated code shown in Fig. 2.7 replaces the automatically generated template source code of the function `sasfit_ff_hammouda_branch()` in the file `sasfit_ff_hammouda_branch.c`, which was generated by filling out the *SASfit* plug-in guide as shown in Fig. 2.6.

The function signature in line 1 was created by the plug-in guide along with the mandatory verification of input parameters in line 6. This function is evaluated for every individual Q value of the scattering vector provided in the first argument `scalar q`. Access to predefined input parameters of the model function is provided by automatically generated variables in upper case, `RG`, `VM` and `C`, along with range checks on them in lines 9-12 which were adjusted to a reasonable domain. Each range check consists of different parts. The first part is the condition which will raise an error; for example, in line 9 an

error is to be raised if the scattering vector magnitude q is smaller than zero. The next part is the name of the common parameter, typically `param`. Subsequent parts of a check define an error message to be forwarded to the user. All variables that will be used are declared at the beginning of each model function. Line 3 in this example declares two floating point variables which will be defined later, while line 4 declares a short-cut name of a function which expects two input values. In line 17 this is set to a specific gamma function provided by the GNU Scientific Library (GSL, http://www.gnu.org/software/gsl/). The model function defined by the formula of Hammouda for branched polymers itself is implemented on lines 15-20, with the scattering variable U_{B} defined in line 15. The inversion of the normalization factor on line 16 replaces two divisions by multiplications in the final formula on line 20.

As demonstrated in the example, model functions in *SASfit* can make use of any function in the GSL but may also use a large set of predefined mathematical functions provided by *SASfit* directly. For example, a convenient wrapper `sasfit_integrate()` is available which simplifies usage of GSL integration routines by managing workspace memory in the background. Additionally, custom routines can make use of model functions defined in other plug-ins by declaring to import them during configuration with the plug-in guide. More information on plug-ins in *SASfit*, as well as an extensive guide on how to start writing custom models for *SASfit* on the Windows, Linux or MacOS platform complete with video guides, can be found online (http://docs.sasfit.org/Overview:_Plugins).

2.9. Example 2: characterization of a bimodal silica particle size distribution

The interpretation of multimodal size distributions of nanoparticles is a demanding typical *SASfit* application. The procedure is demonstrated here by interpreting the scattering pattern of a bimodal size distribution of silica nanoparticles in aqueous solution. Recently, a suitable particle mixture was released as a certified reference material denoted ERM-FD102 (Kestens, Roebben, et al., 2016), which is commercially available as a European reference material. The intended use of ERM-FD102 is the quality control and assessment of performance of nanoparticle size analysis methods, including SAXS.

A sample volume of 20 µl was measured as received for 30 min on a commercial SAXS instrument. Its scattering intensity was converted to absolute scale using water as primary standard according to the procedure described by Orthaber et al. (2000) and was verified using a measurement of bovine serum albumin (Mylonas and Svergun, 2007). The resultant scattering curve with data in the range of $q_{\mathrm{min}} = 0.057\,\mathrm{nm}^{-1}$ to $q_{\mathrm{max}} = 3.0\,\mathrm{nm}^{-1}$ is shown in Fig. 2.8. The obtained data set and a preconfigured *SASfit* project file (`SilicaS2929.pdh`

and `SilicaS2929.sas`) is provided in the supporting information of the original work (Breßler et al., 2015).

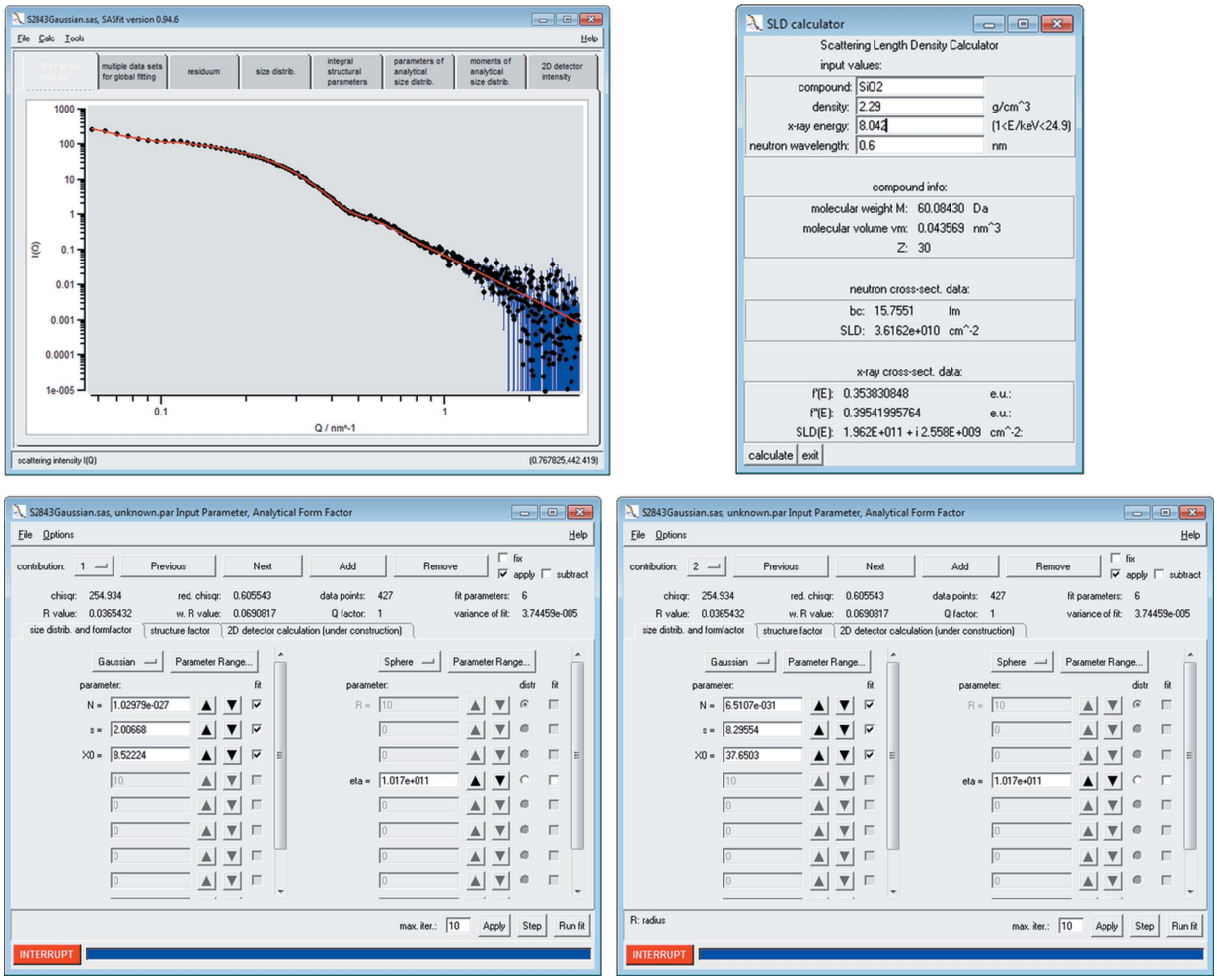

Figure 2.8: Upper left-hand figure: SAXS data of bimodal silica nanoparticles (European reference material ERM-FD102) and a curve fitted using a model of spheres with Gaussian size distribution (black dots and red solid curve, respectively). The uncertainties of the intensity values are displayed as vertical blue lines. Upper right-hand figure: the scattering length density calculator provides a scattering length density of $1.962 \times 10^{11}\,\mathrm{cm}^{-2}$ for SiO_2 particles with a density of $2.29\,\mathrm{g\,cm}^{-3}$. Lower figures: panels of the sphere form factor with Gaussian size distribution for the small particles (contribution 1, left-hand side) and large particles (contribution 2, right-hand side).

In the first step of data evaluation the scattering contrast 'eta' between the silica particles and the solvent is calculated. For this purpose the scattering length density calculator, which is available under the 'Tools' menu entry of *SASfit* was used. The silica particle scattering length density is $1.962 \times 10^{11}\,\mathrm{cm}^{-2}$ by the stoichiometry of silica SiO_2, the density of $2.29(1)\,\mathrm{g\,cm}^{-3}$ (Finsy, Moreels, et al., 1985) and the copper $K\alpha$ X-ray energy of 8.042 keV (see Fig. 2.8). In contrast to water, which has a scattering length density of $9.45 \times 10^{10}\,\mathrm{cm}^{-2}$, the sample scattering contrast is $1.017 \times 10^{11}\,\mathrm{cm}^{-2}$.

Next, a sphere model for the particles' form factor with a Gaussian size distribution was chosen under the 'Calc' → 'Single data set' → 'fit' menu entry. Therein the value of the scattering contrast 'eta' was inserted as a fixed parameter for 'contribution 1' and for

Table 2.1: Parameters of silica nanoparticles ERM-FD102 fitted with a bimodal Gaussian size distribution.

The population of small particles is labelled as 'Particle class A' and that of the large particles as 'Particle class B' in accordance with the ERM-FD102 certification report (Kestens and Roebben, 2014). Fit parameters are the particle number N, the mean radius X_0 and the width of the size distribution s. The values of the mass fraction of the particles φ_m are given as derived from the *SASfit* parameter N and calculated from the data given in the certification report.

Parameter	Particle class A Contribution 1	Particle class B Contribution 2
N (cm^{-3}, $\mathrm{mol\,l}^{-1}$)	$1.02(1) \times 10^{15}$ $1.69(2) \times 10^{-6}$	$6.51(148) \times 10^{11}$ $1.08(25) \times 10^{-9}$
φ_m (*SASfit*) (mg g^{-1})	7.05(7)	0.38(9)
φ_m (certification report)* (mg g^{-1})	8.33	0.42
Mean radius X_0 (*SASfit*) (nm)	8.52(4)	37.65(330)
Mean radius (certification report)† (nm)	9.1(4)	42.0(5)
Width of distribution s (nm)	2.00(3)	8.29(304)
Number ratio N_1/N_2	1567(371) ($\pm$ 24%)	
Mass ratio $\varphi_{\mathrm{m},1}/\varphi_{\mathrm{m},2}$ (*SASfit*)	18.5(46) ($\pm$ 25%)	
$\varphi_{\mathrm{m},1}/\varphi_{\mathrm{m},2}$ (certification report)	19.8	

*Values were calculated from the information on the production data given in the certification report of ERM-FD102 (Kestens and Roebben, 2014).

†Number-weighted modal area-equivalent diameter as obtained by transmission and scanning electron microscopy (Kestens and Roebben, 2014).

'contribution 2'. Next, the fitting procedure described in the curve fitting workflow section (see also Fig. 2.3) was carried out. The resultant best fit curve is shown together with the data points in Fig. 2.8 (red solid curve and points, respectively). The corresponding best fit values are displayed in the fit panels for the particle size contributions 1 and 2, respectively (lower row of Fig. 2.8). The best fit values of the parameters and estimates of their uncertainties are displayed when clicking the button 'parameters of analytical size distribution'.

The parameters of the ERM-FD102 sample are shown in Table 2.1. The estimate for the mean radius of silica particle class A is $X_0 = 8.52(4)\,\mathrm{nm}$ and that for class B is $37.65(330)\,\mathrm{nm}$. These values are in good agreement with the number-weighted modal area-equivalent radii of 9.1(4) and 42.0(5) nm obtained by transmission and scanning electron microscopy (Kestens, Roebben, et al., 2016): Comparing the measurement results with the certified values according to the 'ERM Application Note 1', gives an absolute difference of $\Delta_\mathrm{m} = |9.1 - 8.52| = 0.58\,\mathrm{nm}$ with an expanded uncertainty of $U_\Delta = 2\sqrt{0.4^2 + 0.04^2} = 0.8\,\mathrm{nm}$ satisfying $\Delta_\mathrm{m} \leq U_\Delta$ which shows agreement between the measurement result and the certified value of class A. For class B, $\Delta_\mathrm{m} = |42 - 37.65| = 4.35\,\mathrm{nm}$ with an expanded uncertainty of $U_\Delta = 2\sqrt{5.5^2 + 3.3^2} = 12.82\,\mathrm{nm}$ indicates no significant difference between the measurement result and the certified value as well.

At first glance, it is surprising that the uncertainty of the mean radius is much larger for the larger particles (class B) than for the smaller ones. However, the particle size of class B is close to the instrumental limit of $\pi/q_{\mathrm{min}} = 56\,\mathrm{nm}$, making its accurate determination more challenging, as also discussed by Pauw, Pedersen, et al. (2013). In contrast, the radii of class A particles are far away from both the upper resolution limit and the low resolution limit of $\pi/q_{\mathrm{max}} = 1\,\mathrm{nm}$. Accordingly, the uncertainty of the radii of class B becomes relatively large in comparison to that of the particles of class A.

The widths of the size distributions of class A and B are 2.00(3) and 8.29(304) nm, respectively, which are typical values for commercial silica particles. Also for s, the uncertainty for class B is larger than for class A for the same reason as for X_0. It should be noted that number-weighted size distributions are important for the characterization of nanomaterials, which are defined by the European Commission as 'a natural, incidental or manufactured material containing particles, in an unbound state or as an aggregate or as an agglomerate and where, for 50% or more of the particles in the *number size distribution*, one or more external dimensions is in the size range 1-100 nm' (Potočnik, 2011).

Here, *SASfit* provides direct access to an estimate of number-weighted size distributions of nanoparticles. The implemented formula for curve fitting of spheres, $I_{\mathrm{sphere}}(q, R, \Delta\eta)$, with Gaussian number-weighted size distribution, $\mathrm{Gauss}(R, N, \sigma, R_0)$, is

$$I_{\mathrm{SASfit}}(q) = \int_0^\infty \mathrm{Gauss}(R, N, \sigma, R_0)\, I_{\mathrm{sphere}}(q, R, \Delta\eta)\, \mathrm{d}R, \qquad [2.19]$$

where the Gaussian size distribution is defined as

$$\mathrm{Gauss}(R, N, \sigma, R_0) = N \left((\pi/2)^{1/2}\sigma \left\{1 + \mathrm{erf}\left[R_0/(2^{1/2}\sigma)\right]\right\}\right)^{-1} \times \exp\left[\frac{-(R - R_0)^2}{2\sigma^2}\right] \qquad [2.20]$$

and the scattering of a sphere is given by

$$I_{\mathrm{sphere}}(q, R, \Delta\eta) = \left\{\frac{4}{3}\pi R^3 \Delta\eta \left[3\frac{\sin(qr) - qR\cos(qR)}{(qR)^3}\right]\right\}^2. \qquad [2.21]$$

The approach to estimate the number-weighted distribution is only useful if the distribution is relatively narrow, typically smaller than 20% relative width. Alternatively, a very broad distribution has to be reasonably estimated by other means beforehand. In contrast, the Monte Carlo approach for analysis of SAS data provides good estimates of volume-weighted size distribution but is much less suited for number-weighted determinations, owing to the lack of assumptions for the asymptotic behaviour of the

distributions (see section 3.3 and Pauw, Pedersen, et al. (2013)). *SASfit* calculates the number density distribution internally as long as the form factor is expressed in terms of a size, because it always contains the volume information. By default, the resulting distribution is plotted volume weighted instead for convenience. However, a different parametrization can be implemented if required; for example, a form factor of spheres can be defined to use input parameters of volume, after which volume-weighted Gaussian distribution parameters can be directly determined.

Following the procedure described in Example 1: disperse gold nanoparticles, the fitted N values of class A and B particles were converted to particle number concentrations of $1.02(1) \times 10^{15}$ and $6.51(148) \times 10^{-11}\,\mathrm{cm}^{-3}$, respectively. The molar concentrations were $1.69(2) \times 10^{-6}$ and $1.08(25) \times 10^{-9}\,\mathrm{mol\,l}^{-1}$. Therefore, the number ratio of small to large particles N_1/N_2 is 1567(371). The mass fractions were calculated assuming that the silica particles in the example have a density of $2.29\,\mathrm{g\,cm}^{-3}$ (Finsy, Moreels, et al., 1985), resulting in $\varphi_{\mathrm{m},1} = 7.05(7)\,\mathrm{mg\,g}^{-1}$ for class A and $\varphi_{\mathrm{m},2} = 0.38(9)\,\mathrm{mg\,g}^{-1}$ for class B. Using the composition data given in the certification report of ERM-FD102 (Kestens, Roebben, et al., 2016) the mass fractions are $\varphi_{\mathrm{m},1} = 8.33\,\mathrm{mg\,g}^{-1}$ and $\varphi_{\mathrm{m},2} = 0.42\,\mathrm{mg\,g}^{-1}$. From these values, the mass ratio derived from *SASfit* $\varphi_{\mathrm{m},1}/\varphi_{\mathrm{m},2} = 18.5(46)$ is in good agreement with the values of $\varphi_{\mathrm{m},1}/\varphi_{\mathrm{m},2} = 19.8$ derived from the certification report. It can be concluded that the precision and accuracy of the *SASfit* parameters, and values derived thereof, are in good agreement with the reported values for the silica reference material.

3

McSAS: Monte Carlo fitting of SAS data

McSAS is a user-friendly open source Monte Carlo regression package, which structures the analysis of SAS data using uncorrelated shape-similar particles (or scattering contributions). The underdetermined problem is solvable, provided that sufficient external information is available. Based on this, the user picks a scatterer contribution model (or 'shape') from a comprehensive library and defines variation intervals of its model parameters. A multitude of scattering contribution models are included, such as prolate and oblate nanoparticles, core-shell objects, several polymer models, and a model for densely packed spheres. Most importantly, the form-free Monte Carlo nature of *McSAS* means it is not necessary to provide further restrictions on the mathematical form of the parameter distribution; without prior knowledge, *McSAS* is able to extract complex multimodal or odd-shaped parameter distributions from SAS data. When provided with data on an absolute scale with reasonable uncertainty estimates, the software outputs model parameter distributions and provides the modes of the distribution (*e.g.* mean, variance *etc.*). In addition to facilitating the evaluation of (series of) SAS curves, *McSAS* also helps in assessing the significance of the results through the addition of uncertainty estimates to the result. The *McSAS* software can be integrated as part of an automated reduction and analysis procedure in laboratory instruments or at synchrotron beamlines. By providing the package as open source, universal access is granted and thus, allows it to be reviewed publicly as well as being adjusted by any user to her specific needs.

The program has been used in a variety of works, such as for investigating particle formation during nanoparticle synthesis by pulsed laser ablation in liquids (Letzel et al., 2017), for analysing the evolution of aluminosilicate sol structures (Zhao et al., 2017), characterising zinc phosphate cements and evaluating chemically-bonded ceramics using SANS (Viani, Sotiriadis, Kumpová, et al., 2017; Viani, Sotiriadis, Sasek, et al., 2016), studying the effects of nanostructural evolution on the properties of alloys (Mamiya et al., 2016), understanding iron(III) oxide nucleation (Scheck et al., 2016) and the synthesis of ultra-small core-shell silver nanoparticles (Kästner and Thünemann, 2016), analysing protein binding of colloidal aggregates related to drug discovery (Duan et al.,

2017), investigate the intestinal mobility and cytotoxicity of nanoparticles (Lichtenstein, Ebmeyer, Meyer, et al., 2016; Lichtenstein, Ebmeyer, Knappe, et al., 2015) as well as studying light-responsive supramolecular polymers (Concellón et al., 2016).

3.1. Core concept

The *McSAS* method is a Monte Carlo rejection sampling approach for retrieving model parameter distributions, such as size distributions, from scattering patterns (Pauw, Pedersen, et al., 2013). Central to the method lies a set of independent non-interacting contributions, each of which is an instance of the elementary scatterer model chosen by the user. Depending on the chosen model, one or more parameters can be selected for Monte Carlo optimization (hereafter, referred to as 'fitting parameter').

If the measured data are provided in absolute units, calculations can be performed resulting in absolute volume fractions. In order for this to work, information on the scattering length densities of the phases within the sample needs to be provided. For the included two-phase models, only the differences of the scattering length density of the scatterer and that of the matrix are necessary. These scattering length densities are readily obtained from a variety of online tools (D. Brown and Kienzle, 2015).

The optimization procedure (shown in Fig. 3.1) progresses through replacement of contributions in the set. At the end of the optimization procedure, the spread of fitting parameter values of the contributions in this set defines the final parameter distribution.

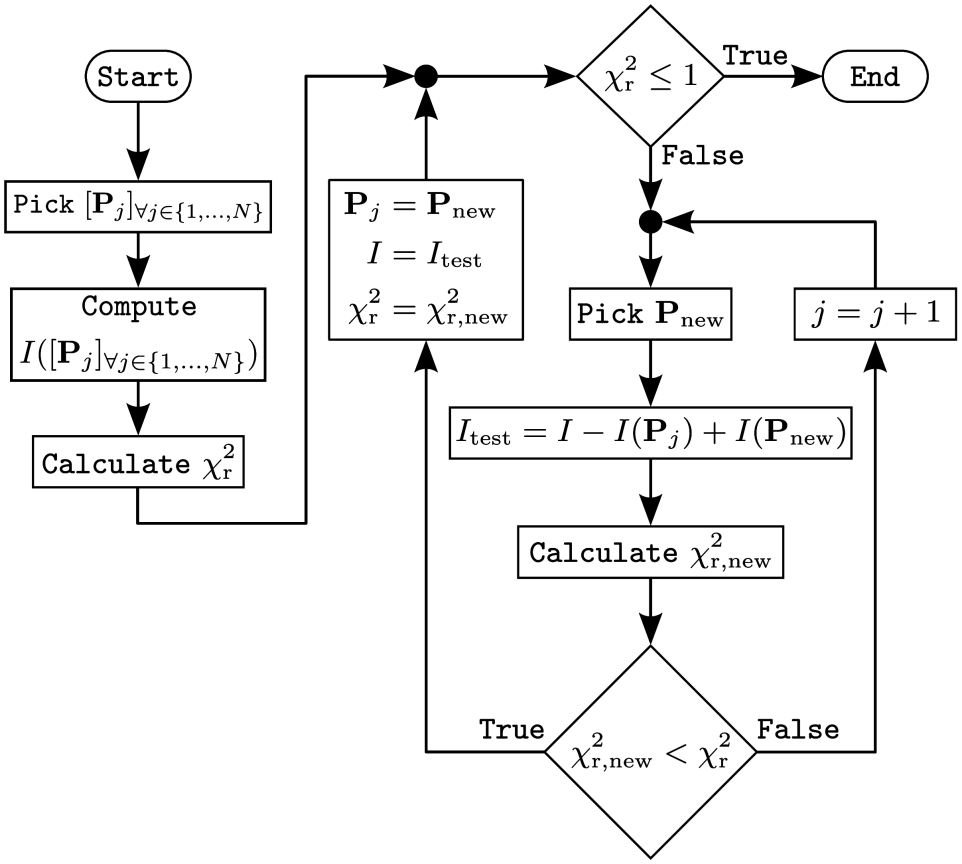

Figure 3.1: The main process of the *McSAS* software for parameter optimization. In each cycle, an attempt is made to replace one of the model contributions in order to improve the agreement between the model and measured data.

3.2. Optimization procedure

The method starts from a set of non-interacting scatterers of predefined shape, such as spheres, rods or ellipsoids, but with random values chosen for the fitting parameters of each contribution in the set. The total model scattering pattern is given by the weighted sum of the scattering patterns of each scatterer in the set. The scattering patterns of the individual contributions of the set are weighted with the inverse of their volume to obtain a more evenly matched impact on the total (Porod, 1952). This process drastically reduces the required number of contributions and improves fitting flexibility. For most practical scattering patterns, about 200-300 contributions are sufficient to describe the scattering patterns with a minimum of optimization iterations.

A goodness of fit criterion (χ^2_r, see eqn. [2.1]) is calculated from the model based on the distance between the model and measured data, weighted by the measured data uncertainty estimates (Pedersen, 1997). In order to obtain this goodness of fit measure, the model intensity is matched to the measured data set through scaling and addition of an optional flat background contribution. The scaling and background parameters are obtained by least-squares minimization of χ^2_r. The goodness of fit thus indicates when the model describes the data on average to within the data uncertainty ($\chi^2_r \leq 1$). Thereby, a

suitable cutoff criterion for optimization is provided, which generally prevents over-fitting of the data and allows for the estimation of uncertainties on the resultant distribution.

Each iteration of the Monte Carlo (MC) procedure consists of replacing one of the scattering objects in the set by another object of the same basic shape but with different, randomly chosen values for its fitting parameter(s). This replacement is accepted if it reduces χ^2_r, *i.e.* if the agreement of the resulting MC scattering pattern with the measured pattern is improved. These iterations continue until the convergence criterion of $\chi^2_r \leq 1$ is reached (the convergence criterion value can be adjusted by the user to support data with over- or underestimated uncertainties). After completion, the model parameter distribution is determined through grouping (binning) of the fitting parameter values in the set.

In addition to this, a 'minimum observability limit' is determined for each contribution in the set, which specifies the minimum volume fraction of scatterers required to make a measurable contribution to the scattering pattern (*i.e.* a contribution exceeding the measurement uncertainty). More specifically, a minimum observability limit $\varphi_{\min,k}$ (in units of volume fraction) can be defined for any method where the total model intensity comprises a set of quantized components, whose partial contributions are $I_k(q)$ for a given component volume fraction φ_k, and where the measurement data uncertainty $\sigma(q)$ is available:

$$\varphi_{\min,k} = \min_{q\in(q_{\min},q_{\max})} \left[\frac{\sigma(q)\varphi_k}{I_k(q)} \right]. \qquad [3.1]$$

Its derivation and use is further explored elsewhere (Pauw, Pedersen, et al., 2013).

Finally, the uncertainty on the resultant parameter distribution is determined through analysis of the sample standard deviation of a multitude of independent MC solutions. These uncertainty estimates and the observability limits are key values in the application of the method. They provide information to distinguish between numerical noise and size distribution components which are shown by the data and, moreover, allow for the assessment of the statistical significance of differences in resultant size distributions. The accuracy of such uncertainty estimates and observability limits in the *McSAS* result are, however, directly reliant on the provision of reasonable uncertainty estimates on the measured data.

With this procedure, *McSAS* is able to retrieve any form-free size distribution provided a basic scatterer shape is given. A test of the retrievability of a wide range of unimodal and multimodal size distributions has been demonstrated for a large variety of simulated size distributions in the supplementary information given by Pauw, Pedersen, et al. (2013). A comparison between size distributions of precipitates in alloys is also available, obtained from electron microscopy and *McSAS* analysis of small-angle X-ray scattering (SAXS) data (Rosalie and Pauw, 2014).

3.3. MC method benefits and drawbacks

Benefits

McSAS has proven to be remarkably useful owing to its ability to work in absolute units and the wide variety of available models. These models include spheres, isotropic cylinders and ellipsoids, core-shell ellipsoids, and core-shell spheres.

For densely packed spheres, a model is included based on the local monodisperse approximation (LMA), which is one of the few structure factors that can be directly implemented given the internal design of the MC method, coupled with the Percus-Yevick (PY) approximation (see 3.5.2) (Kinning and E. L. Thomas, 1984). This particular model combination will, hereafter, be referred to as 'LMA-PY'. Furthermore, two polymer chain models have been added by the author: Kholodenko worm (Kholodenko, 1993) and Gaussian chain (Debye, 1947).

For additional flexibility in available models for analysis, the author implemented an easy to use plug-in system which allows users to use their own scattering models. Since program version 1.3, models written in the *Python* programming language can be provided independently of the program package. Once they are copied to the `models` directory they are loaded upon program start and can be selected from the pulldown menu of the '4. Model' panel. Compatible models are searched for in subdirectories as well. The directory name is used then for grouping models in the GUI. This can be exploited to structure custom models of similar kind in groups.

McSAS can run with or without an user interface, enabling integration into existing data processing procedures. Multiple data files can be provided on the command line for batch fitting. The fitting procedure can then be automatically initiated, inheriting the settings of the previous GUI instance.

Because of the high importance of different weightings for various scientific applications, the author implemented graphical output and population statistics for a user-specified number of parameter ranges (regions of interest). The distributions can be shown with the (horizontal) parameter axes scaled logarithmically or linearly and the (vertical) parameter axes weighted by fractions of volume, number, surface or intensity (which is volume-squared for SAS). For all of the different weightings, estimated uncertainties are provided which allows the user to evaluate the reliability of the chosen distribution immediately.

For broad distributions, however, volume-weighting is strongly recommended (see the note in 3.3). The distributions shown include the minimum observability limit, *i.e.* the minimum required amount for each contribution to be statistically significantly contributing.

Lastly, population statistics of the solution are determined independently of the histogramming procedure. For each selected parameter and range, the total value and the

four distribution modes are provided: the mean, variance, skew and kurtosis. These are number-, volume-, surface- or intensity-weighted depending on the user's choice. Such population statistics simplify the analysis of population trends for *in situ* experiments or other inter-related data sets.

Drawbacks

Owing to its 'brute force', iterative nature, the method is not as fast as some of the alternatives mentioned in the *Introduction*. Optimization speed is strongly dependent on the accuracy of the data. Prior work has shown that reasonable data uncertainty may be estimated as the maximum of the values given (a) by the standard error of the mean (obtained during the averaging or binning procedure), (b) by propagating measurement uncertainties through the data corrections and (c) by limiting the data uncertainty to be no less than 1% of the data value (Pauw, Pedersen, et al., 2013; Rosalie and Pauw, 2014; Schnepp et al., 2013). A data set with 1-2% uncertainty may require a few minutes on a normal modern desktop computer. This is expected to improve in the near future through implementation of multithreading.

Secondly, there is a risk of under-specifying the fitting model when more complex models are chosen. For example, if a cylindrical scatterer model is chosen, and its length and radius are allowed to span the same size range, the solution is no longer unique and a multitude of valid solutions will be found. This manifests itself through excessive uncertainties in the result, originating from large discrepancies between the independent *McSAS* repetitions. Such ambiguity can be easily arrived at when using models such as core-shell objects and anisotropic objects. For these complex shapes, the allowed size ranges for the shape parameters may require the application of strict constraints before a unique solution is obtained.

Two common failure modes of the *McSAS* method can occur. The first happens when data are provided containing unrealistically low uncertainty estimates, which will lead to an attempt by *McSAS* to describe ostensibly significant data variations as features in the size distribution. This will lead either to a failure to reach convergence or to spurious features in the resulting parameter distribution. To alleviate this problem somewhat, the uncertainty is clipped to be at least 1% of the intensity value. Previously, this has been found to be a practical limit from data correction considerations (Pauw, Pedersen, et al., 2013) and is a value supported by experimental results (Hura et al., 2000). This lower limit can be adapted or bypassed if better estimates can be guaranteed.

The second failure mode occurs when the fitting range is set too broad, *i.e.* beyond the range supported by the data. It is recommended to keep size parameters within the limits dictated by the q range of the data, estimated as $\pi/q_{\max} \leq R \leq \pi/q_{\min}$. Exceeding these limits may result in spurious features appearing beyond these limits, as explored in the supplementary information given by Pauw, Pedersen, et al. (2013).

Of further note is that small-angle scattering data represent a volume-weighted distribution according to Porod (Porod, 1952). The implication of this is that volume-weighted size distributions can be easily retrieved using a Monte Carlo approach, as small to medium *volume fractions* of small-sized scatterers can be readily distinguished in scattering patterns. Not so for small *number fractions* of small-sized scatterers in a disperse mixture, for which exceedingly little evidence exists in most measured scattering patterns and in particular for broad size distributions. Therefore, when broad size distributions determined using *McSAS* are shown in their number-weighted form, the values and uncertainties of the small-sized components can be seen to vary excessively, with some values becoming untenably large. Such issues are usually not encountered when using classical fitting methods.

Classical methods circumvent this issue, as the integral equation that is solved strictly constrains the distribution probability at all sizes (*e.g.* the probability distribution typically assumes a value of zero at its smallest size). These methods are, therefore, seemingly capable of determining even small number fractions of small scatterers accurately. However, the evidence for small numbers of small-sized scatterers may be very weak in the data. In summary, the strict assumptions placed in classical methods on the number-weighted size distribution shape may conceal the lack of evidence for the absence or presence of scatterers at the small end of the distribution, thus implying accuracy where there is none.

3.4. Current implementation

The existing program code contained the core algorithm for command line usage within a Python environment. Originally, the standard sphere model was the only particle scattering model and an integral part of the optimization algorithm. To add flexibility towards the actual scattering model being used for optimization and to allow for more complex scattering models, the program code had to be redesigned by the author to be more modular.

For practical use in analyses conducted by scientific users who are not programmers, it was found necessary in the course of this work to add batch processing of a larger number of measured data sets. This leads to the requirement of a structured storage of calculation results and algorithm parameters alongside the original input data. As usually required in scientific work, it ensures that all resulting data can be reproduced based on the original input data at a later point in time.

With a focus on (size) distribution analysis, a flexible postprocessing framework was developed. For each parameter population being subject to optimization, custom histograms over partial ranges can be created without performing a new optimization based on the final results. Additionally, various statistical properties are determined

alongside each histogram. Thus, the additional statistical information help to quickly assess the plausibility of the calculated distributions.

User interface features

The user interface is divided into several panels, each limited to a different aspect of the process (see Fig. 3.2). These consist of a '1. Data Files' panel, a '2. Data Settings' panel, a '3. Optimization' panel, a '4. Model' panel and a '5. Post-fit Analysis' panel and will be discussed in order.

The '1. Data files' panel shows data files loaded upon startup (as command-line arguments) or files added through the right-click menu. All files will be treated identically when the fit is run, though their order of processing can be changed as desired. Available data are read from the input file, which is expected to consist of three semicolon-separated columns of $q\,(\mathrm{nm})$, $I\,[(\mathrm{m\ sr})^{-1}]$ and the uncertainty estimate $\sigma(I)\,[(\mathrm{m\ sr})^{-1}]$.

To help with determining reasonable limits of size parameters in particular, basic analysis is performed when loading each data file. The minimum and maximum values of the provided q vector are used to estimate the maximum and minimum possible scatterer size under the assumption of solid spherical scatterers. Those estimates are displayed next to each data file and, by double-click, can be applied as optimization limits for radius-type model parameters.

The '2. Data Settings' panel offers detailed control over the preprocessing of data during loading, before any optimization is carried out. A lower and upper limit may be specified, for excluding erroneous or dispensable data points from the beginning and/or from the end of the signal, which might improve the numerical stability of the solution or shorten the optimization run time, respectively. Moreover, the number of data points can be reduced effectively by a logarithmic rebinning procedure which maintains the information content qualitatively. It helps to reduce oversampled data sets and improves the numerical stability of the fitting procedure. A minimum uncertainty estimate implements a lower boundary for the uncertainties of the data to avoid underestimating them, which may prevent the algorithm to reach convergence. The last options of this panel allow to compensate for smeared data, caused by the collimation system of the SAS instrument. The degree of smearing of the modelled data can be specified in detail for appropriate comparison with the measured data.

The '3. Optimization' panel contains a subset of MC algorithm settings addressable by the user (see Table 3.1). The most important of these is the chi-squared (convergence) criterion. While this is per default set to 1, it may prevent reaching a state of convergence ($\chi^2_\mathrm{r} \leq 1$) for data whose uncertainty estimates are insufficiently large or poorly estimated. Increasing this value will allow the convergence condition to be reached, after which the fit may be evaluated. This increase directly affects the uncertainties on the resultant

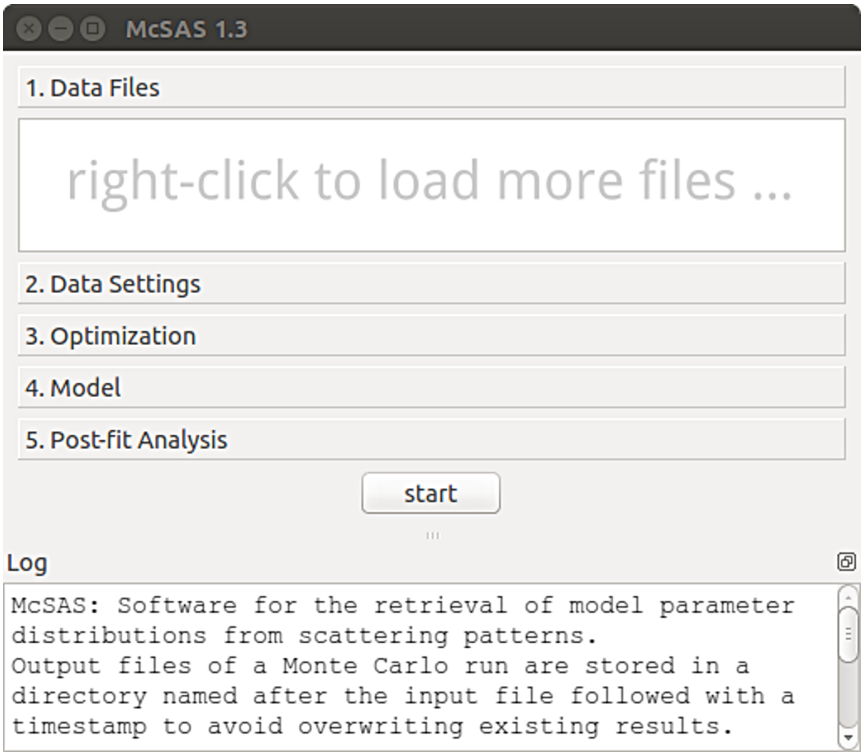

Figure 3.2: The main interface of the *McSAS* software upon startup, showing five configuration panels. The '1. Data Files' panel allows selection and input of the data of interest, the '2. Data Settings' panel allows configuration of the data sets by setting physically meaningful constraints, the '3. Optimization' panel contains settings to adjust the optimization method behaviour, '4. Model' contains all parameters and settings relevant to the chosen morphology, and '5. Post-fit Analysis' holds the settings for histogramming and visualization of intensity-, volume-, surface- and/or number-weighted distributions.

distribution. Additionally, the number of shape contributions can be increased. While the default setting of 300 is sufficiently large to reach the convergence criterion for most scattering patterns, and small enough to reach it rapidly, there may be cases for which an increased number is desired. A detailed discussion on determining the number of shape contributions can be found in Pauw, Pedersen, et al. (2013) (§2.5). Likewise, the number of repetitions can be changed. These independent repetitions are used to estimate the uncertainties on the resultant size distribution, but a reduced number should suffice for initial testing. Lastly, a selection can be made on whether a flat background contribution is to be taken into account when matching the MC intensity to the detected signal.

The '4. Model' panel contains all information on the model used to describe the scatterer morphology. The pulldown menu offers a selection of models that can be used to define the contributions' scatterer shape. The associated parameters and options for the model chosen will then be shown on the right-hand side. Parameters which are selected for fitting require upper and lower bounds to be set.

The '5. Post-fit Analysis' panel offers basic analysis capabilities for interpretation of the MC result. When a range entry is added, the user can select which parameter to histogram, what parameter range to consider (if they decide not to automatically follow the model parameter range) and how many bins to use. Increasing the number of histogram bins

will lead to increased detail in the resulting histogram at the cost of larger uncertainties and evidence requirements *via* observability limits. Furthermore, a choice can be made whether to use a linear or logarithmic parameter scale (useful for distributions spanning several decades) and whether to plot volume-, number-, surface- or intensity-weighted size distributions. When using absolute units, and if scattering contrast is provided, the volume-weighted distribution will contain absolute values; the number-weighted distribution is normalized to an area of one, for lack of information.

Finally, the 'Start' button launches the optimization, and the 'log' shows the output of the program as it runs and is automatically stored in a file. During the optimization process the current goodness of fit value χ^2_r is given along with the user-defined convergence criterion which has to be reached for a successful fit. Additionally, to indicate the presence of systematic errors the author of this work implemented an alternative goodness of fit measure 'aGoFs' (Henn, 2016).

Table 3.1: Selected program parameters and their effects on the computation. For the advanced settings and defaults that can be found in the `mcsasparameters.json` file, only selected values are listed.

Location or Panel	Parameter name	Effect
GUI 'Data Files'	'right click to load files'	Loads multiple measurement files in ASCII format or PDH format, which is basically an ASCII file with an header and trailing meta data.
GUI 'Data Settings'	Lower/Upper q cut off	Defines the range of data to use for evaluation. By default, all data values are used for fitting.
	Minimum uncertainty estimate	Implements a lower boundary for the uncertainties of the data by raising them to the given fraction of the measured signal, if they are not greater than that already. It can be increased or reduced based on a best guess estimate for minimum inter-related data point uncertainty. A value too low may prevent reaching convergence. By default, it is set to 1% of the measured signal.
	Target number of bins	Sets the number of bins for logarithmic rebinning of the data. This reduces the number of data points for evaluation efficiently, which results in faster optimization runs. Due to empty bins, the resulting number of data points may be smaller than the given number of bins. A value of zero disables binning. By default, it is set to 100 bins.
	Apply smearing correction	Activates smearing of the modelled data for comparison with the measured data which was smeared by the collimation system of the SAS instrument. Deactivated by default.
	Slit-smeared data or 2D-averaged data	Defines, which collimation system was used for measuring the data. 2D collimated systems ('point focus') require a different smearing than line-collimated data.
	Number of smearing points around each q	Sets the level of detail of the simulated smearing. The default value of 25 evaluates the scattering model at 24 additional positions for each q given by the data. Therefore, it increases the computation time significantly.
	Variance (σ^2) of Gaussian beam profile	Defines the full width at half maximum (FHWM) of the Gaussian beam profile for line-collimated instruments. The profile is circularly averaged for 2D pinhole and rectangular collimated instruments.

(Continued on the next page)

Location or Panel	Parameter name	Effect
GUI 'Optimization'	Convergence criterion	The least-squares value (χ_r^2) at which the fit is considered a success. For data with realistic uncertainty estimates, this can be set to 1. For a quick fit, it can be set to larger values. Values below 1 can be used if the uncertainties are overestimated.
	Number of repetitions	The number of independent optimizations to be run. Larger values will result in improved uncertainty estimates on the result and a slightly smoother result, but calculation time increases proportionally.
	Automatically close result windows	If activated, it prevents cluttering the screen with many plot windows if data series are fitted. Deactivated by default.
	Advanced Settings	Provides expert users with more detailed options of the optimization procedure for an enhanced investigation of their data.
	Number of contributions	The number of individual contributions whose weighted sum comprises the total model intensity. Too few or too many will result in slow optimization times. Most patterns can be fitted using 300 contributions quickly, but times can be optimized using the timing information shown in the result.
	Weighting compensation	Adjusts internal weighting of scattering pattern contributions. Adjustment between 0.3 and 0.7 may lead to slight speed increases for some samples. Set to $^3/_2$ by default.
	Find background level	If selected, a flat background is fitted during matching of model and data. This speeds up the fit with minimal effect on the result, as many scattering patterns contain a flat scattering component as well, due to density variations or incoherent scattering. Enabled by default.
	Maximum iterations	If convergence has not been reached within this number of iterations, the optimization attempt is aborted. Larger values may allow complex calculations to finish successfully, but often non-convergence can be traced back to poor initialization settings. Increasing this value increases the maximum possible calculation time. The default value is 10^5.
	Plot unsuccessful or incomplete results	Activates plotting of incomplete results after an interrupted run or an optimization that did not reach convergence. Disabled by default.
	Calc. series statistics	Activates a combined plot of the distribution mean for all data sets being optimized. Additionally, all distribution moments are gathered in a common output file for convenient export. Disabled by default.

(Continued on the next page)

Location or Panel	Parameter name	Effect
GUI 'Post-fit Analysis'	Parameter	The parameter to show the distribution histogram of.
	Lower & upper	The distribution will be shown for this parameter range only. This can be used to cut off regions outside the range of interest. Population statistics also apply only to this range.
	Number of bins	The number of divisions to use in the distribution display. By increasing this number, more detail may be visible provided one stays within the Shannon channel limit (Svergun and Pedersen, 1994) (indicated in the '1. Data Files' panel). An increase in the number of divisions will also negatively affect uncertainty estimates and observability limits.
	X-axis scaling	The Scaling (linear or logarithmic) of the parameter axis of the distribution. Logarithmic scaling is recommended for wide parameter ranges.
	Y-axis weighting	The vertical axis can be shown in volume, number, intensity or surface distributions. Volume-weighted distributions are recommended; number-weighted distributions can be used for samples with a narrow dispersity; intensity-weighted distributions reflect the weighting of the measured signal; surface-weighted distributions are relevant for applications concerning the overall particle surface.
`mcsasparameters.json` (file, advanced settings and defaults)	startFromMinimum	Sets the initial guess of the Monte Carlo algorithm constant to $1/2$ of the parameter fit range minimum, rather than a uniform random distribution.

3.5. Experimental evaluation

3.5.1. Example 1: bimodal nanoparticle reference material

To test the ability of the program to retrieve bimodal size distributions, a reference material was measured containing two fractions of silica nanoparticles: ERM-FD102 (Kestens, Roebben, et al., 2016). The first fraction consists of 0.36 vol.% nanoparticles with a certified number-weighted modal area-equivalent radius of 9.1(4) nm obtained by transmission and scanning electron microscopy. Correspondingly, the second fraction consists of nanoparticles with a radius of 42.0(5) nm at a volume fraction of 0.018 vol.%. This makes the volume ratios of the two components 95% and 5%, respectively.

The measurements were performed on an Anton Paar line-collimated instrument utilizing mirror-monochromated copper $K\alpha$ radiation and modified to use a Dectris Mythen detector. The data have been calibrated to absolute intensity using the methods described by Orthaber et al. (2000), the scaling of which was verified using a measurement of bovine serum albumin. The reference material, a water background and an empty capillary background were measured for 30 min each, corrected and desmeared using the software provided by Anton Paar. The data thus collected have been regrouped and averaged over 50 intervals, logarithmically spaced in q. The uncertainty has been set as the largest of (a) the propagated data uncertainty or (b) the standard error on the mean. None of these uncertainty estimates are smaller than 1% of the intensity value. The data span a q range of $0.057 \leq q \leq 2.88\,\mathrm{nm}^{-1}$, corresponding to an estimated size range of $1.09 \leq R \leq 55.4\,\mathrm{nm}$.

Analysis of the data using the *McSAS* program results in the fit shown in Fig. 3.3. To achieve this, the standard settings have been used, except for the number of repetitions (100) and the scattering length density difference ($1.017 \times 10^{-5}\,\mathrm{Å}^{-2}$). The two populations have been correctly resolved in the analysis (see Fig. 3.4), with the first population $1.09\,\mathrm{nm} \leq R \leq 25\,\mathrm{nm}$ having a volume-weighted mean of 9.86(6) nm and a volume fraction of 0.311(3) %. The second population $25\,\mathrm{nm} \leq R \leq 55.4\,\mathrm{nm}$ has a volume-weighted mean of 45(2) nm and a volume fraction of 0.013(4) %.

Comparing the measurement results with the certified values (Table 2.1), gives an absolute difference of $\Delta_\mathrm{m} = |9.1 - 9.86| = 0.76\,\mathrm{nm}$ with an expanded uncertainty of $U_\Delta = 2\sqrt{0.4^2 + 0.06^2} = 0.81\,\mathrm{nm}$ for the first population which agrees to the certified value due to satisfying $\Delta_\mathrm{m} \leq U_\Delta$ according to the 'ERM Application Note 1'. Whereas the measurement of the second population shows an absolute difference of $\Delta_\mathrm{m} = |42 - 45| = 3\,\mathrm{nm}$ which is significantly smaller than the expanded uncertainty $U_\Delta = 2\sqrt{0.55^2 + 2^2} = 4.15\,\mathrm{nm}$ and therefore, agrees with the certified value. Furthermore, as the large-sized component is present at small volume fractions and close to the end of the measurement range, its characterization is not as accurate as that of the small component.

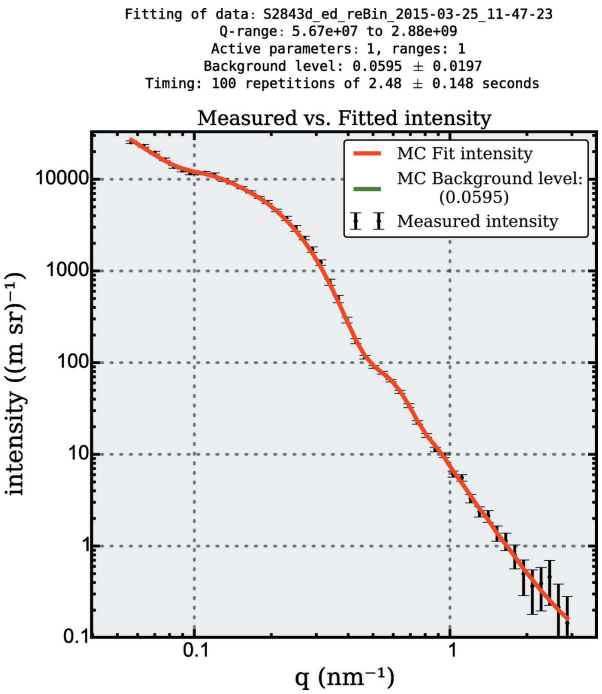

Figure 3.3: *McSAS* graphical output showing the best fit obtained using the MC method to a scattering pattern obtained from a mixture of dilute particles with certified radii of 9.1(4) nm and 42.0(5) nm silica particles. The particle volume ratio of small to large particles is 19:1.

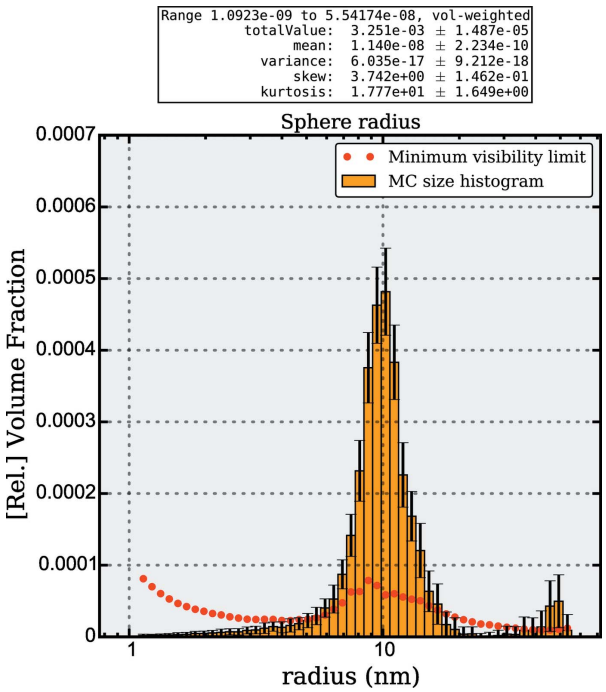

Figure 3.4: *McSAS* graphical output showing the volume-weighted size distribution associated with the MC fit of dilute silica particles shown in Fig. 3.3.

3.5.2. Example 2: densely packed nanoparticles

Dense systems add a degree of complexity to small-angle scattering and are, therefore, interesting as a test case for MC methods. A suitable data set of densely packed, dry SiO_2 spheres (with a stated radius of 75 nm) has been provided by Peter Høghøj of Xenocs, as part of a demonstration data set measured on their Xeuss SAXS instrument. The SiO_2 spheres are packed in a randomly jammed fashion, implying that the volume fraction v_f is approximately 0.63 (Song et al., 2008).

A reasonable fit can be obtained using classical fitting methods implemented in *SASfit* (Breßler et al., 2015), with a model of Gaussian distributed spheres and a structure factor consisting of a PY hard-sphere interaction model assuming the local monodisperse approximation, forming the aforementioned LMA-PY combination. This resulting fit is shown in Fig. 3.5. Most of the intensity can be described well, apart from the region at low q.

A fit to within data uncertainty can be obtained using the same model in *McSAS* (see Fig. 3.6), with the volume fraction v_f set to 0.63. Note that the instrumental resolution has not been considered in either the classical or the MC approach. The main feature in the resultant size distribution (shown in Fig. 3.7) is indeed at the size indicated for the sample [with a number-weighted mean radius of 76.1(2) nm], but a minor component is visible at about half the radius of the main component.

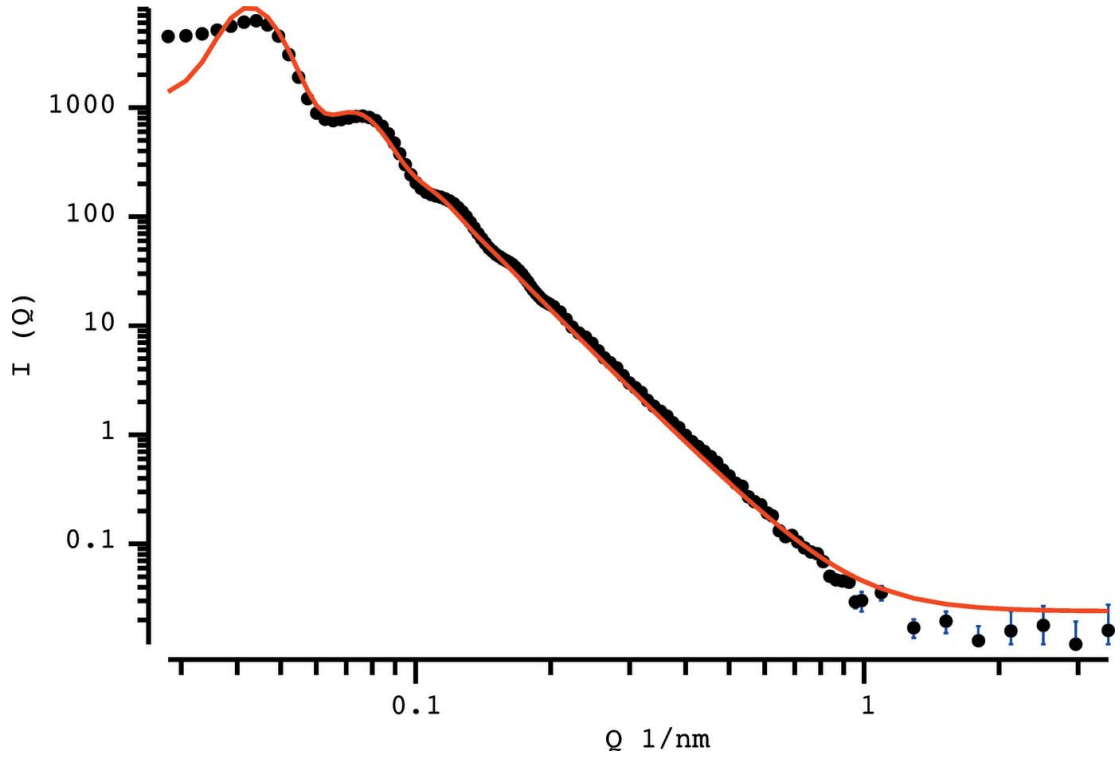

Figure 3.5: Best fit (solid line) using a classical model (implemented in *SASfit*) to a scattering pattern obtained from packed silica spheres (symbols). Model uses a sphere form factor with a LMA-PY structure factor and a Gaussian size distribution.

While the origin of the minor features cannot be established without further investigation, a similar good fit can be obtained when other volume fractions are set (possible between $\sim 0.5 \leq v_{\mathrm{f}} \leq 0.7$). Changing the volume fraction drastically affects the size distribution and demonstrates that there are a multitude of solutions accessible through adjustment of the volume fraction. This highlights once more that information must be provided on the sample to allow SAS analyses to arrive at a unique solution. However, the overall result is quite satisfactory and a clear improvement compared with the classical approach.

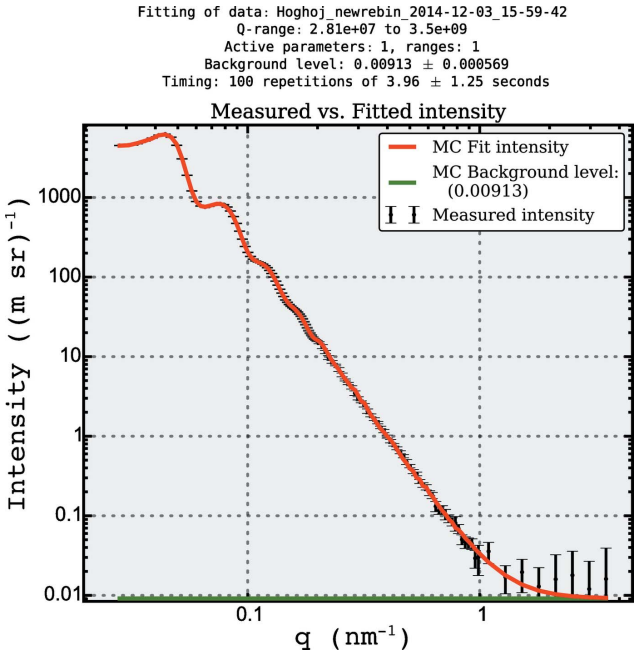

Figure 3.6: The best fit obtained using the MC method to a scattering pattern obtained from packed silica spheres. Model using a sphere form factor with a LMA-PY structure factor.

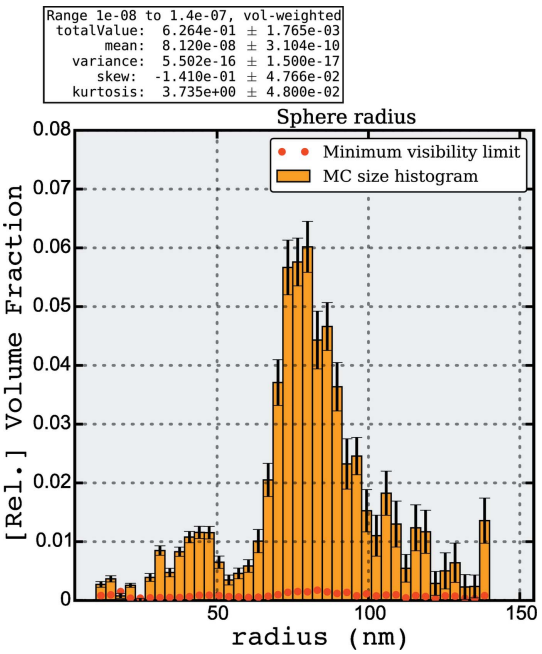

Figure 3.7: *McSAS* graphical output panel showing volume-weighted size distribution associated with the MC fit of Fig. 3.6.

3.6. Conclusion

The *McSAS* package provides the user with a comprehensive model library for advanced SAS data analysis. In addition, it can be extended with additional models. The relatively high computational effort increases with the complexity of new models, but should not be a major concern given the increasing availability of computing power. Concerning the significance of the resulting particle distributions, care must be taken that sufficient external information is provided to ensure a unique solution. This is of particular importance for complex models.

4

McDLS: Monte Carlo fitting of DLS data

The aim of this work is a contribution to improve the shortcomings of currently available DLS analysis methods discussed in section 1.2 and provide users with a Monte Carlo-based curve fitting program for the analysis of relaxation functions (Istratov and Vyvenko, 1999) in particular for reliable DLS data analysis. The program is provided as open source software named *McDLS* in analogy to *McSAS*, which is dedicated to the analysis of small-angle scattering data (section 3.1).

4.1. Modelling DLS data

DLS measures the intensity autocorrelation function $G_2(\tau)$ of the light intensity $I(t)$, which was scattered, for example, by nanoparticles as

$$G_2(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2} \qquad [4.1]$$

where τ is the decay time [see Finsy (1994) for details]. The $G_2(\tau)$ is also called the second-order correlation function as indicated by its index number. It is related to the normalized electric field (or first order) autocorrelation function, $g_1(\tau)$, by the Siegert relation (Borsali and Pecora, 2008) as

$$G_2(\tau) = 1 + \beta\left[g_1(\tau)\right]^2. \qquad [4.2]$$

The meaning of β will be discussed in a separate section. In case of particles of uniform size, *i.e.* Brownian objects moving with the translational diffusion coefficient, D, the first order autocorrelation function simplifies to a mono-exponential decay

$$g_1(\tau) = \exp(-\Gamma\tau) \qquad [4.3]$$

with the characteristic decay rate $\Gamma = Dq^2$. The scattering vector is defined as $q = 4\pi n\lambda^{-1}\sin(\theta/2)$ with the refractive index of the solvent n, the wavelength of the laser light λ and the scattering angle θ. Note that θ is half the angle between incident and scattered beam. For the general case of particles with an arbitrary size distribution, the Laplace

transformation of the distribution function of the decay rates $G(\Gamma)$ provides the electric field correlation function

$$g_1(\tau) = \int_0^\infty G(\Gamma) \exp(-\Gamma\tau)\, \mathrm{d}\Gamma. \qquad [4.4]$$

The 'normal' way to reveal the particles' hydrodynamic radii distribution is to obtain $G(\Gamma)$ with an inverse Laplace transformation from $g_1(\tau)$. Hence, this problem is solved, in principle. Unfortunately, the inversion is an ill posed problem (Istratov and Vyvenko, 1999) on which many efforts were spent for polydispersity analysis (Finsy, 1994). In the case of DLS data interpretation, the inverse Laplace transformation requires first to reveal $g_1(\tau)$ from $G_2(\tau)$, a procedure that includes background subtraction from the data, proper scaling and root finding. Afterwards, a numerical Laplace transformation of the processed data must be performed. This approach often produces questionable results, for one reason in particular: due to very different noise levels of the data. The whole situation is very unsatisfactory for understanding how DLS data can be interpreted in terms of hydrodynamic radii distributions.

A central aspect of this work is not to obtain the particles' R_H distribution through an inverse Laplace transformation. Instead, a distribution of hydrodynamic radii $f(R_\mathrm{H})$ is guessed and compared to the corresponding calculated intensity correlation function iteratively, until it cannot be distinguished from the measured data within the experimental uncertainties. In the following, the proposed *McDLS* method is described and simulations and practical examples are provided.

Calculation of the intensity correlation function

For an ensemble of N particles of an arbitrary size distribution, the discrete field autocorrelation function is a weighted sum of exponential decay functions as

$$g_1(\tau) = \sum_{j=1}^{N} a_j \exp(-\Gamma_j\tau) \qquad [4.5]$$

with the constrained conditions

$$\sum_{j=1}^{N} a_j \leq 1 \quad \text{and} \quad a_j \geq 0, \qquad [4.6]$$

as explained, for example, by Zhu et al. (2011). An a_j is the relative amplitude scattered by the j^{th} particle with the decay constant Γ_j, representing the mean value of each interval $[\Gamma_j, \Gamma_{j+1}]$ from the total number of intervals N. The Γ_j is related to the diffusion coefficient D_j of the j^{th} class of particles with the hydrodynamic radius $R_{\mathrm{H},j}$ by $\Gamma_j = D_j q^2$. Together

with the Stokes-Einstein-equation (Einstein, 1905) $D_j = k_B T/(6\pi\eta R_{H,j})$, the Γ_j values are calculated by:

$$\Gamma_j = \frac{k_B T}{6\pi\eta R_{H,j}} q^2, \qquad [4.7]$$

where k_B is the Boltzmann constant, T is the absolute temperature, η is the viscosity of the solvent and q is the scattering vector.

It can be seen from eqn. [4.7] that $R_{H,j}$ is the only free variable in a given DLS experiment while the other variables are constant. In consequence, a large fraction of recurring calculations during regularly multiple Monte Carlo optimizations, can be avoided by including

$$\Gamma_j = \frac{k_{DLS}}{R_{H,j}} \quad \text{with} \quad k_{DLS} = \frac{k_B T}{6\pi\eta} q^2 \qquad [4.8]$$

Next, abbreviating $R_{H,j}$ by R_j, the scattering amplitude of a sphere is modelled as

$$a(R_j) = V(R_j)\Phi(R_j), \qquad [4.9]$$

where the volume of the sphere is $V(R_j) = \frac{4}{3}\pi R_j^3$ and the scattering amplitude of a sphere is

$$\Phi(R_j) = 3\frac{\sin(qR_j) - qR_j\cos(qR_j)}{(qR_j)^3}. \qquad [4.10]$$

The field correlation function that guarantees the constraints in eqn. [4.6] is calculated by

$$g_1(\tau) = \frac{\sum_{j=1}^{N} a(R_j)^2 \exp(-\Gamma_j \tau)}{\sum_{j=1}^{N} a(R_j)^2}. \qquad [4.11]$$

After inserting eqn. [4.8] and eqn. [4.9],

$$g_1(\tau) = \frac{\sum_{j=1}^{N} V(R_j)^2 \Phi(R_j)^2 \exp(-\frac{k_{DLS}}{R_{H,j}} \tau)}{\sum_{j=1}^{N} V(R_j)^2 \Phi(R_j)^2}. \qquad [4.12]$$

The denominator guarantees normalization at $g_1(0) = 1$. The aim of the *McDLS* tool is to fit directly the R_H values using the experimentally determined $G_2(\tau) - 1$ curves. Consequently, after inserting eqn. [4.12] in eqn. [4.2], one gets

$$G_2(\tau) - 1 = A + B\left[\frac{\sum_{j=1}^{N} V(R_j)^2 \Phi(R_j)^2 \exp(-\frac{k_{DLS}}{R_{H,j}} \tau)}{\sum_{j=1}^{N} V(R_j)^2 \Phi(R_j)^2}\right]^2. \qquad [4.13]$$

The baseline A and the scaling B are automatically determined by least-square fitting in each optimization step, right before comparing the calculated $G_{2,\mathrm{mod}}(\tau) - 1$ with the measured $G_{2,\mathrm{exp}}(\tau) - 1$ data.

At first glance, it is irritating that $V(R_j)$ is squared in the numerator of eqn. [4.13] because that corresponds to R_j^{12} terms in total. But on the second glance it becomes clear that the denominator of $G_2(\tau) - 1$ in eqn. [4.13] ensures the normalization of $G_2(0) - 1 = 1$.

On the usage of the sphere form factor

The use of a sphere from factor for calculation of the scattering amplitude has the practical reason that the hydrodynamic radius of a compact sphere is the same as its geometrical radius. Furthermore, the volume of a sphere is well-defined, in contrast to a Gaussian polymer coil, for example. Additional form factors can be implemented in *McDLS* by the user, if needed. The q-dependency of the scattered light is small for particles in the 'nano'-size range of 1 to 100 nm but becomes complicated for larger particles and needs to be considered. The distinct minima of the sphere form factor can produce an unexpected angular-dependency of the intensity correlation functions, for example, for multimodal polymer latex particles, as discussed in detail by Shibayama et al. (2006). This is illustrated by the scattering form factor of a sphere defined as

$$\begin{aligned} P(q, R_j) &= \Phi(q, R_j)^2 \\ P(q, R_j) &= \left[3 \frac{\sin(qR_j) - qR_j \cos(qR_j)}{(qR_j)^3} \right]^2 \end{aligned} \qquad [4.14]$$

Data in multiangle DLS are typically recorded in a scattering angle range of $(20 \leq 2\theta \leq 150)\,^\circ$, which corresponds to a scattering vector range of $(0.005 \leq q \leq 0.025)\,\mathrm{nm}^{-1}$ when using a laser emitting light with a wavelength of 632 nm. The scattering form factors of spherical particles with radii of 5 nm, 50 nm and 500 nm are shown in Fig. 4.1 to illustrate cases in which it becomes important to care about the particles' form factor. It can be seen that the scattering intensities at $q = 0.019\,\mathrm{nm}^{-1}$ (a scattering angle of 90°) decrease from 1 to 0.998 (5 nm), 0.838 (50 nm) and 0.01 (500 nm), compared to their scattering at $q = 0$. The scattering of compact nanoparticles with radii smaller than 50 nm decrease gradually but deep minima are present in the scattering of larger particles as can be seen in the inset of Fig. 4.1. Therefore, the use of the scattering form factor of spheres for modelling the intensity correlation function in eqn. [4.13] appears reasonable for nanoparticles. But artefacts can be easily produced if the particles sizes are well beyond the nanoscale. If the particles of interest show a strong intensity variation in the relevant q-range, it is recommended to replace $\Phi(q, R_{\mathrm{H},j})$ by expressions for the amplitude adapted to the size and type of the particles. An example is given by Shibayama et al. (2006) for latex particles, where an approximation was used to avoid artefacts due to minima in the

scattering function. Otherwise, it is possible to take advantage of the strong dependency of larger particles for improvement of multimodal particle size analysis with multiangle DLS (Naiim et al., 2014).

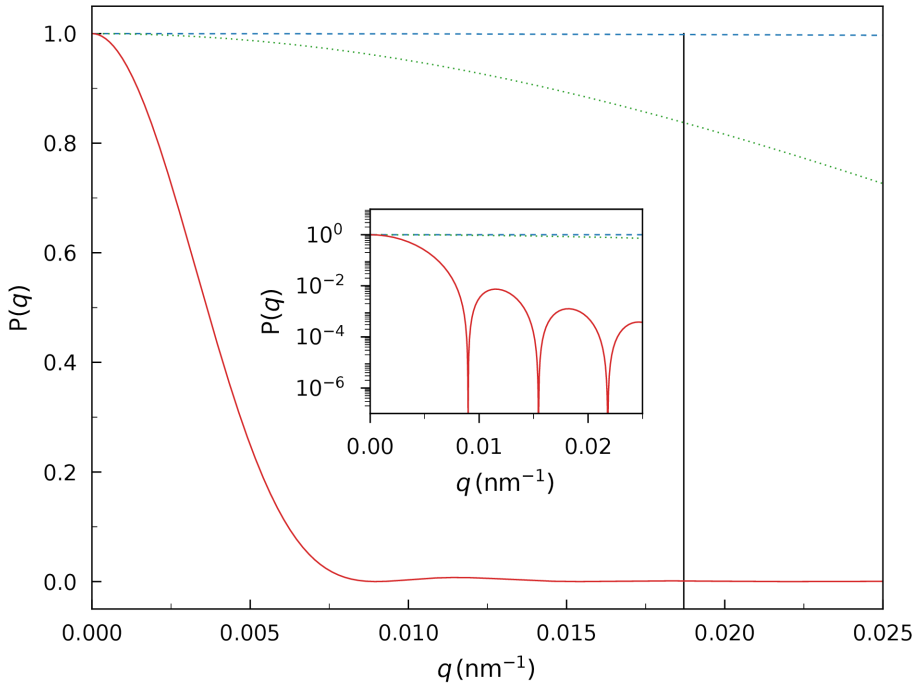

Figure 4.1: Scattering form factors of spherical particles with radii of 5 nm, 50 nm and 500 nm (dashed blue, dotted green and solid red line, respectively). The vertical black line is at a scattering angle of 90°. The inset shows a logarithmic presentation with deep minima in the form factor of the particles with a radius of 500 nm.

Coherence factor and baseline

The β in eqn. [4.2] is called coherence factor and is determined largely by the ratio of the detector area to the coherence area of the scattered light (Vanhoudt and Clauwaert, 1999). In practice, β is regarded as an unknown parameter to be fitted in the data analysis. Experimentally, the intensity correlation function is typically provided in the form of

$$G_2(\tau) - 1 = A + B\left[g_1(\tau)\right]^2, \qquad [4.15]$$

where in ideal cases the A is close to 0 and B is close to 1. In the relevant international standard on particle size analysis (ISO, 1996), these parameters have been recommended to be in the ranges of $0 \leq A \leq 0.1$ and $0.8 \leq B \leq 1$ for practical purposes. When using *McDLS*, it is recommended that the experimental values of A and B are found within these standard conform ranges. Caution is advised for interpretation of data for which A and B are beyond these recommended ranges. The A and B are automatically determined in *McDLS* between distinct optimization steps by a least-square fit of the modelled correlation

function to the measured correlation function. Typically, the value of the baseline A is determined as

$$A = [G_2(\tau) - 1]_{\tau \to \infty} . \qquad [4.16]$$

At large correlation times it can be seen that $G_2(\tau) - 1$ fluctuates around a small value, *i.e.* a linear fit with zero slope for the last third of the data values could be considered to determine A. The topic of how to determine A properly is of vital importance if $g_1(\tau)$ should be extracted correctly for further analysis since it biased the results, as discussed by Finsy (1994) in detail. Since the *McDLS* approach avoids the necessity of extraction of $g_1(\tau)$, the background finding has not the same importance as if an inverse Laplace transformation is foreseen to reconstruct $f(R_\mathrm{H})$. The value of B is determined, in practice, as the value of $G_2(\tau) - 1$ at the shortest meaningful experimental correlation time τ_min, *i.e.*

$$B = [G_2(\tau_\mathrm{min}) - 1] - A. \qquad [4.17]$$

4.2. McDLS fitting procedure

The core optimization procedure is described by the flow chart in Fig. 4.2. It shows that, initially, a discrete random set $[R_{\mathrm{H},j}]_{\forall j \in \{1,\ldots,N\}}$ of hydrodynamic radii of N particles is generated. The user is encouraged to provide reasonable boundaries by adding the lowest and highest meaningful value of the hydrodynamic radii $R_\mathrm{H,min}$ and $R_\mathrm{H,max}$, respectively. The predefined number of R_H-values of $N = 300$ should be suitable for most cases but can be decreased or increased by the user. The effect of the number of contributions in the Monte Carlo method was discussed in detail by Pauw, Pedersen, et al. (2013) (§2.5). In a second step, the radii contributions are used to calculate the first-order correlation function $g_1([R_\mathrm{H}])$. The third step compares the second-order correlation function $G_2([R_\mathrm{H}]) - 1 = Bg_1([R_{\mathrm{H},j}])^2 + A$ to the measured data by calculating the goodness of fit measure χ^2_r. If the convergence criterion $\chi^2_\mathrm{r} \leq 1$ is satisfied, the simulated intensity correlation function cannot be distinguished from the measured one within the experimental uncertainties and the set of radii $[R_\mathrm{H}]$ is considered a reasonable approximation of $f(R_\mathrm{H})$.

If the simulated intensity correlation function fails to match the measured one, the next iteration of the optimization cycle is started by replacing a single particle radius $R_\mathrm{H,j}$ by a randomly selected one $R_\mathrm{H,new}$ from the range $R_\mathrm{H,min} \leq R_\mathrm{H,new} \leq R_\mathrm{H,max}$. To clarify: only one radius value is changed in the set. In the following step, the new second-order intensity correlation function $G_\mathrm{2,new}$ is calculated from the old one by subtracting the field correlation function of a single particle $g_1(R_{\mathrm{H},j})$ from the first-order correlation function g_1

of the previous comparison and adding that of the new particle $g_1(R_{\mathrm{H,new}})$. Therefore, the second-order correlation function of the updated particle ensemble appears to be

$$G_{2,\mathrm{new}} = B(g_1 - g_1(R_{\mathrm{H},j}) + g_1(R_{\mathrm{H,new}}))^2 + A \qquad [4.18]$$

which is subsequently compared to the measured data by calculating the goodness of fit measure $\chi^2_{\mathrm{r,new}}$. It presents an improvement over the previous test if the condition $\chi^2_{\mathrm{r,new}} \leq \chi^2_{\mathrm{r}}$ is satisfied. If it fails, a new iteration is started by selecting another $R_{\mathrm{H,new}}$ from the predefined range. If the test succeeds, the particle ensemble is updated by (a) replacing the old radius $R_{\mathrm{H},j}$ by the new one $R_{\mathrm{H,new}}$, (b) replacing the modelled field correlation function g_1 by $g_1 - g_1(R_{\mathrm{H},j}) + g_1(R_{\mathrm{H,new}})$ and (c) updating the current goodness of fit measure χ^2_{r} by the most recent one $\chi^2_{\mathrm{r,new}}$. The last step of an iteration tests if the current goodness of fit reaches the convergence criterion: $\chi^2_{\mathrm{r}} \leq 1$. On failure, the next iteration starts and on success, the modelled correlation function matches the measured one within the experimental uncertainties and a solution for $f(R_{\mathrm{H}})$ is found. The optimization runs until a maximum number of iterations $n_{\max}$ is reached or a solution is found. The $n_{\max}$ has a predefined value of 10^5 but it can be chosen freely by the user. One optimization run consists of hundreds or thousands of iteration cycles until a convergence between simulation and experiment is found. But it finds only one of many possible solutions for $f(R_{\mathrm{H}})$ which are consistent with the experiment. Many repetitions of the fitting procedure, each resulting in a different $f(R_{\mathrm{H}})$, allowing to estimate the uncertainties of the distribution, which are finally plotted as single histogram with error bars. Typically, 50 of such repetitions are sufficient for reasonable error estimation but the user can chose fewer repetitions for faster results or more repetitions for a more accurate error estimation.

Averaging of data sets

Averaging of correlation curves is activated by default in *McDLS* and can be disabled by unchecking the option 'average each angle' in the 'Data Settings' menu of *McDLS*. If deactivated, no averaging and no outlier filtering is performed, resulting in a single data set for each scattering angle and measurement number combination. For example, ten measurements performed at eight angles, generate eighty data sets listed in the 'Data Files' menu.

With averaging enabled, all remaining correlation curves which satisfy the Z-Score threshold Z are identified by index $l \in L = \{m : Z_m < Z\}$, where index m identifies an

initially loaded correlation curve and Z_m the modified Z-Score of that curve, as defined in eqn. [4.22]. The mean correlation curve is then calculated by

$$\overline{G_2(\tau) - 1} = \frac{1}{|L|} \sum_l (G_{2,l}(\tau) - 1), \tag{4.19}$$

alongside its corrected sample standard deviation

$$\sigma(\tau) = \left(\frac{1}{|L| - 1} \sum_l \left| G_{2,l}(\tau) - 1 - \overline{G_2(\tau) - 1} \right|^2 \right)^{1/2}. \tag{4.20}$$

Over the course of optimizing a model correlation curve against the averaged correlation curve $\overline{G_2(\tau) - 1}$ by the goodness of fit criterion χ^2_r in eqn. [4.21], the $\sigma(\tau)$ of eqn. [4.20] serves as an estimate for the measurement uncertainty.

Optimization

A goodness of fit criterion is minimized similar to eqn. [2.1]:

$$\chi^2_\mathrm{r} = \frac{1}{(N-1)} \sum_{i=1}^{N} \left[\frac{\overline{G_{2,\mathrm{exp}}(\tau_i) - 1} - G_{2,\mathrm{mod}}(\tau_i) - 1}{\sigma_\mathrm{exp}(\tau_i)} \right]^2 \tag{4.21}$$

and according to the procedure described in section 3.1.

Analogous to the intensity of small-angle scattering, *McDLS* optimizes the modelled second order correlation curve $G_{2,\mathrm{mod}}(\tau) - 1$ to match the measured and averaged one, by iteratively adjusting individual values of the parameter contributions of the hydrodynamic radius R_H. The program flow of this procedure is shown by Fig. 4.2 where $[\mathbf{R}_{\mathrm{H},j}]_{\forall j \in \{1,\ldots,N\}}$ represents a set of N radii, used for calculating the modelled curve $G_{2,\mathrm{mod}}(\tau) - 1$ according to eqn. [4.13].

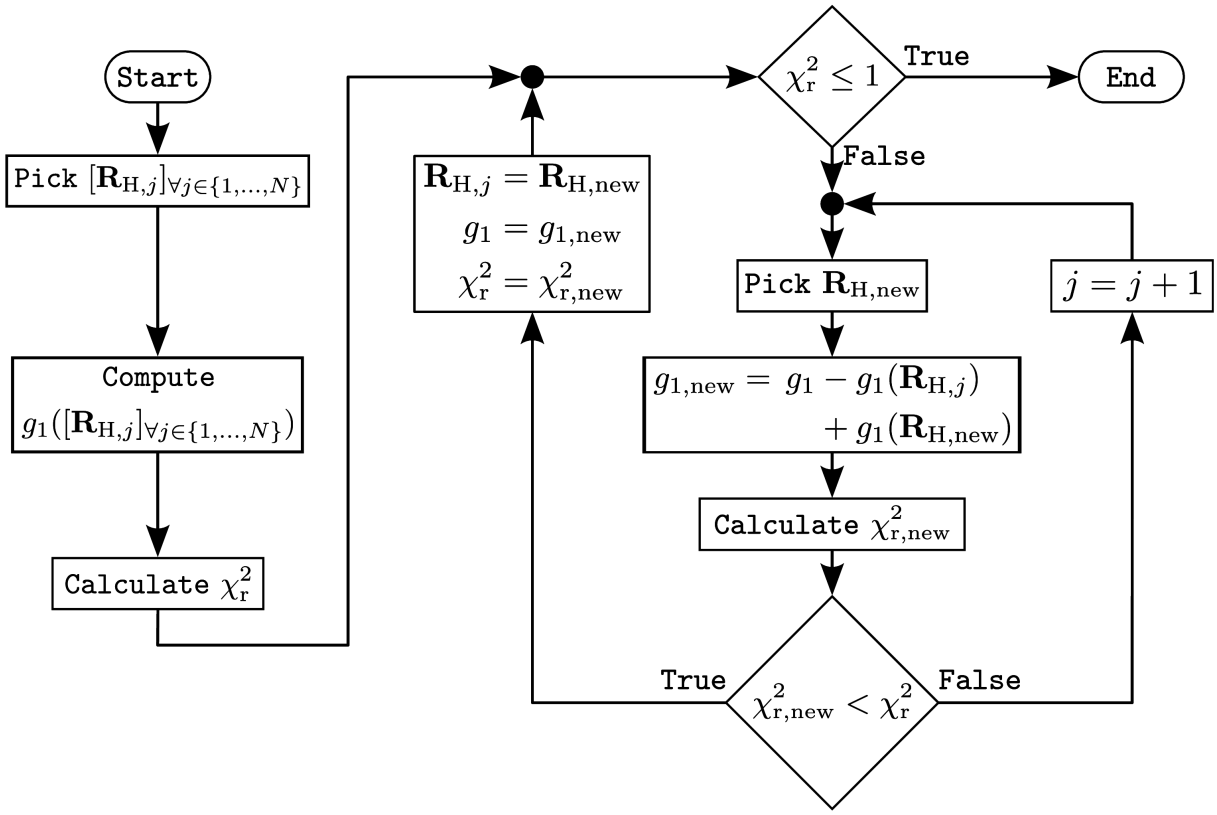

Figure 4.2: The main process of the *McDLS* software for optimization of hydrodynamic radii $f(R_\mathrm{H})$. The process starts with a randomly chosen set of N hydrodynamic radii $[R_{\mathrm{H},j}]_{\forall j\in\{1,\ldots,N\}}$ which is used to calculate the intensity correlation function $G_2([R_\mathrm{H}]) - 1$. Subsequently, it is compared to the measured data by calculating its goodness of fit value χ^2_r. If this value is larger than one and thus not satisfying the convergence criterion, a new single radius $R_{\mathrm{H,new}}$ is chosen randomly to replace $R_{\mathrm{H},j}$ in the set $[R_\mathrm{H}]$. In the following step, an updated correlation function $G_{2,\mathrm{new}}$ is computed and compared to the data by calculating χ^2_r. In each iteration, an attempt is made to replace one of the model contributions of the field correlation functions $g_1(R_{\mathrm{H},j})$ by $g_1(R_{\mathrm{H,\,new}})$ to improve the agreement between the calculated G_2 function and the measured data. This procedure is repeated until no significant difference is found between the modelled and measured G_2 data within the statistical uncertainties, *i.e.* $\chi^2_\mathrm{r} \leq 1$.

4.3. Current implementation

Similar to *McSAS* (section 3.4), the user interface is composed of five panels, each designed for specific tasks of the DLS data analysis as shown in Fig. 4.3. These panels are named '1. Data Files', '2. Data Settings', '3. Optimization', '4. Model' and '5. Post-fit Analysis' and it is recommended to run the program in their consecutive order when starting data analysis. The scientific and technical conceptual aspects lying behind the structure of *McDLS* will be discussed in the following in its numbered order. A summarized description of the program parameters and its effect on the analysis is given in Table 4.2.

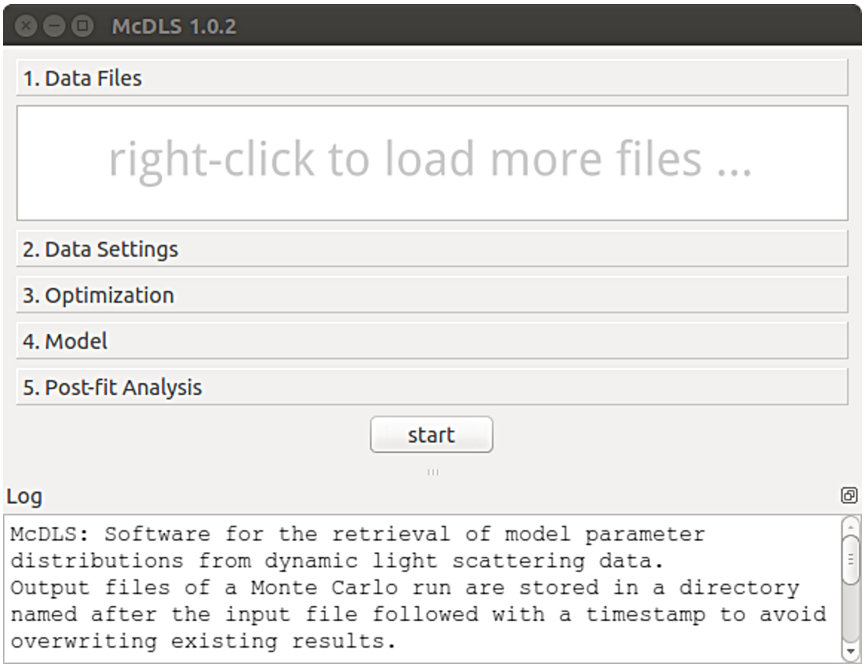

Figure 4.3: The main interface of *McDLS* upon startup showing five configuration panels: The '1. Data Files' panel allows selection and input of the data of interest. The '2. Data Settings' panel allows configuration of the data sets by setting physically meaningful constraints. The '3. Optimization' panel provides basic and advanced settings options useful for orchestrating the simulation procedure. The '4. Model' panel allows to define a range of physically meaningful hydrodynamic radii and decide if a form factor should be taken into account for calculation. The '5. Post-fit Analysis' lets the user configure histogramming of intensity-, volume-, surface- and/or number-weighted distributions.

Panel '1. Data files'

The *McDLS* program reads in `.ASC` data files created by the hardware correlator ALV-7004 (ALV GmbH, Germany) and expects the signature 'ALV-7004 CGS-8F Data, Single Run Data' being present in the header of the file. Moreover, the ASC files store general sample and device properties which were either measured by the device automatically or previously determined and entered by the operator ahead of a measurement. These properties include the temperature, viscosity of the sample, refractive index of the solvent, wavelength of the laser and the eight scattering angles of the multiangle DLS experiment. This information is necessary for calculating the model intensity correlation function as can be seen from eqn. [4.7]. Each data file contains the correlation curves of the eight scattering angles $(\theta_a)_{a\in\{1,\dots,8\}}$ column-wise, preceded by a single column of the corresponding correlation time τ. The original count rates as a function of measurement times for each of the eight angles are contained in the data file as well. The count rate data can be used for detecting and omitting erroneous measurements automatically. Possible reasons for defective measurements are manifold such as dust in the path of light or remains of the sample preparation. Moreover, the particle concentration in the sample being too high or too low may cause erroneous measurements as well. *McDLS* is intended for data evaluation of a broad range of instruments, *i.e.* not only for the ALV instrument used here. Therefore,

users are encouraged to modify the data import herewith adapting to their instrumental requirements.

Regrouping of data

Distribution analysis of DLS data is performed for each scattering angle individually whereas a data file of a single measurement contains correlation curves of multiple scattering angles. To cope with this obstacle, an initial preprocessing of data is provided for grouping correlation curves of the same scattering angle and averaging all curves of each angle of the same sample. This procedure estimates the measurement uncertainties of the correlation curves which are required for calculating the goodness of fit criterion χ^2_r of eqn. [4.21]. Individual outlier curves can be identified and excluded from analysis either automatically or by the user.

Measurement indices

The program shows the measurement indices for each averaged data set of a given scattering angle to make it comprehensible for the user which measurements were averaged and which were left out (see Z-score at panel '2. Data Settings'). This is achieved by setting up the data acquisition program to follow a specific file name pattern when storing the measured data: The basic name, containing the unique experiment number and an arbitrary name describing the sample, followed by a 4-digit index number of the measurement run or group and a trailing 4-digit index number of the individual measurement as shown in Table 4.1.

Table 4.1: File names example of DLS measurements of the silver nanoparticles BAM-N008. Left column: the abbreviated list of files generated by a first multiangle DLS measurement consisting of 6 individual measurements. Mid and right column: File names generated by a second and third measurement run consisting of another 6 individual data files each.

339 N0080000_0001.ASC	339 N0080001_0001.ASC	339 N0080002_0001.ASC
[...]	*[...]*	*[...]*
339 N0080000_0006.ASC	339 N0080001_0006.ASC	339 N0080002_0006.ASC

McDLS extracts the measurement index and group index from the file name and assigns them to each correlation curve when reading in measurement files. The indices of all involved curves are gathered and assigned to the averaged curve. The program groups continuous sequences of indices within a measurement group for display in the GUI. For example, if the first measurement of the first group was filtered as outlier, as well as the second and sixth one of the third group and the third measurement of the fifth group, *McDLS* indicates the involved measurements as "`0'2-6 2'1,3-5 4'1-2,4-6`" (see listing for angle 26° in Fig. 4.4).

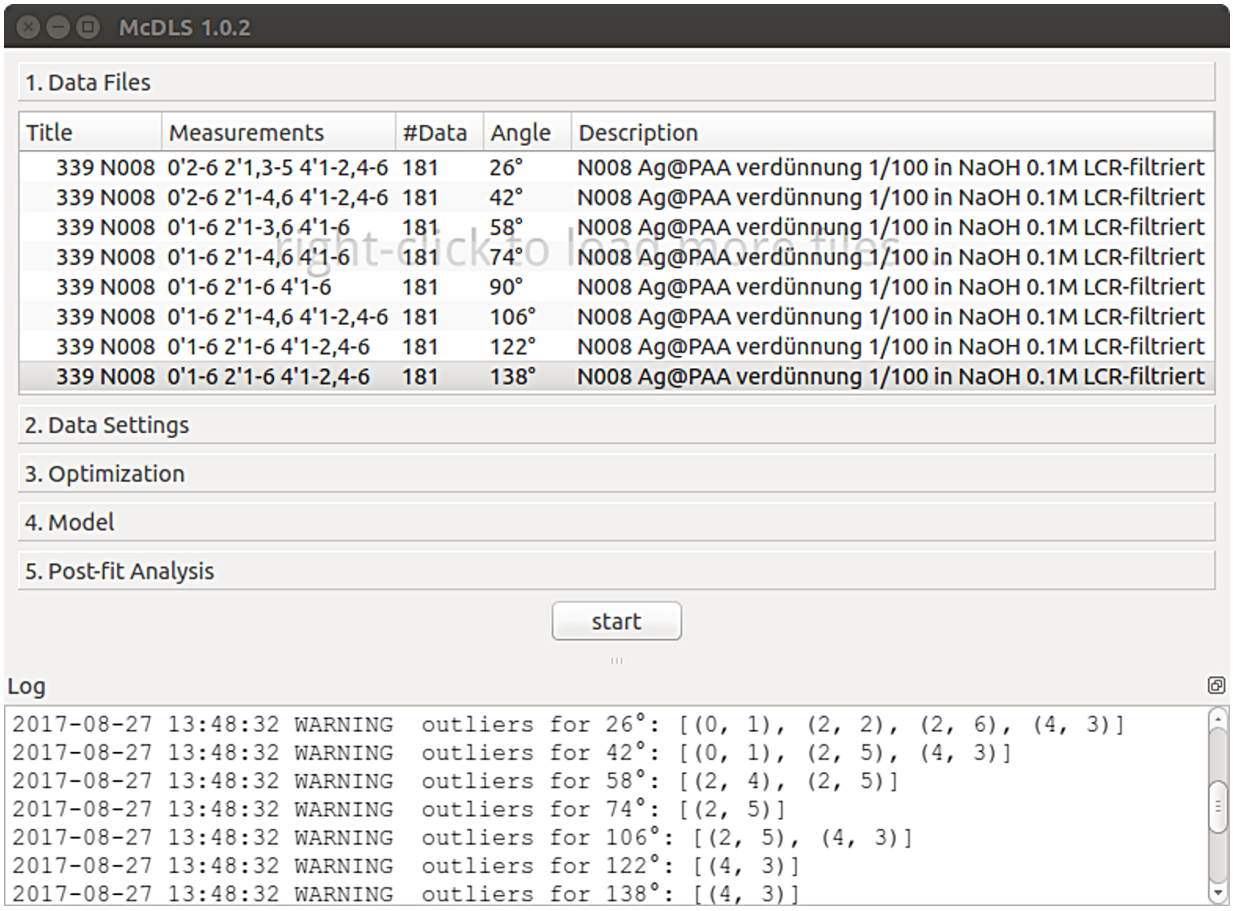

Figure 4.4: *McDLS* listing of regrouped, outlier filtered and averaged multiangle DLS measurement data sets for 8 scattering angles. The 'Measurements' column indicates which data files were included in averaging based on the indices derived from their file names. The log window lists the excluded data files by unique pairs of their group and measurement index.

Panel '3. Optimization'

The third panel contains a subset of algorithm settings addressable by the user. The 'convergence criterion' is the goodness of fit criterion χ^2_r, as described in the model fitting procedure. While this is per default set to 1, it may prevent reaching a state of convergence $\chi^2_r \leq 1$ for data whose uncertainty estimates are too large or too small. In such cases the 'convergence criterion' value needs to be lower or higher than 1, respectively, until the simulations fit the data sufficiently. It should be noted that a change affects the uncertainty estimates of the resultant distribution of hydrodynamic radii. As an alternative to χ^2_r, the program provides 'Test goodness of fit variance?' as optimization criterion. If this checkbox is activated, the variance of goodness of fit values of the last twenty successful improvements has to be reached for completion of an optimization repetition. The results from many repetitions of successful optimization cycles are used to estimate the uncertainties (error bars) of the resultant radii distribution. Its default value of 10 can be decreased (3 to 5) to get a quick result for testing, while it is recommended to be increased to typically 50 to 100 for reliable estimates of the uncertainties. When activating

'Advanced Settings' further default values of several parameters can be changed. The 'number of contributions' of R_H-values is 300 by default and was found sufficiently large for most data. If the number is reduced, the analysis is faster on the cost of accuracy. If it is increased the analysis is slower but with better accuracy. The 'Find background level?' option is activated by default, which means that the program determines the constant A in eqn. [4.15] in accordance to eqn. [4.16]. A correctly determined background level A improves the goodness of fit significantly. In some cases it might not be determined correctly. In this case, its determination can be disabled by this option. The option 'Fix 1^{st} model point to data' is not activated by default but can be activated as constraint by forcing the program to pin the simulated curve at the first experimental value. The first data point is often considered to be of highest relevance for comparing DLS data curves (ISO, 1996). This option allows to evaluate the impact of that assumption on the data evaluation at hand. The 'maximum iterations' option determines how many optimization cycles in a repetition are performed without reaching the convergence criterion before a new repetition starts. The default value of 10^5 can be reduced or increased. Activation of the 'Plot unsuccessful/incomplete results' option allows the user to investigate how simulation and data deviate when the convergence criterion is not reached. The 'Calc. series statistics' option provides access to multiangle DLS data analysis.

Multiangle analysis

Size distributions determined by DLS data analysis often show scattering angle dependence, even with already considered sphere form factor (Shibayama et al., 2006). However, *McDLS* treats a measurement at a certain scattering angle θ as an individual data set, independent of the other scattering angles. To provide an overview of the analysis results across multiple scattering angles, the program supports cumulative series analysis of data sets.

To make use of series analysis, it has to be activated first by checking the box 'Calc. series statistics' in the 'Optimization' menu. Next, one selects multiple entries in the 'Data Files' menu. Each of them corresponds to a measurement at a certain scattering angle. The overall optimization process is started by pressing the 'start' button in the main window of the program. Subsequently, the previously configured model is fitted to the selected data sets, one after another. During this process, the mean and its standard deviation of the resulting distributions are collected and visualised afterwards in a single plot over the selected scattering angles as shown in Fig. 4.5.

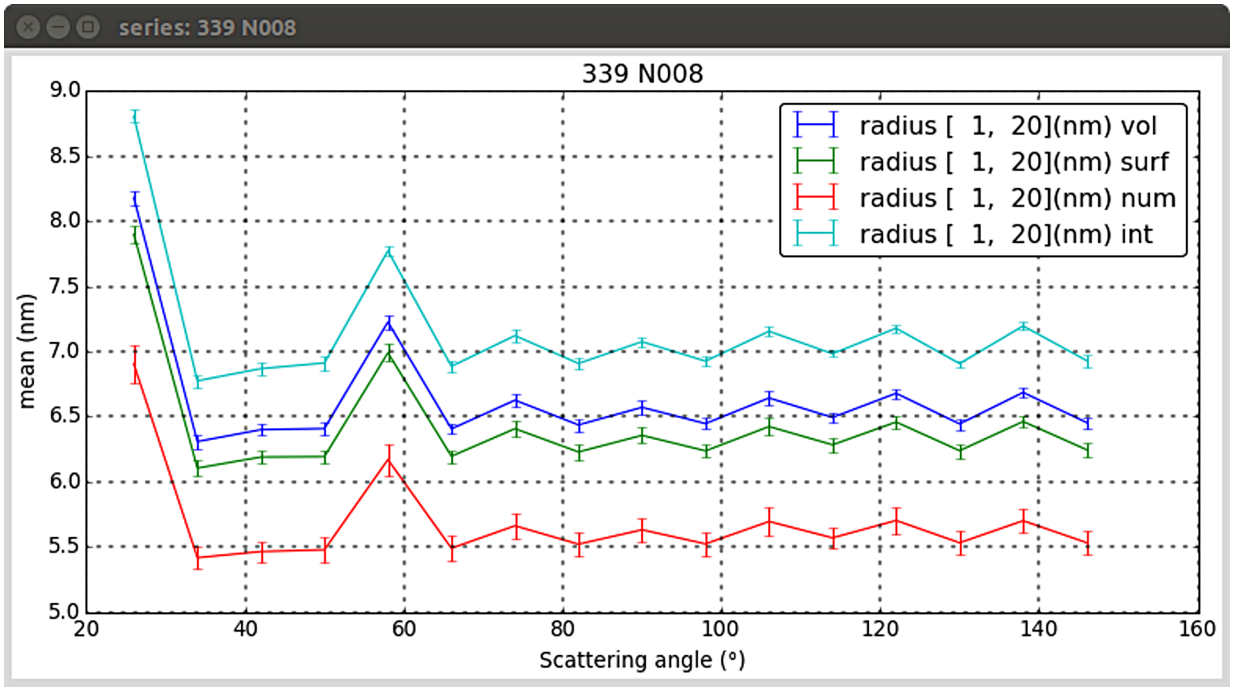

Figure 4.5: *McDLS* plot window of a multiangle analysis result: the volume-, surface-, number- and intensity-weighted size distribution means along with their standard deviations as error bars of an averaged set of 18 measurements at 16 scattering angles of silver nanoparticles BAM-N008 as discussed in section 4.4.1.

Panel '4. Model'

This panel contains all information on the implemented 'Spheres (DLS)' model, for which the program is currently limited but open for adding further models. The 'square the amplitude' is activated by default which means that the particle volume and form factor are squared as in eqn. [4.12]. If deactivated, the amplitude of the field correlation function is used non-squared for data interpretation. This option can be used to assess the effect of the squared amplitude versus the non-squared amplitude. The 'Hydrodynamic Radius (nm)' is activated by default and it is recommended to insert the lowest and the highest expected hydrodynamic radius in the boxes. If the range of meaningful hydrodynamic radii is unknown to the user, it is recommended to deactivate the range. Then the user can explore the range of meaningful radii, by inserting single hydrodynamic radii and click the start button. Now a plot appears with the corresponding intensity correlation function of this single hydrodynamic radius. With some tests of different values the user is able to provide meaningful estimates of the lower and upper bound of possible radii values. The 'Include the SAS sphere form factor' is activated by default and calculates $\Phi(R_{\mathrm{H},j})$ for each radius and takes it into the calculation. Deactivation of this button sets $\Phi(R_{\mathrm{H},j}) = 1$. Comparison of simulation with activated and deactivated form factor allows to study how strong the form factor influences the result.

Panel '5. Post-fit Analysis'

The last panel offers basic analysis capabilities for interpretation of the result. The user can chose different types of histograms for displaying the size distribution within the selected range of hydrodynamic radii. The number of bins has a default value of 50 but can be increased or decreased. A choice can be made whether number-, surface-, volume- and/or intensity-weighted size distribution are calculated. Furthermore, a linear and/or logarithmic scale of the radii axis can be selected. Finally, the 'Start' button starts the simulation process, and the 'log' window shows the output of the program as it runs and stores it in a file automatically. All results are stored in separated ASCII files along with all plots in a PDF file in the directory where the data are located.

Table 4.2: Selected program parameters and their effects on the simulation of intensity correlation functions and the corresponding distribution of the hydrodynamic radii. Only parameters are listed which differ from those described for *McSAS* in Table 3.1.

Location or Panel	Parameter name	Effect
GUI 'Data Files'	'right click to load files'	Loads multiple DLS measurement files in ASC format, regroups, filters, averages and lists them in ascending order, sorted by the scattering angle. The options 'average each angle' and 'outlier threshold' in the 'Data Settings' menu offer more control over data processing.
GUI 'Data Settings'	Mask $G_2(\tau) - 1$ values of 0	Ignores correlation curve values equal to zero during fitting and plotting. Disabled by default.
	Mask negative $G_2(\tau) - 1$ values	Ignores correlation curve values smaller than zero during fitting and plotting. Disabled by default.
	Set uncertainties constant	Sets all uncertainties of the correlation curve to the same value. This option is mutually exclusive with the relative uncertainty option below. Disabled by default.
	Use relative uncertainties $\sigma/[G_2(\tau) - 1]$	Divides all uncertainties of the correlation curve by the corresponding data value to weigh it relatively. This option is mutually exclusive with the constant uncertainties option above. Disabled by default.
	Plot count rate	Plots the mean and median count rate with an error bar for the standard deviation of multiple measurements. It appears behind the correlation plot with grey and pink lines. Enabled by default.
	Average each angle	When loading DLS data files, the correlation data of each angle is averaged if there are multiple measurements with the same sample name. A single data object is created for each angle which contains the correlation mean and its standard deviation, interpreted as measurement uncertainty. If deactivated, the correlation curves are not averaged and a single data object will be created for each combination of measurement and angle. Activated by default.
	Outlier threshold	Decides which measurements are considered as outliers and thus are being excluded from averaging and fitting. The value defines the maximum allowed modified Z-score which uses the median absolute deviation of all measurements at the same angle. The measurement indices displayed in the '1. Data Files' panel indicate which measurements were excluded. Set to 3.5 by default.

(Continued on the next page)

Location or Panel	Parameter name	Effect
GUI 'Optimization'	Test goodness of fit variance?	Instead of trying to reach a given goodness of fit value (convergence criterion), the variance of the last 20 goodness of fit values of successful improvements has to fall below the given value. For experimentally measured DLS data, a value of 10^{-6} is a good starting point.
	Fix 1st model point of data?	Pins the first point of the simulated model data to the first point of the measured data during fitting. Disabled by default.
	Calc. series statistics	Activates multiangle statistics by showing a combined plot of the mean values of all histograms for all data sets being fitted. Additionally, all distribution moments are gathered in a common output file for convenient export. Disabled by default.
GUI 'Model'	Sphere (DLS)	The model of spherical scatterers whose hydrodynamic radii is optimized. The program is open for addition of further models.
	Square the amplitude	By default, the squared amplitude of the field correlation function is used for data interpretation. If this option is deactivated, it allows to investigate the modelled data when not using the square of the particle volume and form factor in the scattering model.
	Hydrodynamic radius	A physically meaningful range of hydrodynamic radii $R_{\mathrm{H,min}}$ and $R_{\mathrm{H,max}}$ should be given as constraint. For example, radii in the range of 1 to 100 nm. The raised 'Active' button deactivates the optimization and leaves a single input field for the R_{H} to compute the model curve of and plot it with the data, once the 'Start' button is pressed.
	Include the sphere form factor	The angular dependence of the scattering of particles is taken into account for calculation of the intensity correlation function using the form factor of a sphere. The influence of the form factor on the retrieved radii distribution can be investigated by deactivating this option.

Outlier filtering

Outliers of the correlation curves are a typical appearance in DLS, and need to be identified and rejected from further data interpretation. The author of this work decided to make use of the count rate $CR(t)$ associated to the correlation curves. The modified Z-score (Iglewicz and Hoaglin, 1993) of $CR(t)$ is compared to the Z-score of count rates from other measurements at the same scattering angle. The count rate quantifies the number of photons detected in a time interval (usually in units of photons per second). The $CR(t)$ is stored in the ALV data file along with the correlation data of all angles measured. The modified Z-score Z_m of measurement m out of M measurements is calculated by:

$$Z_m = \frac{0.6745(x_m - \tilde{x})}{median(|x_m - \tilde{x}|)} \text{ with } x_m = CR_m(t). \quad [4.22]$$

Where $\tilde{x}$ denotes the median count rate of all measurements $[x_m]_{\forall m \in \{1,\ldots,M\}}$ at a certain scattering angle. A correlation curve is excluded from averaging, if its score $Z_m \geq Z$, with Z being a user configurable threshold, initially set to 3.5 as suggested by Iglewicz and Hoaglin (1993). The value of 3.5 may not be suitable for all data and it is in the responsibility of the user how strictly data are considered as outliers. Therefore, the user has the possibility to adjust the Z-score values with respect to the DLS data under investigation. Further discussion of Z-score thresholds can be found in Shiffler (1988) and the reference therein. The Z-score is also used in the interlaboratory size comparison of silica nanoparticles (Lamberty et al., 2011), with data sets being considered as outliers if their Z-score is outside [-3,+3].

In *McDLS*, the median count rate is plotted in the background of the correlation curve plot (pink line), along with the averaged count rate (grey line) and its sample standard deviation as error bars. This can be disabled by unchecking the option 'plot count rate' in the '2. Data Settings' panel of the GUI. Which measurements were rejected is not only indicated by the data listing in the GUI but it is also written to the log window. For example, a rejected fifth measurement of the third measurement group at a scattering angle of 74°, results in the output '`WARNING outliers for 74°: [(2, 5)]`'. While a rejected fourth and fifth measurement of the third measurement group at a scattering angle of 58°, results in '`WARNING outliers for 58°: [(2, 4), (2, 5)]`' (see Fig. 4.4).

4.4. Interpretation of experimental data and simulations

The usage of McDLS is demonstrated in the following by interpretation of typical examples of experimental data. It is complemented by simulations of monomodal and bimodal distributions of hydrodynamic radii.

4.4.1. Monomodal Distributions: Polymer-stabilized silver nanoparticles

Silver nanoparticles are of scientific interest, but they are also widespread in consumer products mostly because of their antibacterial properties (Walczak et al., 2012; Calderón-Jiménez et al., 2017). DLS is a key method to determine their hydrodynamic radii. As a representative example, monomodal polyacrylic acid stabilized silver nanoparticles used as nano-catalyst (Kästner and Thünemann, 2016) were employed. These particles are a candidate of a nanoscale reference material (Stefaniak et al., 2013) labelled 'BAM-N008'. The original BAM-N008 suspension was diluted with distilled water (adjusted to pH 10 with NaOH) to a silver concentration of 3×10^{-2} g/L. Twenty DLS measurements of 60 s duration were performed at 16 scattering angles of $26° \leq 2\theta \leq 146°$ ($0.006\,\mathrm{nm}^{-1} \leq q \leq 0.0253\,\mathrm{nm}^{-1}$), at a temperature of 19.827(4) °C.

For comparison, the R_H-values were determined first by following the standard ISO 13321 which is applicable for particles of narrow size distribution (ISO, 1996). This standard is based on the methods of cumulants derived by Koppel (1972) and provides estimates of an intensity-weighted hydrodynamic radius and a polydispersity index PI. The PI is defined as the ratio of the variance of the relaxation rates (second cumulant = μ) to the square of the mean relaxation rate (first cumulant $\bar{\Gamma}$): $\mathrm{PI} = \mu/\bar{\Gamma}^2$. The width of the distribution of the hydrodynamic radii can be estimated as $\sigma_\mathrm{H} = R_\mathrm{H}\sqrt{PI/2}$, when assuming a Gaussian distribution of relaxation rates (ISO, 1996). A plot of the hydrodynamic radius and the PI of the particles as a function of the scattering vector is shown in Fig. 4.6. The R_H values decrease slightly with decreasing q-values from 6.9 nm at $q = 0.025\,\mathrm{nm}^{-1}$ ($2\theta = 146°$) to 6.2 nm at $q = 0.008\,\mathrm{nm}^{-1}$ ($2\theta = 34°$). A linear extrapolation results in $R_\mathrm{H}(q = 0) = 6.0(2)$ nm. Note that the value of $R_\mathrm{H} = 56$ nm at the lowest q-value of $q = 0.005\,\mathrm{nm}^{-1}$ ($2\theta = 26°$), indicated by an arrow, is rejected as an outlier. The PI values are strongly dependent on the angle of measurement. They are below 0.05 for $0.025\,\mathrm{nm}^{-1} \leq q \leq 0.016\,\mathrm{nm}^{-1}$ but increase for the measurements at lower q-values and display a maximum of PI = 0.18 at $q = 0.008\,\mathrm{nm}^{-1}$ ($2\theta = 34°$).

The angular dependency of the R_H-values determined by the method of cumulants is small but significant. Such finding is typically interpreted as an indication of particle interaction and an extrapolation to zero angle is commonly employed for revealing the 'true' R_H-value (Xu, 2002). Here the difference is $\Delta R_\mathrm{H} = 0.9$ nm between the measured value $R_\mathrm{H}(q = 0.0025\,\mathrm{nm}^{-1})$ and the extrapolated value $R_\mathrm{H}(q = 0)$. More striking than the systematic changes of R_H is the increase of the PI value with decreasing measurement angle. Low PI values mistakenly may lead to the conclusion of a very narrow size distribution when measuring at large angles, typically for instruments that measure only at one high angle of 173°, for example. A considerable variability of PI values (Finsy, Jaeger, et al., 1992) and the problems to derive size distribution widths thereof, have been reported in the literature (Hanus and Ploehn, 1999). The second virial coefficient and therefore the PI are considered

only meaningful by several authors if the corresponding scattering angle, normally 90°, is provided (Koshkina et al., 2015). This indicates that PI values can not be compared when they result from measurements at different angles. A R_H distribution apparently seems to be narrower when DLS is performed at greater scattering angles, in contrast to smaller ones. The reason for this has been discussed by Xu (2002) as follows: Because $G_2(\tau) - 1$ is a function of q, all cumulants are also a function of q. At small τ, for particulate systems that have size distributions with PI < 0.3, the method of cumulants provides reasonable good mean values of the first cumulant but the uncertainty of the second cumulant is substantially higher (see also the original work on the method of cumulants by Koppel (1972)). It is stated that the baseline values must be correctly determined better than 0.1% to avoid large errors in the PI value (Xu, 2002).

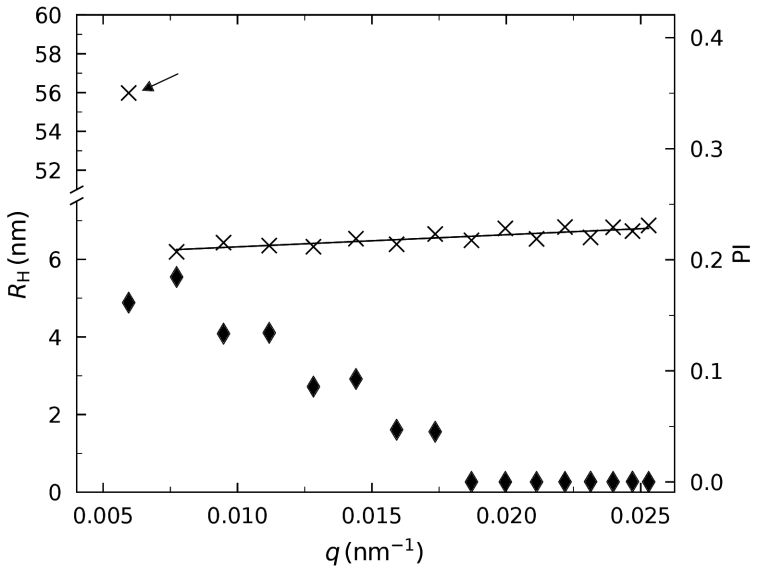

Figure 4.6: Results of the classical cumulant fit for the hydrodynamic radii and polydispersity indices (crosses and diamonds, respectively) of silver nanoparticles BAM-N008. The straight line is a linear fit of the R_H-values for extrapolation to $R_H(q = 0) = 6.0(2)$ nm.

Based on the results from the method of cumulants, it is of interest to examine whether *McDLS* is a more robust data evaluation procedure for these silver nanoparticles. Obviously, the particles' real size distribution should not change when 'looking' at the particles from different angles. Therefore, the same DLS data as employed for the method of cumulants were reanalysed with *McDLS*.

The data was configured to omit the first two data points (panel '2. Data Settings'), resulting in the correlation curve starting at $\tau = 3 \times 10^{-7}$ s. This is typically done for excluding any boundary effect from the analysis, possibly caused by the type of hardware correlator used here. The uncertainties based on the sample standard deviation of the twenty correlation curves were limited to at least 1% of the correlation value. For optimization (panel '3. Optimization'), the convergence criterion of the goodness of fit value (χ^2_r, eqn. [4.21]) was set to 1 along with 50 repetitions for uncertainty estimates.

The number of contributions was kept at the default value of 300. For the sphere model (panel '4. Model') the default settings were used as well, expect for the radius parameter which was set to be fitted in the range 1 nm to 20 nm.

The resulting number-, volume- and intensity-weighted size distributions are shown for $q = 0.02\,\mathrm{nm}^{-1}$ ($2\theta = 90°$) in the upper row of Fig. 4.7. The results for all angles are summarized in the lower row of Fig. 4.7. The mean number-, volume- and intensity-weighted R_H-values increase in the line 5.60(18) nm, 6.55(22) nm and 7.04(24) nm. The corresponding size distribution widths are 1.58(6) nm, 1.08(4) nm and 1.02(4) nm, *i.e.* the relative widths decrease in the line 0.28, 0.17 and 0.15. The determined order of sizes is $R_\mathrm{num} \leq R_\mathrm{vol} \leq R_\mathrm{int}$ and that of the relative widths is $\sigma_\mathrm{R,num} \leq \sigma_\mathrm{R,vol} \leq \sigma_\mathrm{R,int}$ as must be expected. As a first conclusion, it can be stated that this *McDLS* derived size distribution characteristics are angular-independent and seem to be physically meaningful.

Simulation of monomodal correlation data

Next, a simulation of the monomodal distributions could provide insights on possible biasing artefacts of *McDLS*. Therefore, as a test of consistency of the upper results, the distribution was fitted with a suitable analytical function. In the next step, this function was used with the fitted parameters to generate 20 intensity correlation functions, each at eight angles, for a set of N = 182 correlation times corresponding to those of the instrumental ALV correlator used.

For producing reasonable sets of hydrodynamic radii, the lognormal size distribution was employed. It is defined as

$$f_1(R_\mathrm{H}) = \frac{N}{\sqrt{2\pi}\sigma R_\mathrm{H}} \exp\left(-\frac{\ln(R_\mathrm{H}/R_\mathrm{median})^2}{2\sigma^2}\right). \qquad [4.23]$$

Where σ is the width parameter of the size distribution and R_median is the median radius. Using these parameters, the mean radius is given as $R_\mathrm{mean} = R_\mathrm{median}\exp(\sigma^2/2)$ and the standard deviation of the width of the size distribution is $\sigma_\mathrm{width} = R_\mathrm{mean}\sqrt{\exp(2\sigma^2) - \exp(\sigma^2)}$.

For simulation purposes, the correct weighting of the distribution needs to be considered. If, for example, eqn. [4.23] is assumed to be the number-weighted size distribution then the volume-weighted distribution scales with R_H^3 and after area normalization, it becomes

$$f_2(R_\mathrm{H}) = \frac{R_\mathrm{H}^3}{R_\mathrm{median}^3} \exp(-9\sigma^2/2) f_1(R_\mathrm{H}). \qquad [4.24]$$

The intensity-weighted distribution, f_1 and scaled by R_H^6, is

$$f_3(R_H) = \frac{R_H^6}{R_{median}^6} \exp(-18\sigma^2) f_1(R_H). \quad [4.25]$$

Contrary, if eqn. [4.23] is found to represent the intensity-weighted size distribution, then the volume-weighted distribution, based on f_1 and scaled by R_H^{-3}, becomes

$$f_4(R_H) = \frac{R_{median}^3}{R_H^3} \exp(-9\sigma^2/2) f_1(R_H), \quad [4.26]$$

and the number-weighted distribution, based on f_1 and scaled by R_H^{-6}, is

$$f_5(R_H) = \frac{R_{median}^6}{R_H^6} \exp(-18\sigma^2) f_1(R_H). \quad [4.27]$$

Equations [4.24] to [4.27] show that not only the conversion of the lognormal distribution from number- via volume- to intensity-weighting is possible but also from intensity-weighting via volume- to number-weighting. It must be noted, the second scenario cannot be performed for a Gaussian distribution. It should also be mentioned that the error propagation at the conversion of particle size distributions is currently under discussion (Babick and Ullmann, 2016).

In DLS the intensity-weighting must be considered as the most reliable one (Hassan et al., 2015) because DLS is intrinsically weighted by the intensity of the scattered light (Kestens, Roebben, et al., 2016). Consequently, the parameters derived by fitting eqn. [4.23] to the intensity-weighted distribution were used, resulting in $R_{median} = 7.01(2)$ nm, width parameter $\sigma = 0.128(2)$, $R_{mean} = 7.07(2)$ nm and $\sigma_{width} = 0.91(2)$ nm as shown in the upper plot of Fig. 4.7 (black solid line). The fit curve represents the experimentally determined size distribution within the experimentally determined uncertainties. The course of the curves f_1 to f_5 is shown in the left-hand picture of Fig. 4.8. The determined fit parameters seem to be reasonable and thus were used in the simulation by employing f_5 of eqn. [4.27] to evaluate eqn. [4.13]. Herewith 20 synthetic data sets were produced, each for 182 correlation times at sixteen angles. An example is shown in the right-hand picture of Fig. 4.8. These data sets were evaluated with *McDLS* in the same way as the experimental data, resulting in size distributions shown exemplarily for the 90° angle (see symbols of the upper graph of Fig. 4.7). It can be seen, that number-, volume- and intensity-weighted size distributions are well reproduced for the 90° example. Averaging the results from simulation of the data at the eight angles, the mean number-, volume- and intensity-weighted R_H-values are found in the line 5.72(1) nm, 6.60(1) nm and 7.00(1) nm. The corresponding size distribution widths are 1.55(1) nm, 1.00(0) nm and 0.89(1) nm, *i.e.* the relative widths decrease in the

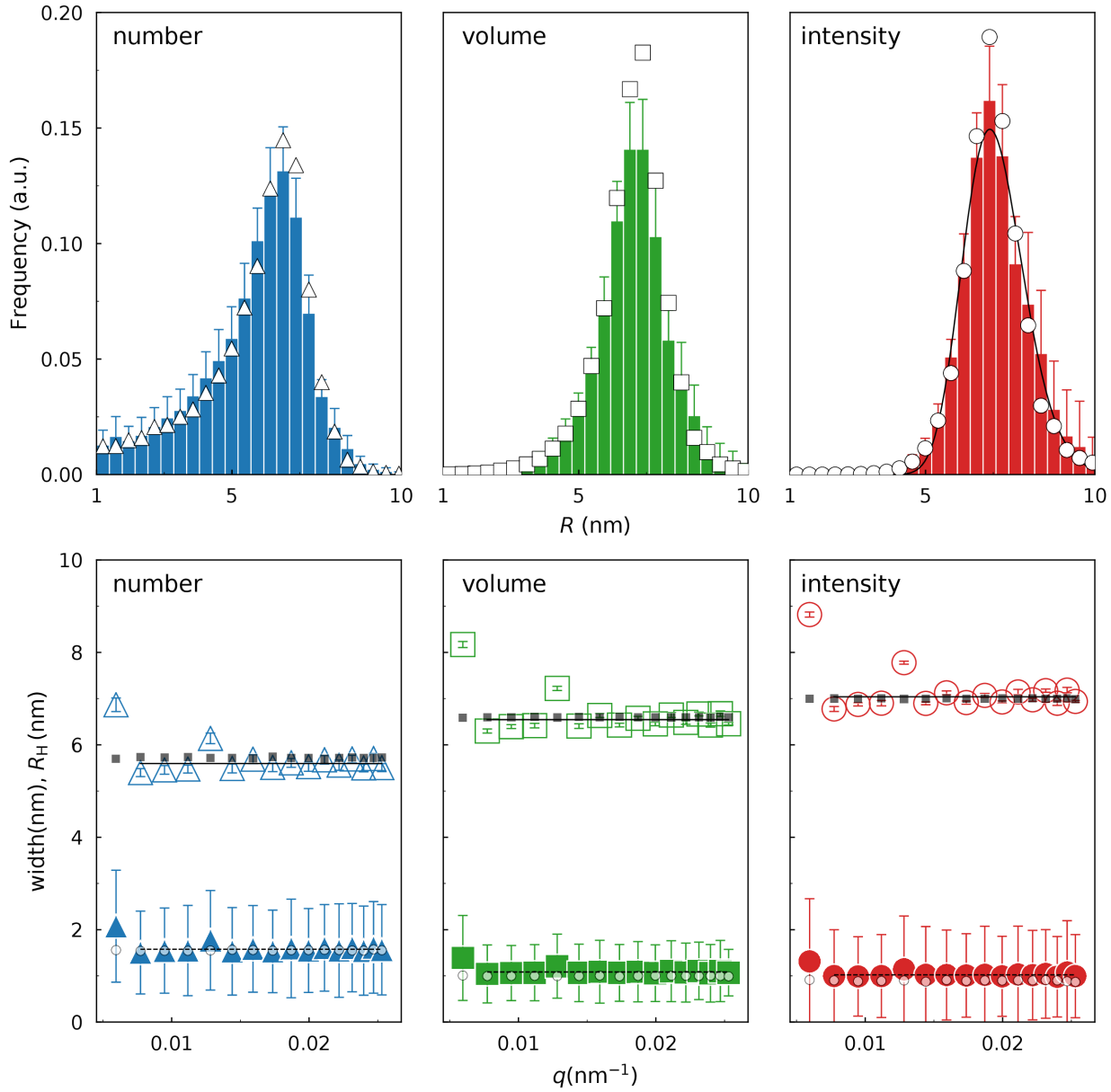

Figure 4.7: Fitting results of the monomodal silver nanoparticles BAM-N008, using 50 repetitions of the Monte Carlo algorithm to determine their number-, volume- and intensity-weighted distribution of hydrodynamic radii. Upper row: resulting size distributions at an scattering angle of 90° with uncertainties. The histograms with error bars are based on experimental data. The best fit curve of a lognormal f_1 function according to eqn. [4.23] represents the intensity-weighted distribution (black solid line). The symbols are distributions resulting from simulated $G_2 - 1$ data derived from the parameters of the fit curve. Lower row: mean hydrodynamic radii and mean widths for multiangle data analysis (large hollow and solid symbols, respectively). The mean values of the means of the hydrodynamic radii and widths are given as solid and dashed lines, respectively. The mean values of the simulations are given as small symbols. All data are listed in Table 4.3.

line 0.27, 0.15 and 0.13. These values are plotted as small symbols in the lower graph of Fig. 4.7. A summary of these results is given in Table 4.3 from which it can be concluded that the evaluation of the simulated data is in agreement with the evaluation of the experimental data within the estimated uncertainties.

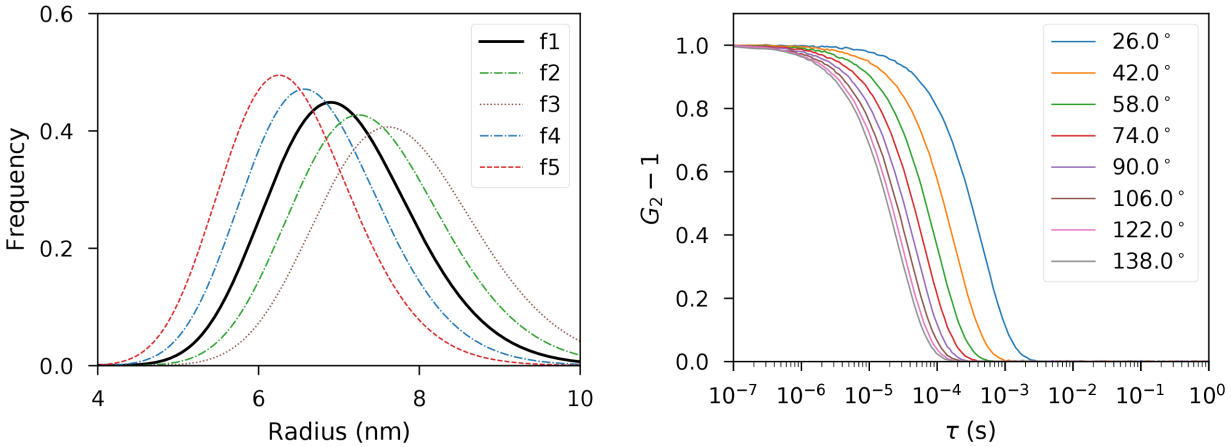

Figure 4.8: Left-hand figure: functions f_1 to f_5 according to eqn. [4.23] to [4.27] with the same lognormal distribution parameters $R_{\text{median}} = 7.01(2)$ nm and $\sigma = 0.128(2)$, which were derived from a curve fit of the experimentally derived intensity-weighted radii distribution as shown in the most right graph of the top row of Fig. 4.7. Right-hand figure: simulated intensity correlation function according to eqn. [4.13] for eight angles employing function f_5.

Table 4.3: Summary of the hydrodynamic radii values of silver nanoparticles BAM-N008.

	Number		Volume		Intensity	
data	R_n (nm)	σ_n (nm)	R_v (nm)	σ_v (nm)	R_i (nm)	σ_i (nm)
measurement	5.60(18)	1.58(6)	6.55(22)	1.08(4)	7.04(24)	1.02(4)
simulation	5.72(1)	1.55(1)	6.60(1)	1.00(0)	7.00(1)	0.89(1)

4.4.2. Bimodal size distributions: Silicon dioxide particles

The bimodal silica certified reference material ERM-FD102 was used as described by Kestens, Roebben, et al. (2016). An overview of the sample characteristics is also given by Breßler et al. (2015) and listed for clarity in Table 2.1.

The DLS measurements were performed using a multi-angle ALV 7004 device from ALV Langen, equipped with a He-Ne laser (wavelength: 632.8 nm). The ERM-FD102 samples were used as delivered without dilution and filtering to prevent changes of the samples. Twenty DLS measurements of 60 s duration were performed at 8 scattering angles of $2\theta \in \{26°, 42°, 58°, 74°, 90°, 106°, 122°, 138°\}$ at a temperature of 22.960(3) °C.

For evaluation with *McDLS*, the same parameters were utilised as for the analysis of the monomodal silver nanoparticles, except for the convergence criterion: it was set to a variance of 10^{-6} of the last twenty successful improvements of the goodness of fit value (χ_r^2, eqn. [4.21]) as stopping criterion. Instead of running the optimization until the given goodness of fit value is reached, the algorithm stops once the last 20 accepted goodness of fit values did not have a higher variance than the specified value. It also results in a slightly different χ_r^2 value finally reached, for each of the 50 repetitions calculated. The radius parameter of the sphere model (panel '4. Model') was set to be fitted in the range 1 nm to 70 nm.

The resulting number-, volume- and intensity-weighted size distributions are shown exemplarily for $q = 0.02\,\mathrm{nm}^{-1}$ ($2\theta = 90°$) in the upper row of Fig. 4.9 and Fig. 4.10 for the large and small particle classes B and A, respectively. The results for all angles are summarized in the lower rows of Fig. 4.9 and Fig. 4.10. For particle class B, the mean number-, volume- and intensity-weighted R_H-values increase in the line 44.98(78) nm, 45.18(79) nm and 45.37(80) nm. The corresponding size distribution widths are 1.72(10) nm, 1.74(10) nm and 1.72(10) nm with relative widths of 0.04. Accordingly, the mean number-, volume- and intensity-weighted R_H-values of particle class A increase in the line 18.60(30) nm, 18.63(28) nm and 18.65(27) nm. The corresponding size distribution widths are 0.38(11) nm, 0.37(10) nm and 0.36(10) nm with relative widths of 0.02. The determined order of sizes is $R_\mathrm{num} \leq R_\mathrm{vol} \leq R_\mathrm{int}$ and that of the relative widths is $\sigma_\mathrm{R,num} \leq \sigma_\mathrm{R,vol} \leq \sigma_\mathrm{R,int}$ as must be expected.

Similar to the results of the monomodal analysis, the *McDLS*-derived size distribution characteristics are found to be angular-independent and seem to be physically meaningful. However, two distinct particle populations centered at 45.0(8) nm and 18.6(3) nm with a volume ratio of approximately 7:1 were identified. While the large particles are found in good agreement with the certified value of 41.7(6) nm for the number-weighted median area-equivalent radii, the size of the small particles differs significantly from the certified and expected values of 9.1(4) nm. Due to the smaller particles being found nearly twice

in radius then expected, the determined volume ratio between both populations does not match the expected volume ratio of 20:1 as well.

The discrepancy between the analysis results and the certified distribution characteristics was investigated. To test the influence of the choice of the initial radius contributions, the same *McDLS* evaluation was performed but with the 'startFromMinimum' option activated. This lets all initial contributions start with their radii set to one half of the lower limit of the radius fitting range (0.5 nm). The resulting volume-weighted size distribution at a scattering angle of 90° is shown in the lower row of Fig. 4.12. The small particles are centered at 16.0(27) nm and the large particles at 41.7(3) nm. While the mean size of the large particles agrees slightly better to the certified value, the sizes of the small particles are still much larger than expected and broader distributed compared to the first evaluation. Qualitatively, the analysis result did not change significantly when initialising all radius contributions with a constant small value.

Simulation of bimodal correlation data

Possible systematic errors of the method can be revealed by a simulation of the bimodal size distributions and an evaluation of synthetic photon correlation data generated thereof, following the same procedure as for the monomodal silver nanoparticle.

For producing reasonable sets of hydrodynamic radii composed of two size classes, a superposition of the lognormal size distribution of eqn. [4.23] was employed. It is defined as

$$\begin{aligned} f_{12}(R_\mathrm{H}, N_1, \sigma_1, R_{1,\mathrm{median}}, N_2, \sigma_2, R_{2,\mathrm{median}}) = \quad & f_1(R_\mathrm{H}, N_1, \sigma_1, R_{1,\mathrm{median}}) \\ + \quad & f_1(R_\mathrm{H}, N_2, \sigma_2, R_{2,\mathrm{median}}) \end{aligned} \qquad [4.28]$$

The constants N_1, σ_1 and $R_{1,\mathrm{median}}$ define the first class of particle sizes, while N_2, σ_2 and $R_{2,\mathrm{median}}$ define the second class of particle sizes. Using these parameters, the mean radius and the standard deviation of the width of the size distribution can be calculated in the same way as stated for the simulation of monomodal distribution data.

Correspondingly, the correct weighting of the distribution is taken into account by employing one of the following adjusted functions. If, for example, eqn. [4.28] is assumed to be the number-weighted size distribution then the volume-weighted distribution scales with R_H^3 and after area normalization, it becomes

$$f_{22}(R_\mathrm{H}) = \frac{(N_1 + N_2) R_\mathrm{H}^3}{N_1 R_{1,\mathrm{median}}^3 \exp(9\sigma_1^2/2) + N_2 R_{2,\mathrm{median}}^3 \exp(9\sigma_2^2/2)} f_{12}(R_\mathrm{H}) \qquad [4.29]$$

The intensity-weighted distribution, f_{12} and scaled by R_H^6, is

$$f_{32}(R_H) = \frac{(N_1 + N_2)R_H^6}{N_1 R_{1,\text{median}}^6 \exp(18\sigma_1^2) + N_2 R_{2,\text{median}}^6 \exp(18\sigma_2^2)} f_{12}(R_H) \qquad [4.30]$$

However, if f_{12} of eqn. [4.28] is found to represent the intensity-weighted size distribution, then the volume-weighted distribution, based on f_{12} and scaled by R_H^{-3}, becomes

$$f_{42}(R_H) = \frac{(N_1 + N_2)R_H^{-3}}{N_1 R_{1,\text{median}}^{-3} \exp(9\sigma_1^2/2) + N_2 R_{2,\text{median}}^{-3} \exp(9\sigma_2^2/2)} f_{12}(R_H) \qquad [4.31]$$

and the number-weighted distribution, based on f_{12} and scaled by R_H^{-6}, is

$$f_{52}(R_H) = \frac{(N_1 + N_2)R_H^{-6}}{N_1 R_{1,\text{median}}^{-6} \exp(18\sigma_1^2) + N_2 R_{2,\text{median}}^{-6} \exp(18\sigma_2^2)} f_{12}(R_H). \qquad [4.32]$$

Following the considerations on simulating monomodal distributions, the parameters were derived by fitting f_{12} of eqn. [4.28] to the intensity-weighted distribution. Later on, the parameters are used in the simulation by employing f_{52} of eqn. [4.32] to evaluate eqn. [4.13] at eight angles. For the large size class, fitting f_{12} found $R_{\text{median}} = 45.01(14)$ nm, width parameter $\sigma = 0.049(3)$, $R_{\text{mean}} = 45.07(14)$ nm and $\sigma_{\text{width}} = 2.23(1)$ nm as shown in the upper right plot of Fig. 4.9 (black solid line). Accordingly for the small size class, $R_{\text{median}} = 18.80(14)$ nm, width parameter $\sigma = 0.045(38)$, $R_{\text{mean}} = 18.81(14)$ nm and $\sigma_{\text{width}} = 0.84(1)$ nm was found as shown in the upper right plot of Fig. 4.10 (black solid line). The fit curve represents the experimentally determined size distribution within the experimentally determined uncertainties. In Fig. 4.11 the course of the curves f_{12} to f_{52} is shown on the left and an example of the simulated curves is shown on the right.

The simulated data sets were evaluated with *McDLS* in the same way as the experimental data, resulting in size distributions shown exemplarily for the 90° angle by symbols in the upper graph of Fig. 4.9 and Fig. 4.10. For the 90° example, it can be seen, that the number-, volume- and intensity-weighted size distributions of the large size class are shifted slightly towards smaller sizes while the small particle class is shifted towards larger sizes. The same effect was observed with the size distributions obtained from measured data. Averaging the results from simulation of the data at the eight angles, the mean number-, volume- and intensity-weighted R_H-values for the large size class are found to be in the line 43.14(35) nm, 43.52(31) nm and 43.84(34) nm. The corresponding size distribution widths are 2.07(35) nm, 2.23(52) nm and 1.86(24) nm, *i.e.* the relative widths are 0.05, 0.05 and 0.04. For the small size class, the mean number-, volume- and intensity-weighted R_H-values are found to be in the line 20.43(30) nm, 20.45(30) nm and 20.48(30) nm. The corresponding size distribution widths are 0.44(11) nm, 0.41(9) nm

and 0.40(8) nm and the relative widths are 0.02. These values are plotted as small symbols in the lower graph of Fig. 4.9 and Fig. 4.10. A summary of these results is given in Table 4.5.

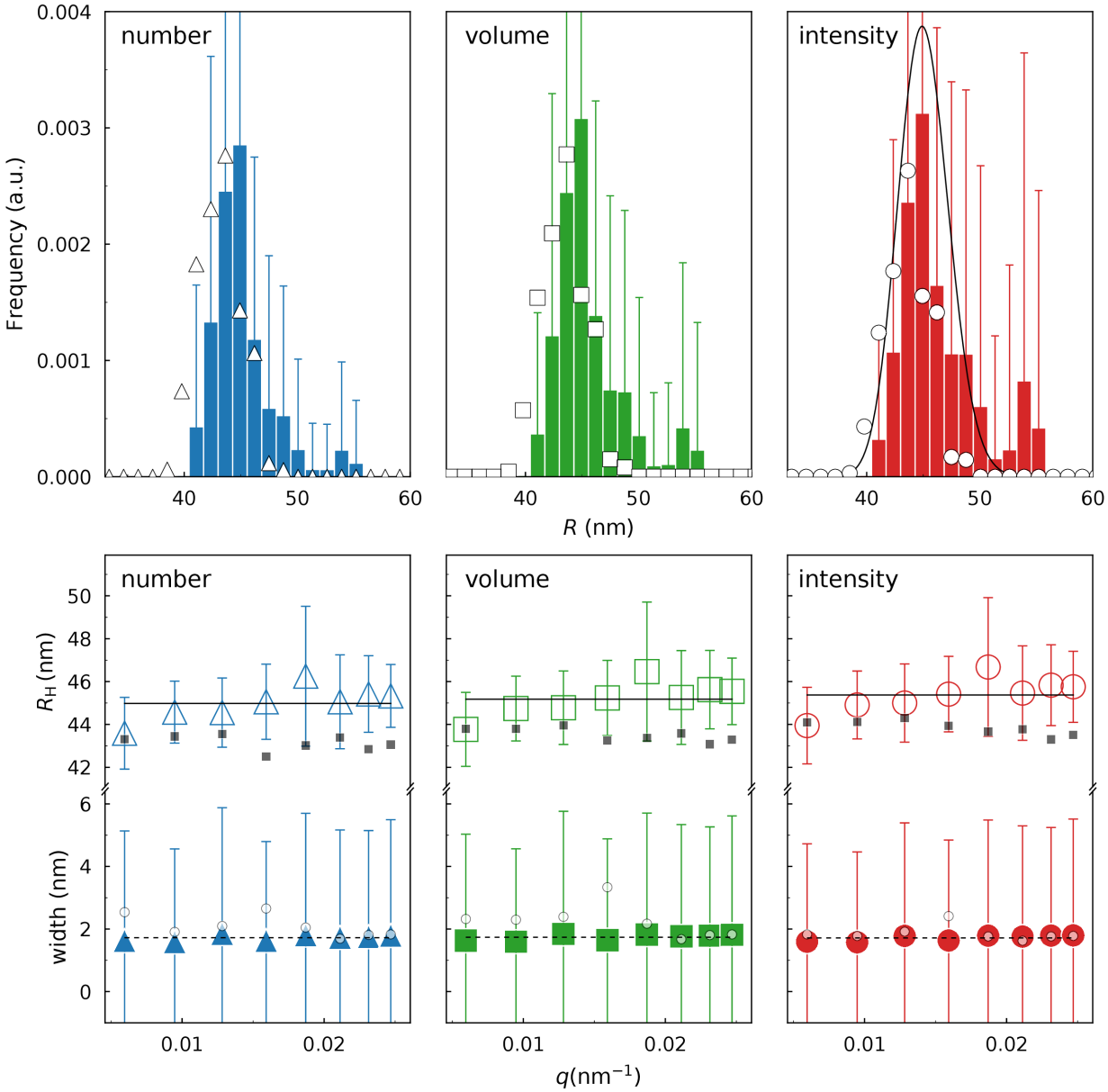

Figure 4.9: Fitting results of the large particle class B of bimodal silica nanoparticles ERM-FD102, using 50 repetitions of the Monte Carlo algorithm to determine their number-, volume- and intensity-weighted distribution of hydrodynamic radii. Upper row: resulting size distributions at an scattering angle of 90° with uncertainties. The histograms with error bars are based on experimental data. The best fit curve of a lognormal f_{12} function according to eqn. [4.28] represents the intensity-weighted distribution (black solid line). The symbols are distributions resulting from simulated $G_2 - 1$ data derived from the parameters of the fit curve. Lower row: mean hydrodynamic radii and mean widths for multiangle data analysis (large hollow and solid symbols, respectively). The mean values of the means of the hydrodynamic radii and widths are given as solid and dashed lines, respectively. The mean values of the simulations are given as small symbols. All data are listed in Table 4.5.

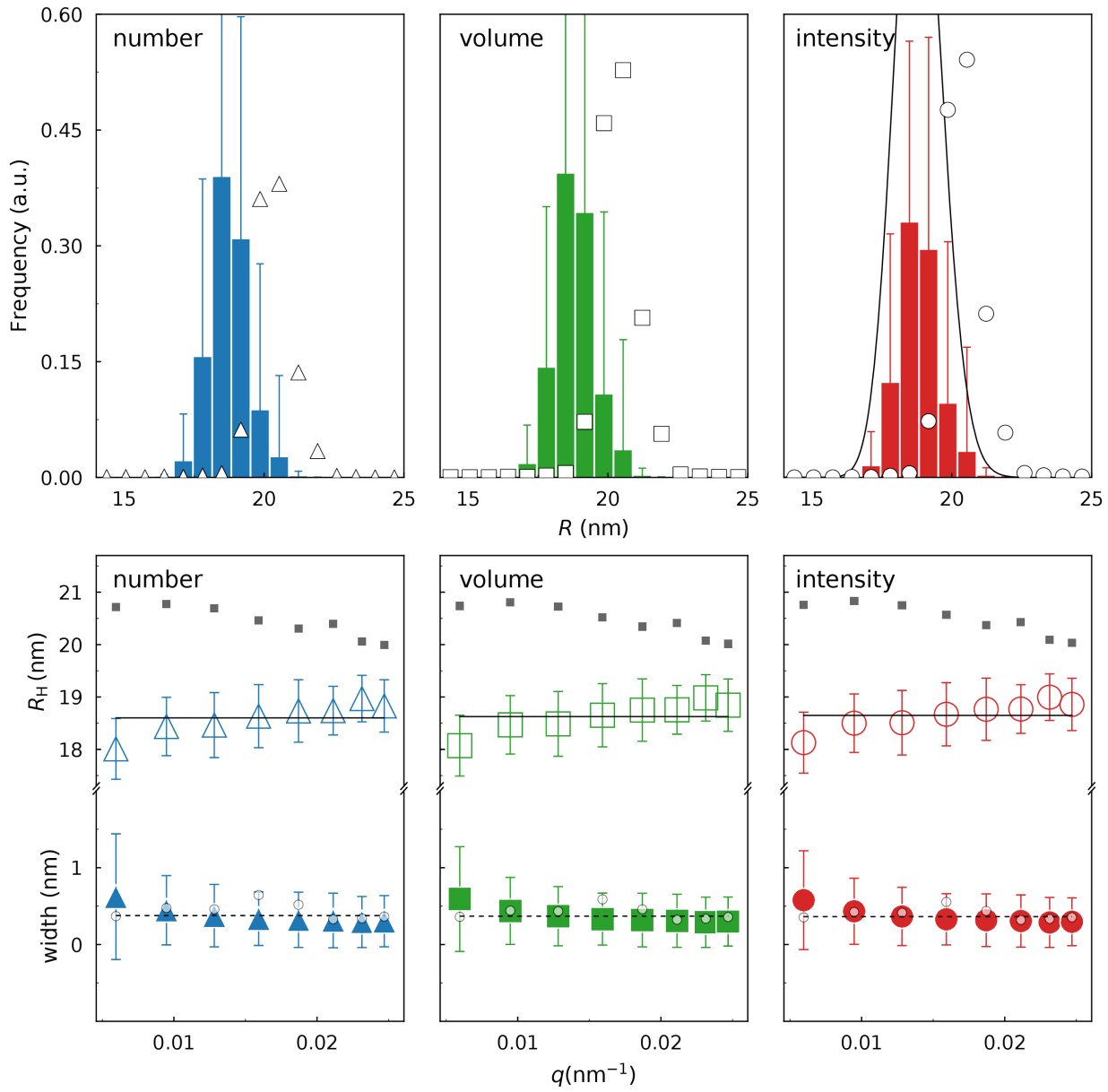

Figure 4.10: Fitting results of the small particle class A of bimodal silica nanoparticles ERM-FD102. For a description, please see Fig. 4.9.

Table 4.5: Summary of the hydrodynamic radii values of silica nanoparticles ERM-FD102.

	Number		Volume		Intensity	
data	R_n (nm)	σ_n (nm)	R_v (nm)	σ_v (nm)	R_i (nm)	σ_i (nm)
class A measurement	18.60(30)	0.38(11)	18.63(28)	0.37(10)	18.65(27)	0.36(10)
class A simulation	20.43(30)	0.44(11)	20.45(30)	0.41(9)	20.48(30)	0.40(8)
class B measurement	44.98(78)	1.72(10)	45.18(79)	1.74(10)	45.37(80)	1.72(10)
class B simulation	43.14(35)	2.07(35)	43.52(31)	2.23(52)	43.84(34)	1.86(24)

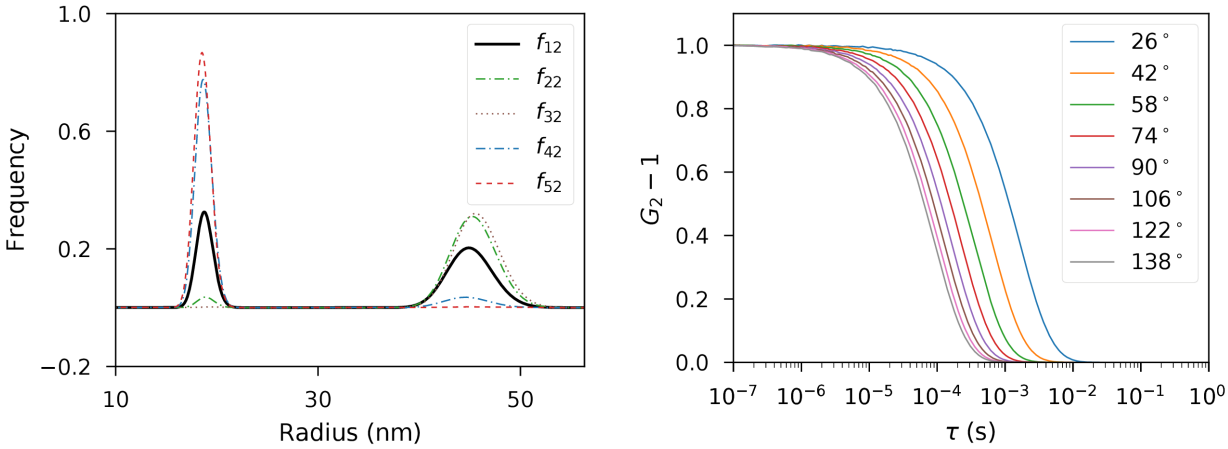

Figure 4.11: Left-hand figure: functions f_{12} to f_{52} according to eqn. [4.28] to [4.32] with the same lognormal distribution parameters $R_{\text{median}} = 45.01(14)\,\text{nm}$ and $\sigma = 0.049(3)$ for the large size class and $R_{\text{median}} = 18.80(14)\,\text{nm}$ and $\sigma = 0.045(38)$ for the small size class. Those parameters were derived from a curve fit of the experimentally derived intensity-weighted radii distribution as shown in the most right graph of the top row of Fig. 4.9 and Fig. 4.10. Right-hand figure: simulated intensity correlation function according to eqn. [4.13] at eight angles employing function f_{52} of eqn. [4.32].

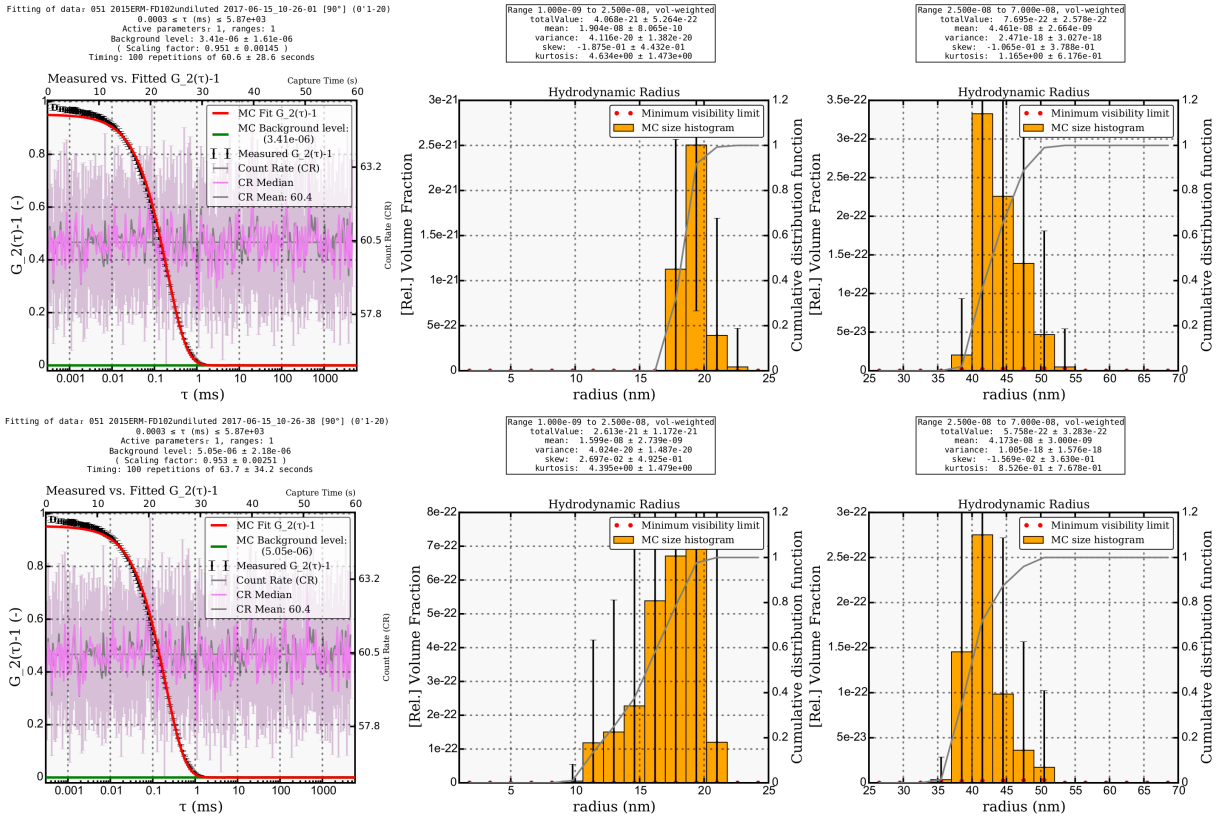

Figure 4.12: Fitting results of the undiluted bimodal reference material ERM-FD102, using 100 repetitions of the Monte Carlo algorithm to determine a distribution of hydrodynamic radii. Upper row: the fitting procedure started with an uniform random distribution of 150 hydrodynamic radii. Lower row: the fitting results of a fitting procedure starting with a constant value of $\frac{1}{2}R_{\text{min}} = 0.5$ nm for 150 hydrodynamic radii, using the same measurements. Left column: the determined model correlation curve above the averaged curve of 20 measurements of the same sample with uncertainties (red solid line and black dots with error bars, respectively). Mid column: the determined volume-weighted distribution histogram of the small particle class with an expected mean radius of 9.1(4) nm. Right column: the determined volume-weighted distribution histogram of the large particle class with an expected mean radius of 41.7(6) nm.

Consideration of noise

The consideration of noise for simulation of DLS data is very diverse in the literature. Roger et al. (2016) added a noise term to the field correlation function, which is constant in q but varies with τ. They argued that this corresponds to the noise from their correlator (Brookhaven BI-9000). The dependence of noise on the delay time in the statistics of photon counting has also been discussed theoretically by Schätzel (1990) and was experimentally studied by Harrison and Fisch (1996). An early study on the noise of photon correlation functions and its effect on data reductions algorithms was provided by R. Peters (W. Brown (1993) (§3)) but its relevance for modern photon correlators is unclear. Recently, Mailer et al. (2015) tried to generate 'realistic' noise of modern photon correlators for their simulated data by multiplying each value of the intensity correlation function by a random number drawn from a Gaussian distribution of mean 1 and standard deviation σ_{noise}. They used $\sigma_{\text{noise}} = 10^{-3}$ for most of their analysis, corresponding to an uncertainty of one part in a thousand at each data point. They also looked at noisier data, up to $\sigma_{\text{noise}} = 10^{-2}$ but found this unrealistically high.

Since the literature does not agree on how to consider noise properly, the bimodal silicium dioxide particle reference material ERM-FD102 was measured and the accuracy of the intensity correlation curves as functions of scattering angle and correlation time was determined. 20 data sets with different random noise were produced, allowing calculation of the mean and standard deviation at each correlation time. Uncertainty estimates for the values of the baseline A and the coherence factor B could be determined as well. The result is presented in Fig. 4.13. The upper figure shows a relatively high absolute noise level for low correlation times with values of about 10^{-2} and decreases strongly towards high correlation times, where it typically is around 10^{-4}. Furthermore, the noise level decreases slightly with increasing scattering angle, *i.e.* the noise is largest at the lowest angle of 26° and is smallest at the highest angle of 138°.

The author of this work believes that a simulation of a 'realistic' noise level for the experiments at hand can be performed by multiplying each value of the coherence factor by a random number drawn from a Gaussian distribution of mean 1 and standard deviation σ_{noise} as $B = 1 + \sigma_{\text{noise},B}$ and for the baseline as $A = 0 + \sigma_{\text{noise},A}$. Here, for simulation a $\sigma_{\text{noise},B} = 10^{-5}$ and $\sigma_{\text{noise},A} = 10^{-3}$ was applied.

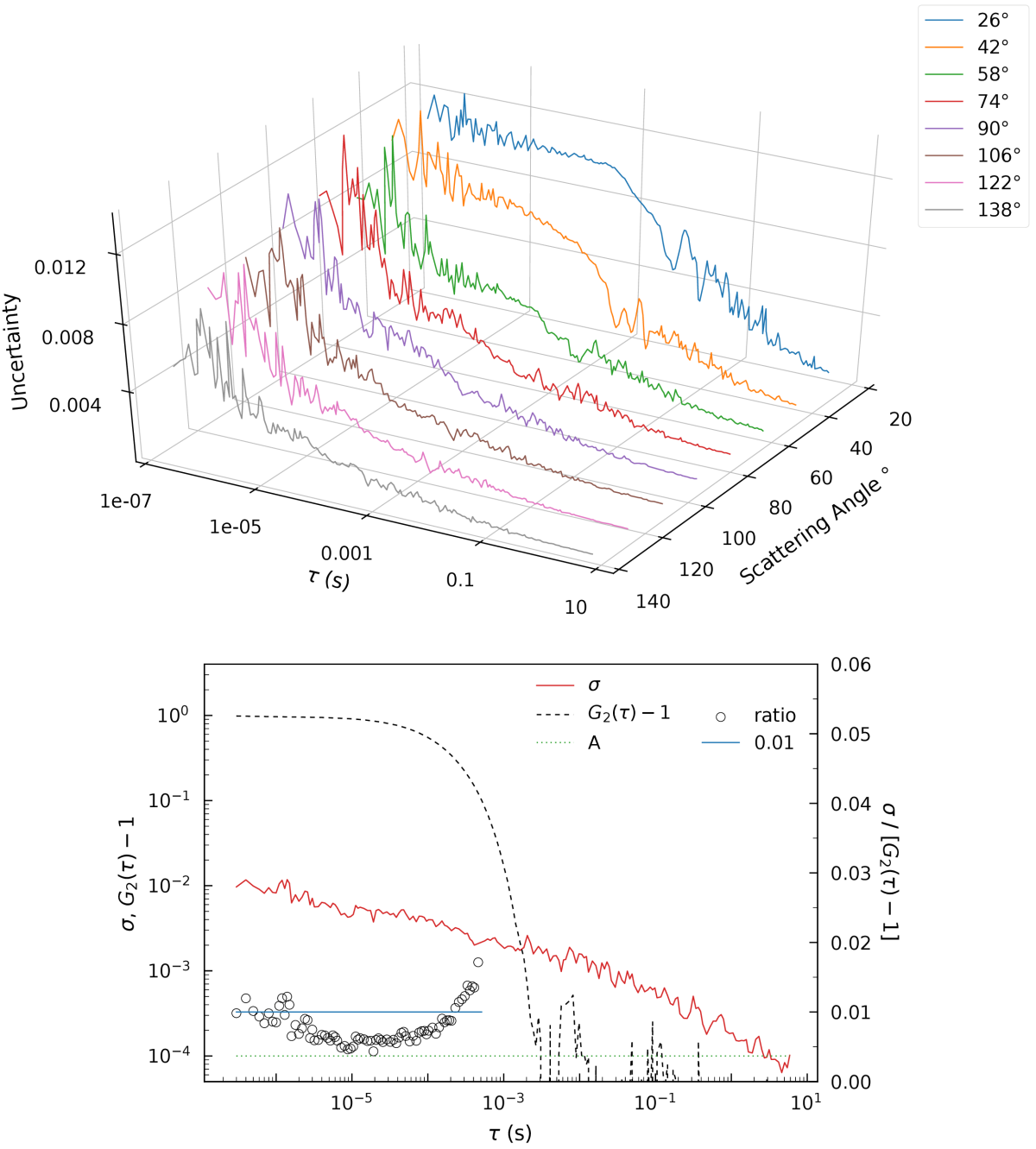

Figure 4.13: Upper figure: The typical mean of uncertainties derived from 20 $G_2(\tau) - 1$ measurements of the same bimodal silica ERM-FD102 sample at eight scattering angles. Lower figure: An example of an averaged $G_2(\tau) - 1$ function from twenty measurements of the undiluted ERM-FD102 bimodal silica nanoparticles at a scattering angle of 90° (black dashed line) and the measurement uncertainties based on the sample standard deviation σ (red solid line). The ratio of σ to $G_2(\tau) - 1$ is approximately 0.01 or smaller (symbols and blue solid line, respectively) for correlation times where $G_2(\tau) - 1$ is greater than 10% of $G_2(\tau_{\min}) - 1$ (approx. down to $\tau = 0.5$ ms). For τ-values larger than 3 s the σ decreases below $A = 10^{-4}$ (green dotted line).

4.5. Conclusion

The program package McDLS was developed with the goal to overcome limitations of existing methods at reliably determining multimodal size distributions from multiangle DLS data. It implements a Monte Carlo regression method which was adapted for the specific requirements of evaluating DLS measurements with a spherical scatterer model. To test the reliability of the method, simulated and experimentally measured DLS data of monomodal and multimodal particle ensembles are analysed. The results were compared with the expected and certified values, respectively.

Regarding monomodal particle ensembles, it can be concluded that *McDLS* determined the size distribution of the experimentally measured silver nanoparticles BAM-N008 as expected: the distribution characteristics are angular-independent and seem to be physically meaningful. Additionally, to verify the consistency of the algorithm, DLS correlation data was simulated based on the previously determined size distributions from measured data with the goal to obtain the same distribution again by *McDLS* evaluation. This consistency test showed that the evaluation of the simulated data is in agreement with the evaluation of the experimental data within the estimated uncertainties.

Given DLS measurements of bimodal particle ensembles, the *McDLS* program clearly identifies separate size distributions qualitatively. The accuracy in determining the size distribution of the larger particle ensemble is good compared to the certified and expected characteristics. However, the method has difficulties with finding the expected size of the smaller particle ensemble in presence of larger particles. Additionally, both resulting particle size classes are slightly shifted towards each other. The smaller class more than the larger one; nevertheless, the latter matches the distribution characteristics in good agreement with the expected and certified values.

Considering that the method is designed to avoid any assumption on the characteristics of the distribution sought after and given the difficulties with correctly determining bimodal distributions, it has to be questioned whether the correlation data used for the analysis of DLS measurements contains enough information required for sufficiently retrieving multimodal distribution characteristics. This is supported by the poor performance of existing methods at precisely determining multimodal size distributions from DLS data.

5

Summary and Outlook

5.1. Summary and Conclusions

The developments on three program packages were presented in this work with a focus on the quantitative determination of particle size distributions. Two of them are designed to analyse small-angle scattering (SAS) data: *SASfit* represents the classical curve fitting toolbox which finds optimal parameters of a scattering model composed of analytical expressions. *McSAS* implements an alternative approach by employing a Monte Carlo regression technique for determining form-free parameter distributions of a scattering model. Finally, *McDLS* was developed over the course of this work to employ the Monte Carlo regression technique for analysing multiangle dynamic light scattering (DLS) data and exploring its capabilities and limitations.

As a classical tool, *SASfit* was already employed in many scientific fields concerning the analysis of scatterers from SAS data on a nanoscale. Several improvements of the program were presented in this work which support the user and improve the quality of analysis results. It was shown that *SASfit* is able to characterise a monomodal gold nanoparticle reference material in reasonable agreement with the certified properties. Additionally, the precision and accuracy of the parameters and associated values determined from measurements of silica reference material ERM-FD102 are in good agreement with the certified values. For assisting the user in finding optimal model parameters, several data reduction algorithms, supplemented by advanced goodness of fit measures and a detailed presentation of the confidence intervals for the fitted parameters were developed and presented in this work. Moreover, the functionality of user provided models as plug-ins was implemented and an example was given on how to create such a custom model in the program which is of interest for advanced users with the need to adjust an existing model or implement a new model for their specific research subject. As a mature and comprehensive data analysis tool, *SASfit* defines the standard range of functionality, competitive tools must provide. However, missing the concept of thorough preparation of uncertainties along the analysis results, it encouraged the development of an alternative analysis method providing the desired statistics.

This work showed that *McSAS* is a reliable and versatile tool for nanoparticle analysis, capable of determining particle size distributions with uncertainty estimates. This was demonstrated by retrieving bimodal size distributions of the reference material solution ERM-FD102 in agreement with the certified values. Additionally, a dry powder of densely packed nanoparticles was characterised correctly as well. Moreover, *McSAS* provides a comprehensive model library which can be extended with additional models by the user. An increasing availability of computing power neutralises the relatively high computational effort of the Monte Carlo method, especially with more complex custom models. Limitations are induced by the information content provided by the SAS measurement method: further information on the sample under investigation is always required for choosing appropriate models and obtaining unique solutions for parameter distributions. Overall, *McSAS* proved to be invaluable for quantitatively determining particle size distributions of nanoparticles from SAS measurements.

With *McDLS*, an attempt was made to transfer the valuable properties of the Monte Carlo approach to the analysis of nanoparticles from DLS measurements. The program was derived from *McSAS* with all necessary modifications required for the evaluation of DLS data. Thus, it offers many identical features which assist with the workflow of scientific data analysis, such as selecting a data range, filtering outlier measurements, averaging data, tweaking optimization parameters and flexible configuration of volume-, number-, intensity- or surface-weighted output histograms over customizable parameter ranges. The reliability of *McDLS* was verified by evaluating measurements and simulations of DLS data of dispersed monomodal silver nanoparticles BAM-N008. It was shown that the determined distributions are consistent with the expected characteristics. Tests with simulated data are in agreement with the results from experimental data. Consecutively, the analysis capabilities of the program were tested by evaluating measurements and simulations of DLS data of bimodal particle ensembles. It was found that the larger class of particles was characterised in agreement with the expected and certified values. Whereas the size of the smaller class of particles was found larger than expected. That results in a diverging volume ratio between both particle classes. Evaluations based on simulated data for the same configuration of particles ensembles confirmed those results. It can be concluded that *McDLS* correctly determines monomodal size distributions from DLS data, but only characterises bimodal distributions qualitatively. The method has difficulties with finding the expected size of the smaller particle ensemble in presence of larger particles. Based on that, it can be argued, that the DLS correlation data used for the analysis might not contain enough information for sufficiently retrieving multimodal distributions.

Three different tools for nanoparticle analysis were presented in this work. While each of them can be used for basic evaluation of reference nanoparticles, they were developed with a certain use case in mind, employing a specific method for handling it at the best. Together with other analysis tools available for nanoparticle research, they will gain relevance and

attention in the future, due to a growing range of industrial applications for nanoparticles and a concurrently growing need for control and regulation.

5.2. Outlook

Since the early versions of the *SASfit* program, the Levenberg-Marquardt algorithm (Levenberg, 1944) has been used to find solutions for multi-dimensional nonlinear optimization problems. Users often experience stability issues or sometimes even crashes of the optimization routine of *SASfit*, in particular when optimizing several parameters of a complex model at once. Those issues may be caused either by correlated parameters within a model (though these are hard to predict) or by instabilities of the optimization algorithm. Further development of the *SASfit* software will, therefore, provide alternative optimization algorithms. These will include modern versions of the Levenberg-Marquardt algorithm with improved numerical stability over the original implementation. Additional options for parameter constraints will improve the overall fit stability, in contrast to the current implementation which interrupts a fit if parameter values leave their defined range. Better minimization routines may be able to automatically account for that as well. This will improve the overall workflow and user experience with the *SASfit* analysis program.

For *McSAS*, several structural improvements are conceivable with a gain in efficiency and flexibility with scientific analyses. Comparing the program with other methods, the computational expensive approach is a limitation which could be reduced significantly by employing parallel processing capabilities of current computers. With a steadily increasing number of cores in current and upcoming processing units, it will be possible to run sufficient multiple optimizations at once, resulting in meaningful statistics in the same time which is required for a single optimization at the current state of the program. Furthermore, since its release, there were several requests of users who want to integrate *McSAS* in a larger, possibly server based, data analysis framework. Currently, that use case does not work out, because the optimization routine requires some parts of the graphical user interface. A modularization which divides the program into self-contained modules, would enable further development and a wider spectrum of the aforementioned applications. This modular approach would automatically lead to a standalone result browser, capable of re-interpreting previously obtained results using different visualisation options. More specifically, it would add new value to existing results by allowing to generate new histograms over customizable ranges and thus exploring already existing results from a new perspective without requiring to run the time consuming Monte Carlo regression again.

Employing *McDLS* with the Monte Carlo approach on the analysis of DLS data proved to be capable of resolving monomodal and to some degree bimodal parameter distributions.

That makes it interesting to adapt it to other related data acquisition methods, such as Fluorescence Correlation Spectroscopy (FCS) or X-ray Photon Correlation Spectroscopy (XPCS). Based on a similar correlation curve, the same principle of data evaluation can be applied. Being useful for a large range of measurement types stresses the importance of investigating the limits of the method for multimodal distributions more in-depth. A comparison with other established analysis tools, such as CONTIN, will shed more light on the performance of *McDLS* and the general limitations for this kind of data.

6

Appendix

6.1. Availability, support and licensing

6.1.1. *SASfit*

All *SASfit* information, the source code repository and downloadable items, can be found at http://sasfit.org. The most recent version of the comprehensive manual is in the repository (http://sasfit.org/blob/master/doc/manual/sasfit.pdf) and additional user guides on creating plug-ins for custom model functions can be found under http://docs.sasfit.org. For core topics in using the *SASfit* program there are also video guides available online (https://www.youtube.com/user/SASfitTeam). The latest packages are available at the repository page (http://sasfit.org/releases).

The numerical part of *SASfit* is written in the C programming language and the user interface in Tcl/Tk. Previous development efforts also included numerous measures to make the program platform-independent. Standalone packages which do not require any additional prerequisites are provided for three main desktop operating systems: Windows, Linux and MacOS. The software is released under the open-source license GPLv3, allowing for academic and commercial adoption given proper attribution. Users for whom the software has been useful may refer to Breßler et al. (2015).

6.1.2. *McSAS*

All *McSAS* information, including instructions and downloadable items, are available through http://mcsas.net. A reasonable degree of support is provided by the author subject to the availability of time. Instructional videos are available to help the user get started.

The software is written in the Python programming language and available as a Git DVCSS repository. The interpreted code should run without issue on any desktop computer running any operating system that supports Python 2.7. It has been tested on the three main operating systems: Windows, Linux and MacOS. Standalone packages of stable versions are also available for these operating systems, which do not require

Python to be available on the host computer. The software is released under an open-source license GPLv3, allowing for academic and commercial adoption given proper attribution. Users for whom the software has been useful may refer to Bressler et al. (2015).

6.1.3. *McDLS*

All *McDLS* information, including instructions and downloadable items, are available through `http://mcdls.net`. As the software is based on *McSAS*, it shares the technical basis and the points regarding availability, support and licensing mentioned above. Users for whom the software has been useful may refer to an upcoming paper in preparation by the author of this work.

7

References

[1] B. Abécassis, C. Bouet, C. Garnero, D. Constantin, N. Lequeux, S. Ithurria, B. Dubertret, B. R. Pauw, D. Pontoni, 'Real-Time in Situ Probing of High-Temperature Quantum Dots Solution Synthesis', *Nano Letters* **2015**, *15*, 2620, DOI 10.1021/acs.nanolett.5b00199 (cited on page 5).

[2] F. Babick, C. Ullmann, 'Error propagation at the conversion of particle size distributions', *Powder Technology* **2016**, *301*, 503, DOI 10.1016/j.powtec.2016.06.039 (cited on page 68).

[3] M. Berghaus, M. Paulus, P. Salmen, S. Al-Ayoubi, M. Tolan, R. Wintert, 'Near-Surface and Bulk Behavior of Bicontinuous Microemulsions under High-Pressure Conditions', *Journal of Physical Chemistry B* **2016**, *120*, 7148, DOI 10.1021/acs.jpcb.6b05639 (cited on page 7).

[4] P. R. Bevington, D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, 3rd edition, **2003**, 320 pages (cited on page 8).

[5] S. Bhattacharjee, 'DLS and zeta potential – What they are and what they are not?', *Journal of Controlled Release* **2016**, *235*, 337, DOI 10.1016/j.jconrel.2016.06.017 (cited on page 3).

[6] H. Borchert, E. V. Shevchenko, A. Robert, I. Mekis, A. Kornowski, G. Grübel, H. Weller, 'Determination of Nanocrystal Sizes: A Comparison of TEM, SAXS, and XRD Studies of Highly Monodisperse CoPt3 Particles', *Langmuir* **2005**, *21*, 1931, DOI 10.1021/la0477183 (cited on page 1).

[7] *Soft-Matter Characterization*, (Eds.: R. Borsali, R. Pecora), Springer Science & Business Media, **2008**, 1490 pages, http://link.springer.com/referencework/10.1007/978-1-4020-4465-6 (cited on page 47).

[8] R. Botet, B. Cabane, 'Simple inversion formula for the small-angle X-ray scattering intensity from polydisperse systems of spheres', *Journal of Applied Crystallography* **2012**, *45*, 406, DOI 10.1107/S0021889812012812 (cited on page 4).

[9] R. H. Bragg, I. Corvin, J. W. Buttrey, 'Small Angle Scattering Investigation of Ludox Mixtures', *Journal of Applied Physics* **1960**, *31*, 1183, DOI 10.1063/1.1735799 (cited on page 1).

[10] B. van den Brandt, H. Glättli, P. Hautle, J. Kohlbrecher, J. A. Konter, A. Michels, H. B. Stuhrmann, O. Zimmer, 'Creating local contrast in small-angle neutron scattering by dynamic nuclear polarization', *Journal of Applied Crystallography* **2007**, *40*, s106, DOI 10.1107/S0021889806052149 (cited on page 8).

[11] I. Bressler, B. R. Pauw, A. F. Thünemann, 'McSAS: software for the retrieval of model parameter distributions from scattering patterns', *Journal of Applied Crystallography* **2015**, *48*, 962, DOI 10.1107/S1600576715007347 (cited on page 86).

[12] I. Breßler, J. Kohlbrecher, A. F. Thünemann, 'SASfit: a tool for small-angle scattering data analysis using a library of analytical expressions', *Journal of Applied Crystallography* **2015**, *48*, 1587, DOI 10.1107/S1600576715016544 (cited on pages 14, 24, 43, 71, 85).

[13] D. Brown, P. Kienzle, Neutron Activation and Scattering Calculator, **2015**, https://www.ncnr.nist.gov/resources/activation/ (visited on 03/23/2017) (cited on page 30).

[14] W. Brown, *Dynamic Light Scattering - The method and some applications*, Oxford University Press, Oxford, **1993** (cited on pages 3, 77).

[15] R. K. Bryan, 'Maximum entropy analysis of oversampled data problems', *European Biophysics Journal* **1990**, *18*, 165, DOI 10.1007/BF02427376 (cited on page 4).

[16] G. Bryant, J. C. Thomas, 'Improved Particle Size Distribution Measurements Using Multiangle Dynamic Light Scattering', *Langmuir* **1995**, *11*, 2480, DOI 10.1021/la00007a028 (cited on page 4).

[17] P. Buchold, R. Schweins, Z. Di, M. Gradzielski, 'Structural behaviour of sodium hyaluronate in concentrated oppositely charged surfactant solutions', *Soft Matter* **2017**, *13*, 2253, DOI 10.1039/C6SM02742C (cited on page 7).

[18] B. Calderón-Jiménez, M. E. Johnson, M. Bustos, A. R, K. E. Murphy, M. R. Winchester, V. Baudrit, J. R, 'Silver Nanoparticles: Technological Advances, Societal Impacts, and Metrological Challenges', *Frontiers in Chemistry* **2017**, *5*, DOI 10.3389/fchem.2017.00006 (cited on page 65).

[19] V. Castelletto, A. Kaur, I. W. Hamley, R. H. Barnes, K.-A. Karatzas, D. Hermida-Merino, S. Swioklo, C. J. Connon, J. Stasiak, M. Reza, J. Ruokolainen, 'Hybrid membrane biomaterials from self-assembly in polysaccharide and peptide amphiphile mixtures: controllable structural and mechanical properties and antimicrobial activity', **2017**, *7*, 8366, DOI 10.1039/C6RA27244D (cited on page 7).

[20] A. Concellón, E. Blasco, A. Martínez-Felipe, J. C. Martínez, I. Šics, T. A. Ezquerra, A. Nogales, M. Piñol, L. Oriol, 'Light-Responsive Self-Assembled Materials by Supramolecular Post-Functionalization via Hydrogen Bonding of Amphiphilic Block Copolymers', *Macromolecules* **2016**, *49*, 7825, DOI 10.1021/acs.macromol.6b01112 (cited on page 30).

[21] P.-J. De Temmerman, E. Verleysen, J. Lammertyn, J. Mast, 'Size measurement uncertainties of near-monodisperse, near-spherical nanoparticles using transmission electron microscopy and particle-tracking analysis', *Journal of Nanoparticle Research* **2014**, *16*, 2628, DOI 10.1007/s11051-014-2628-3 (cited on pages 14, 15).

[22] P. Debye, 'Molecular-weight Determination by Light Scattering.', *The Journal of Physical and Colloid Chemistry* **1947**, *51*, 18, DOI 10.1021/j150451a002 (cited on page 33).

[23] G. Deželić, M. Wrischer, Z. Devidé, J. P. Kratohvil, 'Electron microscopy of Ludox colloidal silica', *Kolloid-Zeitschrift* **1960**, *171*, 42, DOI 10.1007/BF01520323 (cited on page 1).

[24] Y. Dieckmann, H. Cölfen, H. Hofmann, A. Petri-Fink, 'Particle Size Distribution Measurements of Manganese-Doped ZnS Nanoparticles', *Analytical Chemistry* **2009**, *81*, 3889, DOI 10.1021/ac900043y (cited on page 1).

[25] M. Doucet, J. H. Cho, G. Alina, J. Bakker, W. Bouwman, P. Butler, K. Campbell, M. Gonzales, R. Heenan, A. Jackson, P. Juhas, S. King, P. Kienzle, J. Krzywon, A. Markvardsen, T. Nielsen, L. O'Driscoll, W. Potrzebowski, R. F. Leal, T. Richter, P. Rozycko, A. Washington, SasView for Small Angle Scattering Analysis, **2017**, http://www.sasview.org/about.html (visited on 04/04/2017) (cited on page 2).

[26] D. Duan, H. Torosyan, D. Elnatan, C. K. McLaughlin, J. Logie, M. S. Shoichet, D. A. Agard, B. K. Shoichet, 'Internal Structure and Preferential Protein Binding of Colloidal Aggregates', *ACS Chemical Biology* **2017**, *12*, 282, DOI 10.1021/acschembio.6b00791 (cited on page 29).

[27] A. Einstein, 'Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen', *Annalen der Physik* **1905**, *322*, 549, DOI 10.1002/andp.19053220806 (cited on page 49).

[28] J. Eyssautier, P. Levitz, D. Espinat, J. Jestin, J. Gummel, I. Grillo, L. Barré, 'Insight into Asphaltene Nanoaggregate Structure Inferred by Small Angle Neutron and X-ray Scattering', *The Journal of Physical Chemistry B* **2011**, *115*, 6827, DOI 10.1021/jp111468d (cited on page 7).

[29] I. S. Fedorova, P. W. Schmidt, 'A general analytical method for calculating particle-dimension distributions from scattering data', *Journal of Applied Crystallography* **1978**, *11*, 405, DOI 10.1107/S0021889878013503 (cited on page 4).

[30] R. Finsy, E. Moreels, A. Bottger, H. Lekkerkerker, 'Study of the relation between diffusion and sedimentation of charged silica sols by dynamic light scattering, ultracentrifugation, and turbidimetry', *The Journal of Chemical Physics* **1985**, *82*, 3812, DOI 10.1063/1.448869 (cited on pages 24, 27).

[31] R. Finsy, 'Particle sizing by quasi-elastic light scattering', *Advances in Colloid and Interface Science* **1994**, *52*, 79, DOI 10.1016/0001-8686(94)80041-3 (cited on pages 3, 47, 48, 52).

[32] R. Finsy, P. D. Groen, L. Deriemaeker, M. v. Laethem, 'Singular value analysis and reconstruction of photon correlation data equidistant in time', *The Journal of Chemical Physics* **1989**, *91*, 7374, DOI 10.1063/1.457260 (cited on page 4).

[33] R. Finsy, N. de Jaeger, R. Sneyers, E. Geladé, 'Particle Sizing by Photon Correlation Spectroscopy. Part III: Mono and bimodal distributions and data analysis', *Particle & Particle Systems Characterization* **1992**, *9*, 125, DOI 10.1002/ppsc.19920090117 (cited on page 65).

[34] K. Fischer, M. Schmidt, 'Pitfalls and novel applications of particle sizing by dynamic light scattering', *Biomaterials* **2016**, *98*, 79, DOI 10.1016/j.biomaterials.2016.05.003 (cited on page 3).

[35] S. Förster, L. Apostol, W. Bras, 'Scatter: software for the analysis of nano- and mesoscale small-angle scattering', *Journal of Applied Crystallography* **2010**, *43*, 639, DOI 10.1107/S0021889810008289 (cited on page 2).

[36] D. Franke, C. M. Jeffries, D. I. Svergun, 'Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra', *Nature Methods* **2015**, *12*, 419, DOI 10.1038/nmeth.3358 (cited on pages 17, 18).

[37] K. Franks, A. Braun, J. Charoud-Got, O. Couteau, V. Kestens, M. A. Lamberty, T. P. J. Linsinger, G. Roebben, Certification of the equivalent spherical diameters of silica nanoparticles in aqueous solution - Certified Reference Material ERM®-FD304, EUR - Scientific and Technical Research Reports, European Commission, Joint Research Centre, Institute for Reference Materials and Measurements (IRMM), Geel, Belgium, **2012**, http://publications.jrc.ec.europa.eu/repository/handle/JRC67374 (visited on 03/23/2017) (cited on page 2).

[38] A. Fuhrmann, R. Göstl, R. Wendt, J. Kötteritzsch, M. D. Hager, U. S. Schubert, K. Brademann-Jock, A. F. Thünemann, U. Nöchel, M. Behl, S. Hecht, 'Conditional repair by locally switching the thermal healing capability of dynamic covalent polymers with light', *Nature Communications* **2016**, *7*, 13623, DOI 10.1038/ncomms13623 (cited on page 7).

[39] U. M. Garusinghe, V. S. Raghuwanshi, C. J. Garvey, S. Varanasi, C. R. Hutchinson, W. Batchelor, G. Garnier, 'Assembly of nanoparticles-polyelectrolyte complexes in nanofiber cellulose structures', *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2017**, *513*, 373, DOI 10.1016/j.colsurfa.2016.10.068 (cited on page 7).

[40] G. Gelardi, N. Sanson, G. Nagy, R. J. Flatt, 'Characterization of Comb-Shaped Copolymers by Multidetection SEC, DLS and SANS', *Polymers* **2017**, *9*, 61, DOI 10.3390/polym9020061 (cited on page 7).

[41] R. Gilles, D. Mukherji, L. Karge, P. Strunz, P. Beran, B. Barbier, A. Kriele, M. Hofmann, H. Eckerlebe, J. Roesler, 'Stability of TaC precipitates in a Co-Re-based alloy being developed for ultra-high-temperature applications', *Journal of Applied Crystallography* **2016**, *49*, 1253, DOI 10.1107/S1600576716009006 (cited on page 7).

[42] O. Glatter, 'A new method for the evaluation of small-angle scattering data', *Journal of Applied Crystallography* **1977**, *10*, 415, DOI 10.1107/S0021889877013879 (cited on page 5).

[43] G. Gleber, L. Cibik, S. Haas, A. Hoell, P. Müller, M. Krumrey, 'Traceable size determination of PMMA nanoparticles based on Small Angle X-ray Scattering (SAXS)', *Journal of Physics: Conference Series* **2010**, *247*, 012027, DOI 10.1088/1742-6596/247/1/012027 (cited on page 7).

[44] L. M. Gugliotta, G. S. Stegmayer, L. A. Clementi, V. D. G. Gonzalez, R. J. Minari, J. R. Leiza, J. R. Vega, 'A Neural Network Model for Estimating the Particle Size Distribution of Dilute Latex from Multiangle Dynamic Light Scattering Measurements', *Particle & Particle Systems Characterization* **2009**, *26*, 41, DOI 10.1002/ppsc.200800010 (cited on page 4).

[45] V. M. Gun'ko, A. V. Klyueva, Y. N. Levchuk, R. Leboda, 'Photon correlation spectroscopy investigations of proteins', *Advances in Colloid and Interface Science* **2003**, *105*, 201, DOI 10.1016/S0001-8686(03)00091-5 (cited on page 3).

[46] W. C. Hamilton, 'Significance tests on the crystallographic R factor', *Acta Crystallographica* **1965**, *18*, 502, DOI 10.1107/S0365110X65001081 (cited on pages 16, 17).

[47] I. W. Hamley, S. Burholt, J. Hutchinson, V. Castelletto, E. R. da Silva, W. Alves, P. Gutfreund, L. Porcar, R. Dattani, D. Hermida-Merino, G. Newby, M. Reza, J. Ruokolainen, J. Stasiak, 'Shear Alignment of Bola-Amphiphilic Arginine-Coated Peptide Nanotubes', *Biomacromolecules* **2017**, *18*, 141, DOI 10.1021/acs.biomac.6b01425 (cited on page 7).

[48] B. Hammouda, 'Small-Angle Scattering From Branched Polymers', *Macromolecular Theory and Simulations* **2012**, *21*, 372, DOI 10.1002/mats.201100111 (cited on page 21).

[49] S. Hansen, J. S. Pedersen, 'A comparison of three different methods for analysing small-angle scattering data', *Journal of Applied Crystallography* **1991**, *24*, 541, DOI 10.1107/S0021889890013322 (cited on page 5).

[50] S. Hansen, 'Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering', *Journal of Applied Crystallography* **2000**, *33*, 1415, DOI 10.1107/S0021889800012930 (cited on page 5).

[51] L. H. Hanus, H. J. Ploehn, 'Conversion of Intensity-Averaged Photon Correlation Spectroscopy Measurements to Number-Averaged Particle Size Distributions. 1. Theoretical Development', *Langmuir* **1999**, *15*, 3091, DOI 10.1021/la980958w (cited on page 65).

[52] D. Harrison, M. R. Fisch, 'Experimental Study of Noise on Photon Autocorrelation Functions', *Langmuir* **1996**, *12*, 6691, DOI 10.1021/la9606310 (cited on page 77).

[53] P. A. Hassan, S. Rana, G. Verma, 'Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering', *Langmuir* **2015**, *31*, 3, DOI 10.1021/la501789z (cited on pages 3, 68).

[54] J. Henn, 'An alternative to the goodness of fit', *Acta Crystallographica Section A: Foundations and Advances* **2016**, *72*, 696, DOI 10.1107/S2053273316013206 (cited on pages 17, 38).

[55] C. Herfurth, A. Laschewsky, L. Noirez, B. von Lospichl, M. Gradzielski, 'Thermoresponsive (star) block copolymers from one-pot sequential RAFT polymerizations and their self-assembly in aqueous solution', *Polymer*, Self-Assembly **2016**, *107*, 422, DOI 10.1016/j.polymer.2016.09.089 (cited on page 7).

[56] M. J. Hollamby, K. Aratsu, B. R. Pauw, S. E. Rogers, A. J. Smith, M. Yamauchi, X. Lin, S. Yagai, 'Simultaneous SAXS and SANS Analysis for the Detection of Toroidal Supramolecular Polymers Composed of Noncovalent Supermacrocycles in Solution', *Angewandte Chemie* **2016**, *128*, 10044, DOI 10.1002/ange.201603370 (cited on page 8).

[57] C. Huck-Iriart, J. Montes-de-Oca-Ávalos, M. L. Herrera, R. J. Candal, C. L. Pinto-de-Oliveira, I. Linares-Torriani, 'New insights about flocculation process in sodium caseinate-stabilized emulsions', *Food Research International* **2016**, *89, Part 1*, 338, DOI 10.1016/j.foodres.2016.08.026 (cited on page 7).

[58] G. Hura, J. M. Sorenson, R. M. Glaeser, T. Head-Gordon, 'A high-quality x-ray scattering experiment on liquid water at ambient conditions', *The Journal of Chemical Physics* **2000**, *113*, 9140, DOI 10.1063/1.1319614 (cited on page 34).

[59] B. Iglewicz, D. C. Hoaglin, *How to detect and handle outliers*, Milwaukee, Wis. : ASQC Quality Press, **1993**, 87 pages, http://trove.nla.gov.au/version/12350384 (visited on 04/28/2017) (cited on page 64).

[60] J. Ilavsky, P. R. Jemian, 'Irena: tool suite for modeling and analysis of small-angle scattering', *Journal of Applied Crystallography* **2009**, *42*, 347, DOI 10.1107/S0021889809002222 (cited on page 2).

[61] S. Isabettini, M. Liebi, J. Kohlbrecher, T. Ishikawa, E. J. Windhab, P. Fischer, P. Walde, S. Kuster, 'Tailoring Bicelle Morphology and Thermal Stability with Lanthanide-Chelating Cholesterol Conjugates', *Langmuir* **2016**, *32*, 9005, DOI 10.1021/acs.langmuir.6b01968 (cited on page 7).

[62] ISO, ISO 13321:1996(en), Particle size analysis — Photon correlation spectroscopy, International Organization for Standardization, **1996**, https://www.iso.org/obp/ui/?_escaped_fragment_=iso:std:21707:en#!iso:std:21707:en (visited on 10/24/2016) (cited on pages 2, 3, 51, 59, 65).

[63] ISO, ISO/FDIS 17867:2014(E), Particle size analysis — Small-angle X-ray scattering [Draft], International Organization for Standardization, **2014** (cited on page 2).

[64] A. A. Istratov, O. F. Vyvenko, 'Exponential analysis in physical phenomena', *Review of Scientific Instruments* **1999**, *70*, 1233, DOI 10.1063/1.1149581 (cited on pages 3, 47, 48).

[65] D. A. Jacques, J. M. Guss, D. I. Svergun, J. Trewhella, 'Publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution', *Acta Crystallographica. Section D Biological Crystallography* **2012**, *68*, 620, DOI 10.1107/S0907444912012073 (cited on page 3).

[66] JCGM, Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement, **2008**, http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (visited on 03/22/2017) (cited on page 18).

[67] L. Jin, V. M. Garamus, F. Liu, J. Xiao, H. Eckerlebe, R. Willumeit-Roemer, B. Mu, A. Zou, 'Interaction of a biosurfactant, Surfactin with a cationic Gemini surfactant in aqueous solution', *Journal of Colloid and Interface Science* **2016**, *481*, 201, DOI 10.1016/j.jcis.2016.07.044 (cited on page 7).

[68] C. Kästner, A. F. Thünemann, 'Catalytic Reduction of 4-Nitrophenol Using Silver Nanoparticles with Adjustable Activity', *Langmuir* **2016**, *32*, 7383, DOI 10.1021/acs.langmuir.6b01477 (cited on pages 7, 29, 65).

[69] V. Kestens, G. Roebben, The certification of the equivalent spherical diameters of silica nanoparticles in aqueous solution, ERM®-FD102, European Commission, Joint Research Centre, Institute for Reference Materials and Measurements (IRMM), Geel, Belgium, **2014**, http://publications.jrc.ec.europa.eu/repository/handle/JRC90438 (visited on 03/23/2017) (cited on page 25).

[70] V. Kestens, G. Roebben, J. Herrmann, Å. Jämting, V. Coleman, C. Minelli, C. Clifford, P.-J. D. Temmerman, J. Mast, L. Junjie, F. Babick, H. Cölfen, H. Emons, 'Challenges in the size analysis of a silica nanoparticle mixture as candidate certified reference material', *Journal of Nanoparticle Research* **2016**, *18*, 171, DOI 10.1007/s11051-016-3474-2 (cited on pages 23, 25, 27, 42, 68, 71).

[71] M. F. Khan, M. K. Singh, S. Sen, 'Measuring Size, Size Distribution, and Polydispersity of Water-in-Oil Microemulsion Droplets using Fluorescence Correlation Spectroscopy: Comparison to Dynamic Light Scattering', *The Journal of Physical Chemistry B* **2016**, *120*, 1008, DOI 10.1021/acs.jpcb.5b09920 (cited on page 3).

[72] A. L. Kholodenko, 'Analytical calculation of the scattering function for polymers of arbitrary flexibility using the Dirac propagator', *Macromolecules* **1993**, *26*, 4179, DOI 10.1021/ma00068a017 (cited on page 33).

[73] J. Kieffer, D. Karkoulis, 'PyFAI, a versatile library for azimuthal regrouping', *Journal of Physics: Conference Series* **2013**, *425*, 202012, DOI 10.1088/1742-6596/425/20/202012 (cited on page 3).

[74] D. J. Kinning, E. L. Thomas, 'Hard-sphere interactions between spherical domains in diblock copolymers', *Macromolecules* **1984**, *17*, 1712, DOI 10.1021/ma00139a013 (cited on page 33).

[75] S. Kirkham, V. Castelletto, I. W. Hamley, K. Inoue, R. Rambo, M. Reza, J. Ruokolainen, 'Self-Assembly of the Cyclic Lipopeptide Daptomycin: Spherical Micelle Formation Does Not Depend on the Presence of Calcium Chloride', *CHEMPHYSCHEM* **2016**, *17*, 2118, DOI 10.1002/cphc.201600308 (cited on page 7).

[76] T. Klein, E. Buhr, K.-P. Johnsen, C. G. Frase, 'Traceable measurement of nanoparticle size using a scanning electron microscope in transmission mode (TSEM)', *Measurement Science and Technology* **2011**, *22*, 094002, DOI 10.1088/0957-0233/22/9/094002 (cited on page 2).

[77] S. Kogikoski, M. S. Liberato, I. M. Factori, E. R. da Silva, C. L. P. Oliveira, R. A. Ando, W. A. Alves, 'Polycaprolactone–Polyaniline Blend: Effects of the Addition of Cysteine on the Structural and Molecular Properties', *The Journal of Physical Chemistry C* **2017**, *121*, 863, DOI 10.1021/acs.jpcc.6b10011 (cited on page 7).

[78] J. Kohlbrecher, I. Breßler, SASfit Manual, **2016**, https://github.com/SASfit/SASfit/blob/master/doc/manual/sasfit.pdf (visited on 03/23/2017) (cited on page 12).

[79] J. Kohlbrecher, J. Buitenhuis, G. Meier, M. P. Lettinga, 'Colloidal dispersions of octadecyl grafted silica spheres in toluene: A global analysis of small angle neutron scattering contrast variation and concentration dependence measurements', *The Journal of Chemical Physics* **2006**, *125*, 044715, DOI 10.1063/1.2220564 (cited on page 8).

[80] D. E. Koppel, 'Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants', *The Journal of Chemical Physics* **1972**, *57*, 4814, DOI 10.1063/1.1678153 (cited on pages 3, 65, 66).

[81] O. Koshkina, T. Lang, R. Thiermann, D. Docter, R. H. Stauber, C. Secker, H. Schlaad, S. Weidner, B. Mohr, M. Maskos, A. Bertin, 'Temperature-Triggered Protein Adsorption on Polymer-Coated Nanoparticles in Serum', *Langmuir* **2015**, *31*, 8873, DOI 10.1021/acs.langmuir.5b00537 (cited on page 66).

[82] A. Lak, A. F. Thünemann, M. Schilling, F. Ludwig, 'Resolving particle size modality in bi-modal iron oxide nanoparticle suspensions', *Journal of Magnetism and Magnetic Materials*, 10th International Conference on the Scientific and Clinical Applications of Magnetic Carriers 10-14 June, 2014, Dresden, Germany **2015**, *380*, 140, DOI 10.1016/j.jmmm.2014.08.050 (cited on page 5).

[83] A. Lamberty, K. Franks, A. Braun, V. Kestens, G. Roebben, T. P. J. Linsinger, 'Interlaboratory comparison for the measurement of particle size and zeta potential of silica nanoparticles in an aqueous suspension', *Journal of Nanoparticle Research* **2011**, *13*, 7317, DOI 10.1007/s11051-011-0624-4 (cited on page 64).

[84] J. Langowski, R. Bryan, 'Maximum entropy analysis of photon correlation spectroscopy data using a Bayesian estimate for the regularization parameter', *Macromolecules* **1991**, *24*, 6346, DOI 10.1021/ma00023a045 (cited on page 4).

[85] A. Letzel, B. Gökce, P. Wagener, S. Ibrahimkutty, A. Menzel, A. Plech, S. Barcikowski, 'Size Quenching during Laser Synthesis of Colloids Happens Already in the Vapor Phase of the Cavitation Bubble', *The Journal of Physical Chemistry C* **2017**, *121*, 5356, DOI 10.1021/acs.jpcc.6b12554 (cited on page 29).

[86] L. Leu, A. Georgiadis, M. J. Blunt, A. Busch, P. Bertier, K. Schweinar, M. Liebi, A. Menzel, H. Ott, 'Multiscale Description of Shale Pore Systems by Scanning SAXS and WAXS Microscopy', *Energy & Fuels* **2016**, *30*, 10282, DOI 10.1021/acs.energyfuels.6b02256 (cited on page 7).

[87] K. Levenberg, 'A method for the solution of certain non-linear problems in least squares', *Quarterly of Applied Mathematics* **1944**, *2*, 164, DOI 10.1090/qam/10666 (cited on pages 18, 83).

[88] D. Lichtenstein, J. Ebmeyer, P. Knappe, S. Juling, L. Böhmert, S. Selve, B. Niemann, A. Braeuning, A. F. Thünemann, A. Lampen, 'Impact of food components during in vitro digestion of silver nanoparticles on cellular uptake and cytotoxicity in intestinal cells', *Biological Chemistry* **2015**, *396*, 1255, DOI 10.1515/hsz-2015-0145 (cited on page 30).

[89] D. Lichtenstein, J. Ebmeyer, T. Meyer, A.-C. Behr, C. Kästner, L. Böhmert, S. Juling, B. Niemann, C. Fahrenson, S. Selve, A. F. Thünemann, J. Meijer, I. Estrela-Lopis, A. Braeuning, A. Lampen, 'It takes more than a coating to get nanoparticles through the intestinal barrier in vitro', *European Journal of Pharmaceutics and Biopharmaceutics* **2016**, DOI 10.1016/j.ejpb.2016.12.004 (cited on page 30).

[90] F. Liebig, A. F. Thünemann, J. Koetz, 'Ostwald Ripening Growth Mechanism of Gold Nanotriangles in Vesicular Template Phases', **2016**, http://pubs.acs.org/doi/full/10.1021/acs.langmuir.6b02662 (visited on 11/29/2016) (cited on page 7).

[91] D. Mahl, J. Diendorf, W. Meyer-Zaika, M. Epple, 'Possibilities and limitations of different analytical methods for the size determination of a bimodal dispersion of metallic nanoparticles', *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2011**, *377*, 386, DOI 10.1016/j.colsurfa.2011.01.031 (cited on page 1).

[92] A. G. Mailer, P. S. Clegg, P. N. Pusey, 'Particle sizing by dynamic light scattering: non-linear cumulant analysis', *Journal of Physics: Condensed Matter* **2015**, *27*, 145102, DOI 10.1088/0953-8984/27/14/145102 (cited on page 77).

[93] H. Mamiya, J. Rabajczyk, N. Watanabe, A. Kowalska, H. Kitazawa, 'Aging-treatment-induced soft magnetism in nickel–chromium-based superalloy X-750', *Journal of Alloys and Compounds* **2016**, *681*, 367, DOI http://dx.doi.org/10.1016/j.jallcom.2016.04.237 (cited on page 29).

[94] Y. Meir, E. Jerby, Z. Barkay, D. Ashkenazi, J. B. Mitchell, T. Narayanan, N. Eliaz, J.-L. LeGarrec, M. Sztucki, O. Meshcheryakov, 'Observations of Ball-Lightning-Like Plasmoids Ejected from Silicon by Localized Microwaves', *Materials* **2013**, *6*, 4011, DOI 10.3390/ma6094011 (cited on page 5).

[95] F. Meli, T. Klein, E. Buhr, C. G. Frase, G. Gleber, M. Krumrey, A. Duta, S. Duta, V. Korpelainen, R. Bellotti, G. B. Picotto, R. D. Boyd, A. Cuenat, 'Traceable size determination of nanoparticles, a comparison among European metrology institutes', *Measurement Science and Technology* **2012**, *23*, 125005, DOI 10.1088/0957-0233/23/12/125005 (cited on pages 2, 7, 14).

[96] I. D. Morrison, E. F. Grabowski, C. A. Herb, 'Improved techniques for particle size determination by quasi-elastic light scattering', *Langmuir* **1985**, *1*, 496, DOI 10.1021/la00064a016 (cited on page 4).

[97] E. Mylonas, D. I. Svergun, 'Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering', *Journal of Applied Crystallography* **2007**, *40*, s245, DOI 10.1107/S002188980700252X (cited on page 23).

[98] M. Naiim, A. Boualem, C. Ferre, M. Jabloun, A. Jalocha, P. Ravier, 'Multiangle dynamic light scattering for the improvement of multimodal particle size distribution measurements', *Soft Matter* **2014**, *11*, 28, DOI 10.1039/C4SM01995D (cited on page 51).

[99] V. Nguyen-Kim, S. Prévost, K. Seidel, W. Maier, A.-K. Marguerre, G. Oetter, T. Tadros, M. Gradzielski, 'Solubilization of active ingredients of different polarity in Pluronic® micellar solutions – Correlations between solubilizate polarity and solubilization site', *Journal of Colloid and Interface Science* **2016**, *477*, 94, DOI http://dx.doi.org/10.1016/j.jcis.2016.05.017 (cited on page 7).

[100] J. E. Noble, E. De Santis, J. Ravi, B. Lamarre, V. Castelletto, J. Mantell, S. Ray, M. G. Ryadnov, 'A De Novo Virus-Like Topology for Synthetic Virions', *Journal of the American Chemical Society* **2016**, *138*, 12202, DOI 10.1021/jacs.6b05751 (cited on page 7).

[101] S.-L. Nyeo, R. R. Ansari, 'Data inversion for dynamic light scattering using Fisher information', *Laser Physics* **2015**, *25*, 075703, DOI 10.1088/1054-660X/25/7/075703 (cited on page 3).

[102] Y. Oba, S. Morooka, H. Sato, N. Sato, K. Ohishi, J.-i. Suzuki, M. Sugiyama, 'Simultaneous Characterisation of Precipitates and Matrix in a Steel Using Small-Angle Neutron Scattering and Bragg-Edge Transmission Analysis', *ISIJ International* **2015**, *55*, 2618, DOI 10.2355/isijinternational.ISIJINT-2015-315 (cited on page 5).

[103] D. Orthaber, A. Bergmann, O. Glatter, 'SAXS experiments on absolute scale with Kratky systems using water as a secondary standard', *Journal of Applied Crystallography* **2000**, *33*, 218, DOI 10.1107/S0021889899015216 (cited on pages 14, 16, 23, 42).

[104] B. R. Pauw, C. Kästner, A. F. Thünemann, 'Nanoparticle Size Distribution Quantification: Results of a SAXS Inter-Laboratory Comparison', *arXiv:1702.03902 [cond-mat physics:physics]* **2017**, http://arxiv.org/abs/1702.03902 (visited on 03/22/2017) (cited on page 7).

[105] B. R. Pauw, M. Ohnuma, K. Sakurai, E. A. Klop, '2D anisotropic scattering pattern fitting using a novel Monte Carlo method: Initial results', *arXiv:1303.2903 [physics]* **2013**, http://arxiv.org/abs/1303.2903 (visited on 05/06/2014) (cited on page 5).

[106] B. R. Pauw, J. S. Pedersen, S. Tardif, M. Takata, B. B. Iversen, 'Improvements and considerations for size distribution retrieval from small-angle scattering data by Monte Carlo methods', *Journal of Applied Crystallography* **2013**, *46*, 365, DOI 10.1107/S0021889813001295 (cited on pages 2, 3, 5, 26, 27, 30, 32, 34, 37, 52).

[107] J. S. Pedersen, 'Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting', *Advances in Colloid and Interface Science* **1997**, *70*, 171, DOI 10.1016/S0001-8686(97)00312-6 (cited on pages 4, 31).

[108] M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev, D. I. Svergun, 'New developments in the ATSAS program package for small-angle scattering data analysis', *Journal of Applied Crystallography* **2012**, *45*, 342, DOI 10.1107/S0021889812007662 (cited on page 3).

[109] G. Porod, 'Die Röntgenkleinwinkelstreuung von dichtgepackten Kolloiden Systemen 2. Teil Fortsetzung', *Kolloid-Zeitschrift und Zeitschrift Für Polymere* **1952**, *125*, 108, DOI 10.1007/BF01526289 (cited on pages 31, 35).

[110] J. Potočnik, Commission Recommendation of 18 October 2011 on the Definition of Nanomaterial, OJ L 275, European Union, Brussels, **2011**, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011H0696 (visited on 12/09/2014) (cited on pages 1, 26).

[111] S. W. Provencher, 'A Constrained Regularization Method For Inverting Data Represented By Linear Algebraic Or Integral-Equations', *Computer Physics Communications* **1982**, *27*, 213, DOI 10.1016/0010-4655(82)90173-4 (cited on page 4).

[112] S. W. Provencher, 'Contin - A General-Purpose Constrained Regularization Program For Inverting Noisy Linear Algebraic And Integral-Equations', *Computer Physics Communications* **1982**, *27*, 229, DOI 10.1016/0010-4655(82)90174-6 (cited on page 3).

[113] R. P. Rambo, *ScÅtter - software for the analysis of biological SAXS datasets*, **2017**, http://www.bioisis.net/tutorial/9 (cited on page 3).

[114] V. Roger, H. Cottet, L. Cipelletti, 'A New Robust Estimator of Polydispersity from Dynamic Light Scattering Data', *Analytical Chemistry* **2016**, *88*, 2630, DOI 10.1021/acs.analchem.5b03584 (cited on page 77).

[115] J. M. Rosalie, B. R. Pauw, 'Form-free size distributions from complementary stereological TEM/SAXS on precipitates in a Mg-Zn alloy', *Acta Materialia* **2014**, *66*, 150, DOI 10.1016/j.actamat.2013.11.029 (cited on pages 4, 5, 32, 34).

[116] W. Schärtl, *Light Scattering from Polymer Solutions and Nanoparticle Dispersions*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2007**, 191 pages, http://link.springer.com/10.1007/978-3-540-71951-9 (visited on 10/02/2015) (cited on page 3).

[117] K. Schätzel, 'Noise on photon correlation data. I. Autocorrelation functions', *Quantum Optics: Journal of the European Optical Society Part B* **1990**, *2*, 287, DOI 10.1088/0954-8998/2/4/002 (cited on page 77).

[118] J. Scheck, B. Wu, M. Drechsler, R. Rosenberg, A. E. S. Van Driessche, T. M. Stawski, D. Gebauer, 'The Molecular Mechanism of Iron(III) Oxide Nucleation', *The Journal of Physical Chemistry Letters* **2016**, *7*, 3123, DOI 10.1021/acs.jpclett.6b01237 (cited on page 29).

[119] T. Schindler, J. Walter, W. Peukert, D. Segets, T. Unruh, 'In Situ Study on the Evolution of Multimodal Particle Size Distributions of ZnO Quantum Dots: Some General Rules for the Occurrence of Multimodalities', *The Journal of Physical Chemistry B* **2015**, *119*, 15370, DOI 10.1021/acs.jpcb.5b08005 (cited on page 5).

[120] Z. Schnepp, Y. Zhang, M. J. Hollamby, B. R. Pauw, M. Tanaka, Y. Matsushita, Y. Sakka, 'Doped-carbon electrocatalysts with trimodal porosity from a homogeneous polypeptide gel', *J. Mater. Chem. A* **2013**, *1*, 13576, DOI 10.1039/C3TA12996A (cited on pages 5, 34).

[121] M. Shibayama, T. Karino, S. Okabe, 'Distribution analyses of multi-modal dynamic light scattering data', *Polymer* **2006**, *47*, 6446, DOI 10.1016/j.polymer.2006.06.060 (cited on pages 50, 59).

[122] R. E. Shiffler, 'Maximum Z Scores and Outliers', *The American Statistician* **1988**, *42*, 79, DOI 10.2307/2685269 (cited on page 64).

[123] J. A. Small, R. L. Watters, Reference Material 8011, National Institute of Standards & Technology, Gaithersburg, MD 20899, USA, **2015**, https://www-s.nist.gov/srmors/reports/8011.pdf (visited on 03/23/2017) (cited on pages 14, 15).

[124] B. M. Soares, A. M. Aguilar, E. R. Silva, M. D. Coutinho-Neto, I. W. Hamley, M. Reza, J. Ruokolainen, W. A. Alves, 'Chiral organocatalysts based on lipopeptide micelles for aldol reactions in water', **2017**, *19*, 1181, DOI 10.1039/C6CP08135E (cited on page 7).

[125] C. Song, P. Wang, H. A. Makse, 'A phase diagram for jammed matter', *Nature* **2008**, *453*, 629, DOI 10.1038/nature06981 (cited on page 43).

[126] A. B. Stefaniak, V. A. Hackley, G. Roebben, K. Ehara, S. Hankin, M. T. Postek, I. Lynch, W.-E. Fu, T. P. J. Linsinger, A. F. Thünemann, 'Nanoscale reference materials for environmental, health and safety measurements: needs, gaps and opportunities', *Nanotoxicology* **2013**, *7*, 1325, DOI 10.3109/17435390.2012.739664 (cited on page 65).

[127] D. I. Svergun, 'Mathematical methods in small-angle scattering data analysis', *Journal of Applied Crystallography* **1991**, *24*, 485, DOI 10.1107/S0021889891001280 (cited on page 5).

[128] D. I. Svergun, J. S. Pedersen, 'Propagating errors in small-angle scattering data treatment', *Journal of Applied Crystallography* **1994**, *27*, 241, DOI 10.1107/S0021889893008337 (cited on page 41).

[129] D. I. Svergun, M. H. J. Koch, P. A. Timmins, R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*, Oxford University Press, Oxford, **2013**, 368 pages (cited on page 3).

[130] W. Szczerba, R. Costo, S. Veintemillas-Verdaguer, M. d. P. Morales, A. F. Thünemann, 'SAXS analysis of single- and multi-core iron oxide magnetic nanoparticles', *Journal of Applied Crystallography* **2017**, *50*, DOI 10.1107/S1600576717002370 (cited on page 7).

[131] M. Sztucki, T. Narayanan, G. Beaucage, 'In situ study of aggregation of soot particles in an acetylene flame by small-angle x-ray scattering', *Journal of Applied Physics* **2007**, *101*, 114304, DOI 10.1063/1.2740341 (cited on page 10).

[132] K. Tiede, A. B. A. Boxall, S. P. Tear, J. Lewis, H. David, M. Hassellöv, 'Detection and characterization of engineered nanoparticles in food and the environment', *Food Additives & Contaminants: Part A* **2008**, *25*, 795, DOI 10.1080/02652030802007553 (cited on page 1).

[133] J. Trewhella, 'Small-angle scattering and 3D structure interpretation', *Current Opinion in Structural Biology* **2016**, *40*, 1, DOI 10.1016/j.sbi.2016.05.003 (cited on page 3).

[134] J. S. Tsuji, A. D. Maynard, P. C. Howard, J. T. James, C.-w. Lam, D. B. Warheit, A. B. Santamaria, 'Research Strategies for Safety Evaluation of Nanomaterials, Part IV: Risk Assessment of Nanoparticles', *Toxicological Sciences* **2006**, *89*, 42, DOI 10.1093/toxsci/kfi339 (cited on page 1).

[135] J. Vanhoudt, J. Clauwaert, 'Experimental Comparison of Fiber Receivers and a Pinhole Receiver for Dynamic and Static Light Scattering', *Langmuir* **1999**, *15*, 44, DOI 10.1021/la980747r (cited on page 51).

[136] Z. Varga, Y. Yuana, A. E. Grootemaat, E. van der Pol, C. Gollwitzer, M. Krumrey, R. Nieuwland, 'Towards traceable size determination of extracellular vesicles', *Journal of Extracellular Vesicles* **2014**, *3*, 23298, DOI 10.3402/jev.v3.23298 (cited on page 7).

[137] R. Vavrin, J. Kohlbrecher, A. Wilk, M. Ratajczyk, M. P. Lettinga, J. Buitenhuis, G. Meier, 'Structure and phase diagram of an adhesive colloidal dispersion under high pressure: A small angle neutron scattering, diffusing wave spectroscopy, and light scattering study', *The Journal of Chemical Physics* **2009**, *130*, 154903, DOI 10.1063/1.3103245 (cited on page 8).

[138] A. Viani, K. Sotiriadis, I. Kumpová, L. Mancini, M.-S. Appavou, 'Microstructural characterization of dental zinc phosphate cements using combined small angle neutron scattering and microfocus X-ray computed tomography', *Dental Materials* **2017**, *33*, 402, DOI 10.1016/j.dental.2017.01.008 (cited on page 29).

[139] A. Viani, K. Sotiriadis, P. Sasek, M.-S. Appavou, 'Evolution of microstructure and performance in magnesium potassium phosphate ceramics: Role of sintering temperature of MgO powder', *Ceramics International* **2016**, *42*, 16310, DOI 10.1016/j.ceramint.2016.07.182 (cited on page 29).

[140] A. P. Walczak, R. Fokkink, R. Peters, P. Tromp, Z. E. H. Rivera, I. M. C. M. Rietjens, P. J. M. Hendriksen, H. Bouwmeester, 'Behaviour of silver nanoparticles and silver ions in an in vitro human gastrointestinal digestion model', *Nanotoxicology* **2012**, *7*, 1198, DOI 10.3109/17435390.2012.726382 (cited on page 65).

[141] A. Wiedenmann, R. Gaehler, C. D. Dewhurst, U. Keiderling, S. Prevost, J. Kohlbrecher, 'Relaxation mechanisms in magnetic colloids studied by stroboscopic spin-polarized small-angle neutron scattering', *Physical Review B* **2011**, *84*, DOI 10.1103/PhysRevB.84.214303 (cited on page 8).

[142] E. Woźniak, M. Špírková, M. Šlouf, V. M. Garamus, M. Šafaříková, I. Šafařík, M. Štěpánek, 'Stabilization of aqueous dispersions of poly(methacrylic acid)-coated iron oxide nanoparticles by double hydrophilic block polyelectrolyte poly(ethylene oxide)-block-poly(N-methyl-2-vinylpyridinium iodide)', *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2017**, *514*, 32, DOI 10.1016/j.colsurfa.2016.11.044 (cited on page 7).

[143] R. Xu, *Particle Characterization: Light Scattering Methods*, (Ed.: B. Scarlett), Kluwer Academic Publishers, Dordrecht, **2002**, http://link.springer.com/10.1007/0-306-47124-8 (visited on 10/02/2015) (cited on pages 65, 66).

[144] X. Zhao, R. Liu, H. Zhang, Y. Shang, Y. Song, C. Liu, T. Wang, Y. Gong, Z. Li, 'Structure evolution of aluminosilicate sol and its structure-directing effect on the synthesis of NaY zeolite', *Journal of Applied Crystallography* **2017**, *50*, 231, DOI 10.1107/S1600576716020409 (cited on page 29).

[145] X. Zhu, J. Shen, Y. Wang, J. Guan, X. Sun, X. Wang, 'The reconstruction of particle size distributions from dynamic light scattering data using particle swarm optimization techniques with different objective functions', *Optics & Laser Technology* **2011**, *43*, 1128, DOI 10.1016/j.optlastec.2011.02.010 (cited on pages 4, 48).

8

Acknowledgements

First, I thank Prof. Dr. Andreas F. Thünemann for the opportunity to work and research in a fascinating and emerging field, for all guidance, many helpful discussions and continuous encouragement along the way. For supervising my work as a doctoral supervisor and for always standing by with advice, I would like to thank Prof. Dr. Michael Gradzielski of the TU Berlin. I thank Prof. Dr. Thorsten Ressler for chairing the doctoral committee.

I am very grateful for a productive and enjoyable long-term collaboration with Dr. Brian R. Pauw and especially for proofreading the manuscript and many invaluable tips for scientific writing. Moreover, I thank Dr. Joachim Kohlbrecher for introducing me to the programming of SASfit and the field of small-angle scattering many years ago, as well as the fruitful collaboration since then. Similarly, I thank Dr. Sylvain Prévost for many enlightening discussions, his continuous support and valuable feedback. For the last years of my work at the BAM, I thank Dr. Wojciech Szczerba for many insightful scientific discussions alike, and for sharing his virtually inexhaustible general knowledge with me.

Moreover, I thank Petra Fengler, Kerstin Brademann-Jock, Maximilian Ebisch and Tina Garchow for measuring most of the DLS and SAXS data and the laboratory work. Additionally, I would like to thank my fellow PhD students Claudia Kästner and Patrick Saloga for their valuable feedback on the software and their support. For the organizational work and being the heart of the working group I would like to thank Dana Ziegel and Marianne Haske.

I am grateful for the financial support by the MIS program of BAM which funded my work and this thesis. Furthermore, I acknowledge Peter Høghøj and Xenocs for their provision of the dense spheres data set in section 3.5.2.

Finally, I want give the most special thanks to Mirjam for her love, support and patience through all times.