Soft Matter

PAPER

View Article Online
View Journal | View Issue

Cite this: *Soft Matter*, 2025, **21**, 8060

Accelerated small angle neutron scattering algorithms for polymeric materials

Kexin Dai and Bradley D. Olsen *

Small-angle neutron scattering (SANS) is an extremely powerful technique for characterizing a wide variety of soft, biological, magnetic, and quantum materials, but it is often throughput-limited. This work proposes an algorithm to accelerate small angle neutron scattering (SANS) experiments by estimating the minimum number of counts to perform parameter estimation and model differentiation tasks to a specified level of certainty. Three classes of model polymer materials were examined and analyzed, and time slices of SANS data were used to model a reduced number of counts. The scattering data with reduced numbers of counts were fitted to SANS model functions to perform parameter estimation and model differentiation tasks. For parameter estimation, estimators accurate to within 5–10% of the full count estimator can be produced with only 1–50% of the full counts depending upon the sample and parameter of interest. In order to project parameter uncertainties at lower number of counts prior to the completion of experiments, it is crucial to have a robust error quantification method that reflects the true uncertainty associated with each parameter. Uncertainties from Monte Carlo (MC) bootstrapping are shown to in general overestimate the error from fitting many experimental replicates. For most parameter estimation techniques, the weighted least squares estimator is unbiased; however, certain models yield biased estimators. To differentiate between models, both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) can be used, and with either criterion, reduced numbers of counts can still identify the best model for our samples from a group of related candidate models for each material. The proposed algorithm can help SANS users optimize valuable beamtime and accelerate the use of SANS for structural characterization of libraries of materials while obtaining reasonable parameter estimation and model differentiation when scattering models are available.

Received 14th November 2024,
Accepted 12th September 2025

DOI: 10.1039/d4sm01350f

rsc.li/soft-matter-journal

Introduction

Neutron scattering is a useful technique to study the static and dynamic structures in a variety of materials such as polymers,[1] colloids,[2] biomacromolecules,[3] metals,[4] glasses,[5] and ceramics.[6] Among many neutron scattering techniques, small angle neutron scattering (SANS) is one of the most widely used in probing the microstructure of polymer and soft materials.[7] SANS has been successfully used to characterize the structural formation in polymer solutions, melts, and networks and has advanced the understanding of fundamental polymer physics.[8,9] The use of neutrons provide several advantages over other radiation sources, such as X-ray or light, because neutrons are nondestructive to soft matter[10] and allow contrast variation using isotopic labeling.[11] However, neutron count rates are inherently lower because neutrons do not interact very strongly with many materials, and neutron sources generally have lower flux than other radiation sources. These effects make SANS experiments slower. Typically, SANS data is acquired for a pre-determined number of counts above the background. After data acquisition and reduction, models are fit to experimental data to extract relevant structural parameters or identification of structures based on the goodness of fit. The measurement time to obtain SANS data for each sample can range from ten minutes to tens of hours depending on the facility, the scattering power of the sample, and instrument configuration.[12] Because neutron sources are expensive to construct and maintain, neutron scattering is only available at a handful of facilities worldwide, limiting access to this characterization technique.

With the surge of machine learning[13,14] and data driven research in material discovery,[15] the need to generate databases of materials with experimental data has led to many automated high throughput material syntheses in daily experimental workflow.[16–18] With the importance of neutron scattering in studying the microstructure and dynamics of many materials, many large facilities have been improving the hardware of the instruments, upgrading the software at the beamline,[19] making

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. E-mail: bdolsen@mit.edu; Tel: +1 617 715 4548

the data reduction process automated, and building higher flux neutron facilities[20] to increase the accessibility of neutrons to users in the world.

There have been increasing efforts in literature to optimize the use of beam time through experimental design, statistical analysis, and machine learning approaches. Steinhart and Pleštil proposed an algorithm to balance the tradeoff between measurement time and resolution by using the scattering intensity data from background measurement to optimize the measurement time of the sample at each angle for small angle X-ray experiments.[21] Saito and colleagues applied kernel density estimation in reducing measurement time for anisotropic samples, achieving a significant reduction without fully addressing the smearing effect.[22] Kanazawa *et al.* proposed accelerated small-angle scattering experiments with simulation-based machine learning.[23] They created a database of virtual experiments by simulation and use those data adaptively during real experiments to aid the selection of optimal wavevector (q) ranges sequentially to achieve the best sampling at each measurement. Another attempt to accelerate the data acquisition by Chang and coworkers involved increasing the size of binning of the detector pixels at the sacrifice of resolution and reconstructing the high-resolution image using a deep-learning based super resolution technique trained by data on EQ-SANS at Oak Ridge National Lab.[12] Their work shows that the reconstructed 2D scattering image is comparable in resolution to the true experimental data and effective at saving beamtime if implemented online during SANS acquisition. In addition, Chen *et al.* introduced a model-free statistical inference framework for SANS that uses Gaussian process regression (GPR) to reconstruct smooth, noise-reduced scattering intensity profiles directly from sparse, low-count measurements. By leveraging the inherent smoothness and continuity of scattering functions, their Bayesian approach infers missing information without requiring an analytical model or prior structural assumptions. This method can significantly reduce the number of counts required to obtain a smooth scattering function.[24] Recent work by Do *et al.* applies this GPR method to propose a model-free convergence metric for measurement sufficiency in time-resolved SANS. This convergence metric guides the user when additional counting yields negligible information gain.[25] However, there remains an opportunity to critically analyze the question of how many counts are required for data acquisition based on the type of structural information being obtained from the scattering experiment.

Herein, a novel small angle neutron scattering workflow is proposed that focuses on experimental information content rather than total counts, allowing counts to be reduced with minimal loss of knowledge gained by an experiment. To do this, two common tasks are considered: parameter estimation and structural model differentiation. The algorithm for counts optimization uses parameter uncertainty estimated in real time for both tasks, and decisions can be made regarding the optimal number of counts to achieve a targeted uncertainty level. Three different model polymer samples for SANS experiments examined in this paper include a polymer in solution, associative protein hydrogel, and micellar solution; however, the proposed algorithm can be generally applied to many other systems.

Methods

Sample preparation

All deuterated solvents at 99.9% purity were purchased from Cambridge Isotopic Laboratories, Inc. End-functional poly(ethylene glycol) (PEG) polymer was synthesized following procedures in literature[26,27] and dissolved in deuterated DMF at a concentration of 16 mM. The associative protein hydrogel is formed by an artificial protein denoted as P4, which contains four associative coiled-coil domains (P) connected by flexible polyelectrolyte linkers (C_{10}) as illustrated in Fig. 1(a). The gene sequence has been reported in literature,[28] and the protein expression protocols have been developed previously.[29] The P4 hydrogel was prepared at 6.5% (w/v) by dissolving P4 in 100 mM deuterated phosphate buffer at pD = 7.6. The blends were equilibrated overnight at 5 °C before stirring and centrifugation at 13 100 × g at 10 °C to ensure homogeneous gel formation. Pluronic F-127 was purchased from Millipore Sigma. The polymer was dissolved in D_2O at 1.0 wt%.

Small angle neutron scattering (SANS)

The P4 gel was loaded into a 1 mm demountable copper cell by spreading the gel on one side of the quartz window and compressing the second quartz window onto the sample carefully to prevent bubble formation. The Pluronic F-127 and the 16 mM PEG solutions were loaded into 1 mm quartz banjo cells. The small-angle neutron scattering experiments were performed on GP-SANS at Oak Ridge National Lab (ORNL).[30] The lower q configuration has a sample-to-detector distance of 15 m, which corresponds to q range of 0.003698 to 0.04988 Å. The higher q configuration has a sample to detector distance of 2 m, which corresponds to q range of 0.03 to 0.43 Å. The wavelength for both configurations is 4.75 Å. The measurements of P4 gel and 16 mM PEG solutions were performed at room temperature. The measurement of Pluronic F-127 was performed at 60 °C. P4 gel and 16 mM PEG solutions had approximately 500 000 counts above the background, and Pluronic F-127 had approximately 1 000 000 counts above the background.

SANS data analysis

Data reduction was performed on jupyter.sns.gov using the data reduction script associated with the software developed at ORNL.[19] The SANS data was further reduced by subtracting the solvent for the two configurations and merging the two configurations together by scaling the higher q data to the lower q data by minimizing the weighted least square residual in the overlap region. The overlap region is chosen to be 0.033 to 0.049 Å for P4 gel and 16 mM PEG solution and 0.033 to 0.0613 Å for Pluronic F-127. Time slicing was performed by calculating the count rates and estimating the time step to slice

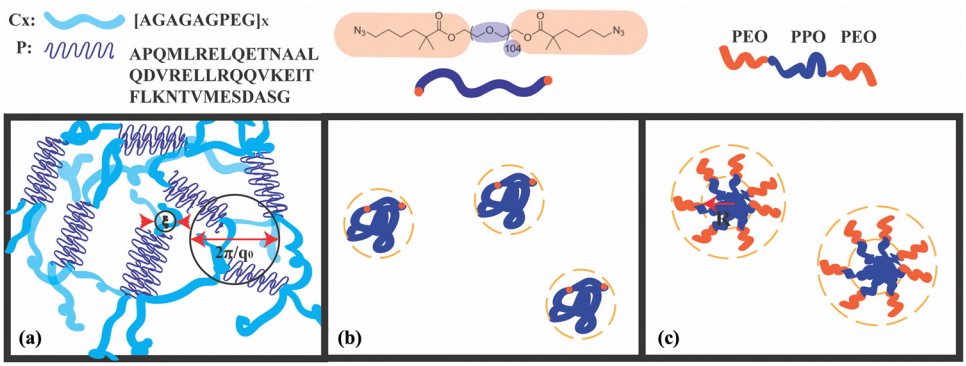

Fig. 1 (a) Illustrative figure of the P4 network formed by coiled-coil association. The coiled-coil domains (P4) are in dark blue and the polyelectrolyte linkers are in light blue. (b) Structure of polymer solution forming a Gaussian coil in dilute theta solution. (c) Pluronic F-127 self assembles into spherical micelle with PPO in the core and PEO in the corona above the critical micelle concentration.

such that the counts are 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1 of the total counts with respect to lower q and higher q. Count fractions lower than 0.01 required time resolution that was beyond the capacity of the data reduction software.

SANS model functions

Debye function. When uncharged polymers are dissolved in theta solvents, they form Gaussian coils in solution which can be described by the Debye function[31] as illustrated in Fig. 1(b). The Debye function is commonly used to obtain the radius of gyration of the polymer. The form factor of the Debye function to describe monodisperse polymer chains is

$$P(q) = \frac{2\left(e^{-q^2R_g^2} + q^2R_g^2 - 1\right)}{\left(q^2R_g^2\right)^2} \quad (1)$$

where q is the scattering wavevector, and R_g is the radius of gyration of the polymer. The bounds for R_g are set between 10 and 80 Å. For dilute solution, form factor fitting was used to perform parameter estimation. The scattering intensity can be expressed as

$$I(q) = A \times P(q) + b \quad (2)$$

where A is the scale and b is the background scattering intensity in cm^{-1}. The bounds for A are set between 0 and 100 and the bounds for b are set between 0 to 10 for the multiple initializations.

Broad peak model. For physical gel systems, the correlation length and the peak components are important to determine the material properties. The characteristic structural information for many polymer gels can be obtained by fitting to the empirical broad peak model[32–34] based on the correlation length model[34–41] to demonstrate the effectiveness of this algorithm on empirical models. The fitting function for scattering intensity is

$$I(q) = \frac{A'}{q^n} + \frac{C'}{1 + (|q - q_0|\xi)^m} + b \quad (3)$$

where q_0 is the peak position of the primary peak in q, ξ is a local correlation length, n and m are Porod and Lorentzian exponents, respectively, A' and C' are empirical parameters, and b is the incoherent background scattering intensity in cm^{-1}. The range of q_0 is set to 0.01 to 0.07. The range of ξ is set to 0 to 100. The range of m and n is set to 0 to 5. The range of A' is set to 0 to 0.1. The range of C' is set to 0 to 10 times the maximum value of intensity. Following fitting methods of the same system and model functions,[29] the fitting is first performed on the low q region using $\frac{A'}{q^n}$. After obtaining values of A' and n, those two values are fixed, and the fitting is performed over the entire q range with eqn (3) to obtain the rest of the fitting parameters. The peak position q_0 can be used to calculate the domain spacing d_0, as illustrated in Fig. 1(a).

$$d_0 = \frac{2\pi}{q_0} \quad (4)$$

Spherical micelle model with polydispersity. Pluronic F-127 is a triblock copolymer poly(ethylene oxide)$_{99}$–poly(propylene oxide)$_{69}$–poly(ethylene oxide)$_{99}$[42] (PEO_{99}–PPO_{69}–PEO_{99}). Above the critical micelle concentration, this polymer self-assembles into spherical micelles as illustrated in Fig. 1(c). Spherical micelle model with polydispersity has been chosen based on literature.[43] Here, the parameters of interest are the radius of gyration R_g of the PEO and the radius of the PPO core R. The bounds for R_g are set between 10 to 80 Å and the bounds for R are set between 20 to 80 Å based on literature.[43] The analytical expression of the form factor $P(q)$ in the spherical micelle model[44,45] contains four terms that contribute to the scattering intensity: the self-correlation term of the core, the self-correlation term of the chains, the cross-term between the core and the chains, and the cross term between different chains:[46]

$$P(q,R) = N^2\beta_s^2F_s(q) + N\beta_c^2F_c(q) + 2N^2\beta_s\beta_cS_{sc}(q) + N(N-1)\beta_c^2S_{cc}(q) \quad (5)$$

Here, β_s and β_c are the total scattering length of the core and the corona, respectively, calculated as

$$\beta_s = V_s(\rho_s - \rho_{solvent}) \quad (6)$$

$$\beta_c = V_c(\rho_c - \rho_{solvent}) \quad (7)$$

where V_s and V_c are the volumes of a block in the core and in the corona of a single copolymer molecules. Based on literature reported values,[47] V_s is set to 6283 Å^3 and V_c is set to 14 667 Å^3.

N is the aggregation number of the micelle and can be calculated as follows

$$N = \frac{\frac{4}{3}\pi R^3}{V_s} \quad (8)$$

ρ_s and ρ_c are the corresponding scattering length densities and $\rho_{solvent}$ is the scattering length density of the solvent, which is deuterated water in this case. ρ_s and ρ_c are fitting parameters, since they can vary depending on the solvent distribution. Based on the scattering length density calculations of PEO and PPO, the bounds for ρ_s are set between 4.0×10^{-6} Å^{-2} to 5.0×10^{-6} Å^{-2} and the bounds for ρ_c are set between 4.5×10^{-6} Å^{-2} to 5.5×10^{-6} Å^{-2}. $\rho_{solvent}$ is set to 6.3×10^{-6} Å^{-2}. The self-correlation term of the core is given as

$$F_s(q) = \Phi^2(qR) \quad (9)$$

where

$$\Phi(qR) = \frac{3[\sin(qR) - qR\cos(qR)]}{(qR)^3} \quad (10)$$

The self-correlation term $F_c(q)$ of the Gaussian chains is given by the Debye function[31] in eqn (1).

The interference cross term between the core and the chains is then

$$S_{sc}(q) = \Phi(qR)\psi(qR_g)\frac{\sin(q[R+dR_g])}{q[R+dR_g]}, \quad (11)$$

where $\psi(x) = [1 - e^{-x}]/x$. The interference term between the chains in the corona is

$$S_{cc}(q) = \psi^2(qR_g)\left[\frac{\sin(q[R+dR_g])}{q[R+dR_g]}\right]^2 \quad (12)$$

where d is the factor to mimic non-penetration of Gaussian chains. The bounds for d are set between 0.5 to 2.0.

The scattering intensity with polydispersity can be expressed as

$$I(q) = n_{density} \times P_{poly}(q) + b, \quad (13)$$

$$P_{poly}(q) = \int f_{SZ}(R,z)P(q,R)dR \quad (14)$$

where $n_{density}$ is the number density of micelle, which is defined as the number of micelles per unit volume in cm^3 and calculated using mass concentration $C_{wt\%}$, aggregation number N, solution density ρ_{sol}, and Avogadro's number N_A.

$$n_{density} = \frac{C_{wt\%}}{100} \times \frac{\rho_{sol}}{N \times \text{Munimer}} \times N_A \quad (15)$$

$P_{poly}(q)$ is the form factor with polydispersity, b is the background from incoherent scattering. The bounds for b are set between 0 to 1, given the low background noise in the data. The distribution $f_{SZ}(R,z)$ is chosen as a Schulz–Zimm distribution of the radius of the PPO core R following literature procedures[48] defined as follows.

$$f_{SZ}(R,z) = \frac{R^z}{\Gamma(z+1)}\left(\frac{z+1}{\langle R\rangle}\right)^{z+1} e^{\left(-(z+1)\frac{R}{\langle R\rangle}\right)} \quad (16)$$

$$z = \frac{1}{\sigma^2} - 1 \quad (17)$$

where $\langle R\rangle$ is the expectation value of the micelle core radius. z is the parameter related to the width of the distribution and σ is the polydispersity factor. Here σ has bounds between 0.1 to 1. The lower bound is set to 0.1 because small values of σ yields very large z, which can cause numerical instability (*e.g.*, overflow or underflow) when evaluating terms such as $(z+1)^{z+1}$. These instabilities can lead to NaN values in numerical implementations, especially in floating-point arithmetic.

The selected models were fit to reduced data as a part of parameter estimation experiments by minimizing the least squared residual weighted by the error in scattering intensity using MATLAB command lsqnonlin. The parameter constraints were set based on the physics of the parameter and literature values. Within the bounds, 100 different initial conditions were randomly generated and supplied to the optimization algorithm to assess the global convergence. The confidence interval was output by nlparci, which calculates 95% confidence intervals from asymptotic normal distributions.[49] Briefly, orthogonal-triangular decomposition is performed on the Jacobian matrix J computed by lsqnonlin,

$$J = QU \quad (18)$$

where Q is the orthogonal matrix and U is the upper triangular matrix.

U is inverted to M.

$$M = U^{-1} \quad (19)$$

The diagonal of the Fisher information matrix F can be computed as

$$F_{i,i} = \sum_j (M_{i,j})^2 \quad (20)$$

The root mean square error (RMSE) is computed as

$$\text{RMSE} = \frac{\|\text{resid}\|_2}{\sqrt{\nu}} \quad (21)$$

where resid is the residual, and ν is the degrees of freedom.

$$\nu = n - p \quad (22)$$

where n is the number of data points and p is the number of parameters.

The sample variance se_i for parameter i is then calculated as

$$se_i = \sqrt{F_{i,i}} \times \text{RMSE} \quad (23)$$

The margin of error delta can be computed as

$$\text{delta}_i = \text{se}_i \times t_{0.025,\nu} \quad (24)$$

where t is the value for a confidence interval calculated from the student t-distribution at 95% confidence with ν degrees of freedom. The confidence interval can then be calculated from

$$\text{CI} = \theta \pm \text{delta} \quad (25)$$

where θ is the vector of parameter values obtained from weighted nonlinear least square fitting.

Other model functions used for model differentiation can be found in the SI, eqn (S1)–(S7).

Monte Carlo bootstrapping

Monte Carlo bootstrapping is a technique used to assess parameter uncertainties in experiments characterized by high cost and a limited number of replicates. In this approach, replicas of the scattering intensity are generated by drawing samples from a normal distribution centered around the measured scattering intensity at each q-value with a variance determined by the squared error in the measured scattering intensity at that q-value. For each scattering condition, 99 replicas were generated using this method.[50,51] The mean and standard deviation of each parameter distribution from fitting 100 datasets consisting of one experimental dataset and 99 bootstrapping replicates were calculated to provide error estimates.

Analysis on SANS intensity errors

From time slicing into 0.01 fraction of counts, each full count SANS data set generates 100 time-sliced SANS experimental replicates. The variance of the 100 intensity values at a given q was calculated and compared against the square of the 100 errors of intensity (ΔI^2), which were output by the SANS reduction file provided by ORNL. The percent difference is calculated as

$$\frac{\text{Var}[I(q)] - \Delta I^2}{\text{Var}[I(q)]} \quad (26)$$

where Var[$I(q)$] denotes the calculated variance from intensities and ΔI^2 denotes the variance outputted by the SANS reduction file.

The scaling analysis on ΔI was selected for low q (q = 0.0151), mid q (q = 0.03999), and high q (q = 0.250). Each ΔI was plotted as a function of the fraction of counts at 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 1. The ΔI values for low q, mid q, and high q were fitted to

$$\Delta I = \frac{\text{scale}}{\sqrt{a}} \quad (27)$$

where a is the fraction of counts, and scale is a fitting parameter. The objective function f for fitting scale to minimize the sum of least squares is

$$f(\text{scale}) = \log(\Delta I) - \log\left(\frac{\text{scale}}{\sqrt{a}}\right) \quad (28)$$

Checking the normality of weighted residuals

Experimentally measured SANS data, y, can be described as

$$y = f(X) + \varepsilon \quad (29)$$

where $f(X)$ is the analytical model function with parameters X, and ε is the residual. For a perfect fit, the residual is random error that can be modeled using a Gaussian distribution with a mean of 0 and standard deviation of ΔI. The weighted residuals, which are residuals normalized by ΔI, should follow a normal distribution with mean of 0 and standard deviation of 1. The weighted residual from fitting the P4 experimental and MC bootstrapping SANS data to the broad peak model were computed and checked for normality. Histograms were generated to examine the distribution of weighted residuals from fitting both the experimental data and the MC bootstrapping data.

Estimating the bias of weighed least squares

After performing parameter estimation with weighted nonlinear least square fitting using full count experimental SANS data, the best fit parameters were used to generate an analytical fitted SANS intensity curve. To simulate measurement noise, Gaussian noise with a mean of 0 and standard deviation of ΔI was added to the analytical fitted intensity curve. Following this workflow, 100 replicates with Gaussian noise were generated to simulate SANS data and fitted to the same model. If the estimator is unbiased, parameters' expectation values from the 100 fitting results should be equal to the parameter values used to generate the simulated SANS curve. The expectation value was computed by taking the mean of each parameter distributions. The bias was computed as

$$\text{Bias } [\%] = \frac{E[X] - X_{\text{true}}}{X_{\text{true}}} \times 100 \quad (30)$$

where X_{true} is the parameter value used to generate the simulated SANS data and $E[X]$ is the expectation value of parameter distribution from fitting the simulated SANS data.

Calculating the Fisher information matrix and its eigenvalue decomposition

The variance is estimated by

$$\sigma^2 = \frac{\text{RSS}}{N - k} \quad (31)$$

where N is the number of data points for the sample. RSS is the residual sum of squares. k is the number of free fitting parameters.

The Fisher information matrix F can be computed from the Jacobian matrix J[52] computed by lsqnonlin

$$F = \frac{1}{\sigma^2} \times J' \times J \quad (32)$$

Eigenvalue decomposition is performed on F to obtain the corresponding eigenvalues and eigenvectors.

Calculating spatial correlation of intensity values

From time slicing into 0.01 fraction of counts, full count SANS data generates 100 time-sliced SANS experiment replicates. The residual function RES(q) of intensity is computed as

$$\mathrm{RES}(q) = I(q) - E[I(q)] \tag{33}$$

where $I(q)$ is the intensity at q, and $E[I(q)]$ is the expectation value or mean of the 100 replicates of intensity at each q. The correlation coefficient matrix of RES(q) as each pairwise q is computed based on the Pearson coefficient using the corr function in MATLAB.

The correlation function $G(\Delta q)$ is calculated as

$$G(\Delta q) = \frac{\sum_{100\ \text{replicates}} \sum_{\text{all } q} \mathrm{RES}(q)\mathrm{RES}(q+\Delta q)}{\sum_{100\ \text{replicates}} \sum_{\text{all } q} \mathrm{RES}(q)^2} \tag{34}$$

Information criteria calculations

The Akaike information criteria (AIC), derived from the asymptotic approximation of the Kullback–Leibler divergence, is often useful for model selection by the maximum likelihood method. AIC considers not only the goodness of fit, but also penalizes the number of parameters. The expression for calculating AIC is

$$\mathrm{AIC} = -2\ln(\mathcal{L}) + 2k \tag{35}$$

where $\mathcal{L}$ is the maximum likelihood estimate and k is the number of free parameters. By assuming independently and identically distributed Gaussian distribution of the data points, the maximum log-likelihood can be approximated as[53]

$$\ln(\mathcal{L}) = -\frac{1}{2}N\log\left(\frac{\mathrm{RSS}}{N}\right) \tag{36}$$

where N is the number of data points for the sample and RSS is the residual sum of squares.

Bayesian information criterion (BIC), on the other hand, is derived from Bayesian probability theory and assumes a Bayesian approach to model selection.[54] It introduces a stronger penalty for model complexity than AIC, aiming to select the model that is most likely, given the data, but also has the fewest parameters.

$$\mathrm{BIC} = -2\ln(\mathcal{L}) + k\log(N) \tag{37}$$

Results and discussion

Parameter estimation task

Fitting the SANS curves of dilute polymer solution (16 mM PEG solution) and spherical micelle (Pluronic F-127) shows that a reduced number of counts can achieve good parameter estimation. Literature has shown that d-DMF is well-approximated as a theta solvent for PEG.[26] The 16 mM PEG solution demonstrated good parameter estimation when fit to the Debye model to extract R_g at a reduced number of counts. The R_g value with 95% confidence interval from fitting the full number of counts is 29.17 ± 0.248 Å. Fig. 2(a)–(d) shows the 1D SANS scattering pattern and nonlinear least square fitting curve as the fraction of counts increases. One hundred random initializations are used for assessing the global convergence of the parameter estimation. The same values of fitted parameters and the weighted least square values were obtained at each fraction of counts as shown in SI. Even when the counts are reduced to 0.01 fraction of the full counts, the value of the R_g is within 5% of R_g from full number of counts as shown in Fig. 3(a). At low fraction of counts, the mean parameter value R_g from fitting bootstrapping replicates deviates from the ground truth obtained from fitting experimental replicates. As the number of counts starts to increase, the parameter values obtained from fitting experimental replicates converge with those obtained from MC bootstrapping due to decreases in ΔI. The mean R_g value from MC bootstrapping is consistent with the experimental data it is simulated from.

The uncertainty of R_g from fitting experimental replicates decreases faster than that of MC bootstrapping and 95% confidence interval as shown in Fig. 3. Three distinct notions of uncertainty are readily accessible: margins of error from 95% confidence intervals of nonlinear least squares fitting, standard deviations of parameters from fitting MC bootstrapping replicates, and standard deviations of parameters from fitting experimental time sliced replicates. As shown in Fig. 3(b), MC bootstrapping overestimates the uncertainty of R_g at any fraction of counts compared to experimental time-sliced replicates. Margin of error from 95% confidence interval underestimates the uncertainty obtained from experimental replicates at 0.01, 0.025, 0.1 fraction of counts, and overestimates at 0.05, 0.25, 0.5 fraction of counts. The uncertainties of R_g in 16 mM PEG solution indicate that the 95% confidence interval is closer to the experimental uncertainty of R_g between 0.01 and 0.1 fraction of counts than MC bootstrapping. Above 0.1 fraction of counts, MC bootstrapping and margin of error from 95% confidence intervals give similar uncertainty values, which both overestimate the true experimental uncertainty.

The uncertainties on the parameter R_g from MC bootstrapping in general follows a $\frac{1}{\sqrt{a}}$ scaling law, where a is the fraction of counts as shown in Fig. 3(b). The exception occurs at 0.01 fraction of counts, where the standard deviation of R_g obtained from MC bootstrapping is significantly higher than the scaling prediction and that from the experimental replicates. This can be attributed to the fact that the variances output from the SANS reduction file exceeds the empirical variances from time slicing. Variances output by the SANS reduction file (ΔI^2) are 485% higher at q = 0.015 and 5.63% higher at q = 0.25 compared to variances from the empirical intensity distributions (Var[$I(q)$]) at corresponding q values. The bootstrapping parameter error scaling allows users to decide the optimal number of counts to reach a desired parameter uncertainty level from a short SANS scan. Since the true uncertainty from fitting experimental replicates is always lower than those from

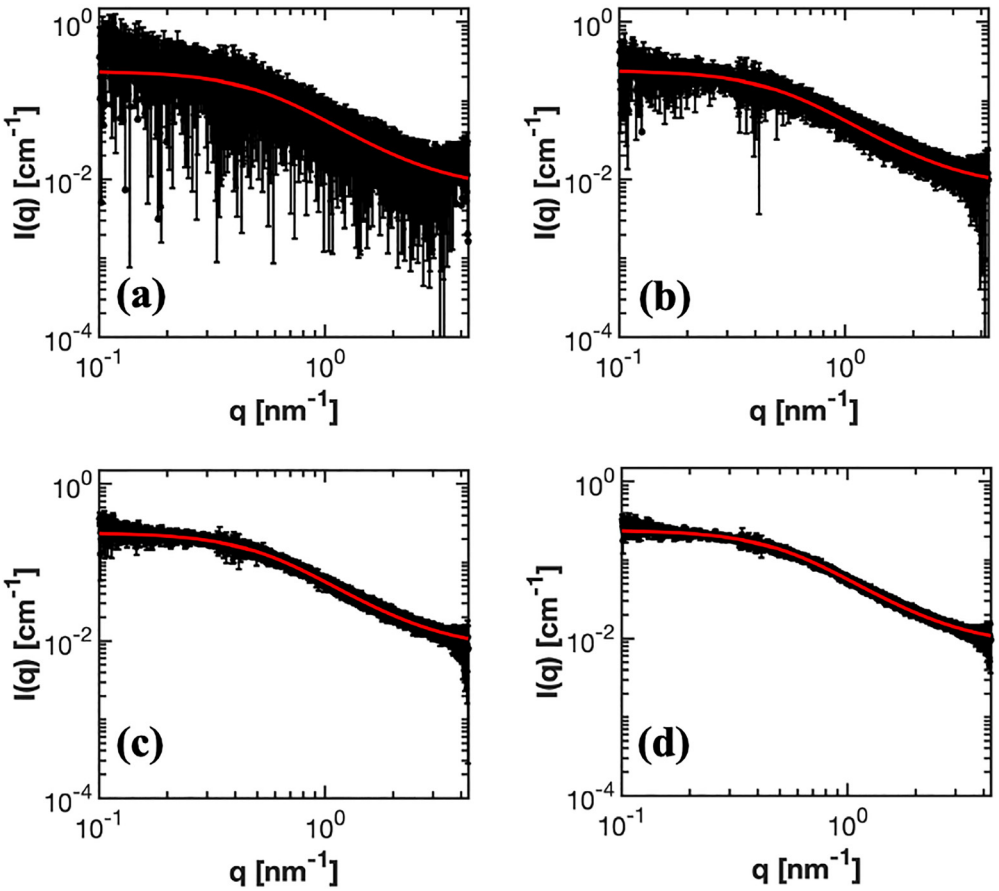

Fig. 2 SANS intensity curves for 16 mM PEG solution in deuterated DMF. (a)–(d) 0.01, 0.05, 0.25, 1 fraction of the total counts. The red line illustrates the fit of each dataset to the Debye model.

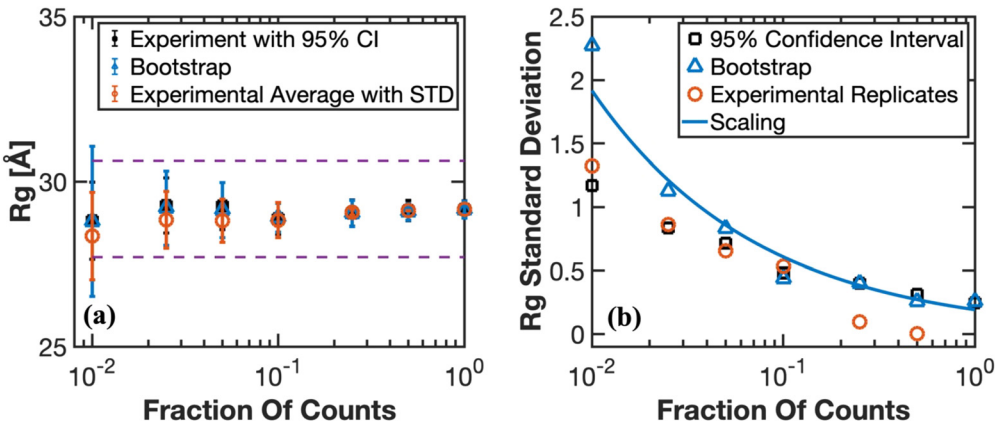

Fig. 3 (a) Radius of gyration, R_g as a function of the fraction of total counts. The black square data points show the R_g values from one experimental dataset and the error bars denote 95% confidence interval. The blue hollow circle shows the mean R_g value fitted from 100 bootstrapping replicates. The error bars denote one standard deviation of the R_g distributions. The orange hollow circles are mean R_g values fitted from averaging the parameters obtained from experimental replicates from time slicing. The dashed blue line represents 5% of the R_g value extracted from the total counts' dataset. (b) The standard deviation of R_g from fitting MC bootstrapping replicates (blue circles), experimental replicates (orange circles), and scaling fitting with respect to bootstrap standard deviation (blue line) as a function of the fraction of total counts.

MC bootstrapping, using the MC bootstrapping error can provide a reliable and stopping criterion.

Parameter estimation analysis at reduced number of counts is further investigated on a block copolymer micelle system which has found many important applications in drug delivery,[55] imaging,[56] biosensing,[57] removing hydrophobic pollutants.[58] The spherical micelle model with polydispersity provides a model for the SANS data from a Pluronic F-127

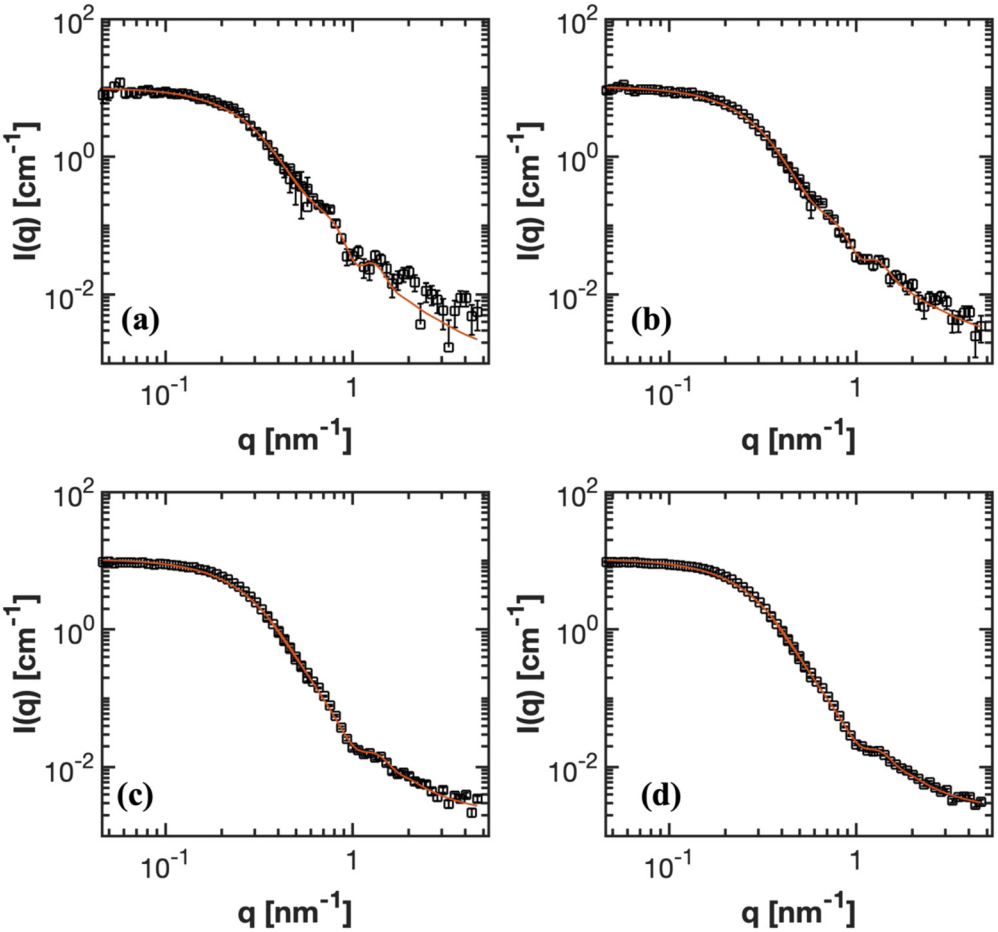

Fig. 4 SANS intensity curves for Pluronic F-127 in deuterated water. (a)–(d) 0.01, 0.05, 0.25, 1 fraction of the total counts. The red line illustrates the fit of each dataset to the spherical micelle model.[44]

micellar solution.[44] Fig. 4(a)–(d) shows the 1D scattering curves for Pluronic F-127 with increasing number of counts fitted to the spherical micelle model to estimate R, the micelle core radius, and R_g, the radius of gyration of the polymer chain in the corona. One hundred random initializations are used for assessing the global convergence of the parameter estimation. The same values of fitted parameters and the weighted least square values were obtained at each fraction of counts as shown in SI. Among the scattering models tested, the spherical micelle model with polydispersity contains the largest number of fitting parameters (7 fitting parameters) with numerical integrations and is the most challenging to converge. The fact that all optimizations still converged to the global minimum demonstrates that the fitting algorithm is capable of robust parameter estimation given that the parameters are identifiable from eigenvalue and eigenvector analysis, even in the presence of high data uncertainty.

For Pluronic F-127, the micelle core radius (R) shows close agreement at all fractions of counts from 0.025 to 1 from fitting experimental replicates and MC bootstrapping replicates, and the parameter value R_g, radius of gyration of the polymer chain in the corona, shows agreement as the fraction of counts approaches 0.25. The value of R is 42.26 $\pm$ 1.20 Å, and the R_g value is 28.92 $\pm$ 2.78 Å from fitting the SANS data at the full number of counts, where the error is the margin of error from a 95% confidence interval. The value of R for all fraction of counts except for 0.01 remains within 5% of the full counts as shown in Fig. 5(a). As shown in Fig. 5(b), R_g remains 10% of the full counts starting at 0.1 fraction of counts. R captures the position of the peak at 1.32 nm^{-1}, making the fitting parameter more robust against noise than R_g. R_g displays larger error even with fitting experimental replicates. Therefore, for the Pluronic F-127 block copolymer micelle, the number of counts can be reduced by a factor of four while still achieving good parameter estimation for both R and R_g.

For Pluronic F-127 fitted to the spherical micelle model, MC bootstrapping and margin of error from 95% confidence intervals can either underestimate or overestimate the experimental uncertainty depending on the parameters of interest and the fraction of counts. MC bootstrapping estimates the experimental uncertainty with greater similarity for parameter R to experimental time-slicing replicates compared to the margin of error from 95% confidence intervals at lower count values, while at higher count values all estimators become comparable.

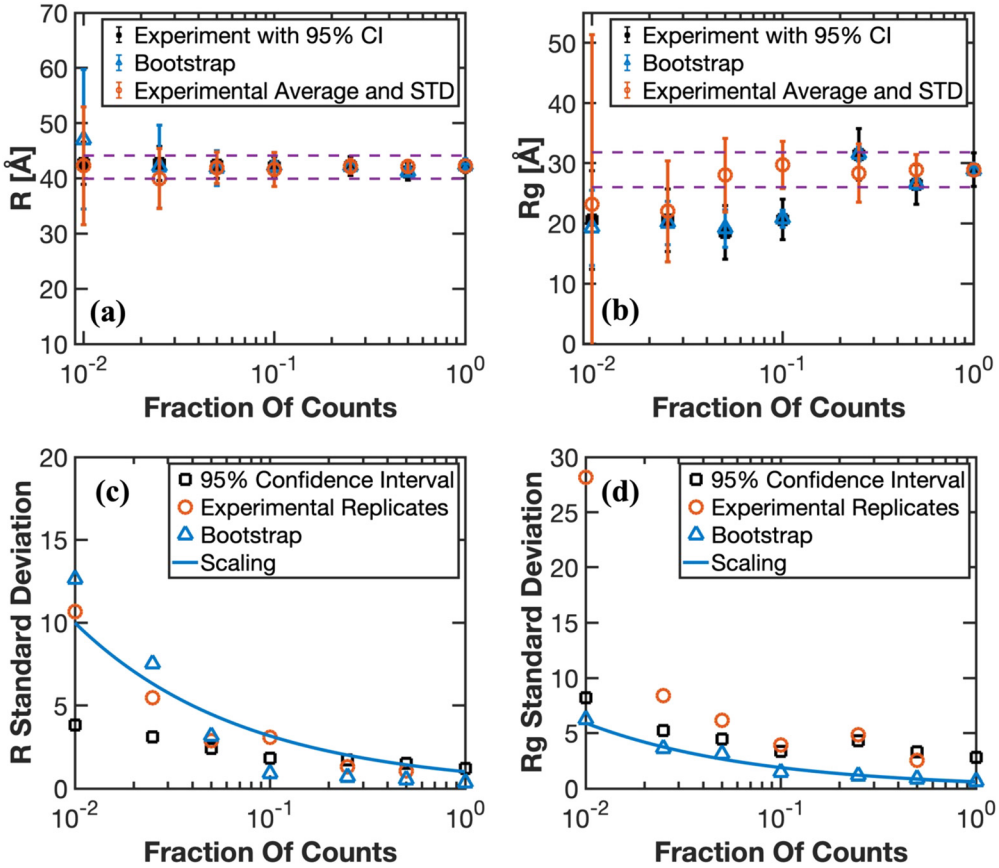

Fig. 5 (a) Micelle radius (R) as a function of the fraction of total counts. The error bars are 95% confidence intervals from the curve fitting. The dashed line represents 5% of the R value extracted from the total counts' dataset. (b) Radius of gyration (R_g) as a function of the fraction of total counts. The error bars are 95% confidence intervals from the curve fitting. The dashed yellow line represents 10% of the R_g value extracted from the total counts' dataset. (c) The standard deviation of R from fitting MC bootstrapping replicates (blue circles), experimental replicates (orange circles), and scaling fitting (blue line) as a function of the fraction of total counts. (d) The standard deviation of R_g from fitting MC bootstrapping replicates (blue circles), experimental replicates (orange circles), and scaling fit with respect to bootstrap standard deviation (blue line) as a function of the fraction of total counts.

On the contrary, for parameter R_g, the margin of error from 95% confidence intervals is closer to experimental uncertainty values at lower fraction of counts.

Similar to the uncertainty scaling of R_g in 16 mM PEG solution, R and R_g in the spherical micelle also follow a $\frac{1}{\sqrt{a}}$ scaling law when fitted to Pluronic F-127 data as shown in Fig. 5(c) and (d), where a is the fraction of counts. The scaling law can reasonably capture the trend of decrease in parameter uncertainties associated with experimental replicates and MC bootstrapping replicates, allowing users to quickly estimate the optimal counts at the beamline.

This parameter estimation analysis enables systematic optimization of neutron counts in SANS measurements based on a user-defined certainty level for the parameters of interests. By performing a brief initial acquisition ($\sim$5000 counts) and fitting the data to relevant model functions, parameter values can be estimated, and parameter uncertainties assessed *via* MC bootstrapping. Leveraging the scaling relationship of parameter uncertainty with neutron count number $\left(\text{uncertainty} \sim \frac{1}{\sqrt{\text{neutron count number}}}\right)$, the algorithm quickly predicts the optimal count required to a user-defined certainty level, such as 5% of the parameter values, providing systematic optimization of valuable SANS beamtime while maintaining desired signal to noise ratio.

This algorithm for parameter estimation has two limitations: (1) computational and (2) requires prior knowledge of the structural model. Computationally, the algorithm requires online reduction of the SANS data, applying MC bootstrapping and model fitting for the short SANS scans to predict additional measurement time/counts needed for the experiment. MC bootstrapping and model fitting take less than 20 seconds to compute on average, which is smaller than the time scale of measurements. Therefore, online implementation should be possible. Second, if the structural model is unknown, this algorithm cannot be applied for parameter estimation. In this case, the user is faced instead with the model differentiation task described below and would use that approach instead.

Parameter estimation with biased estimators

Protein hydrogels "P4", made of molecules that consist of four rodlike associating coiled-coil domains ("P") linked by flexible

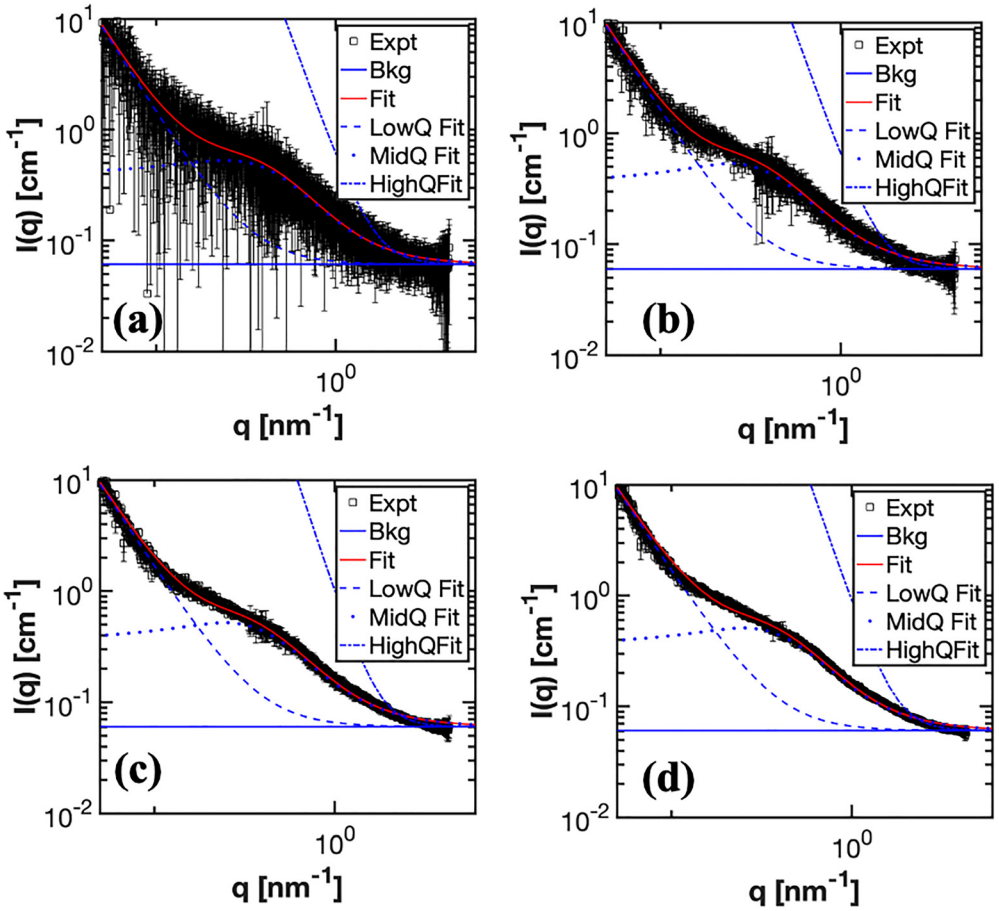

Fig. 6 SANS intensity curves for P4 hydrogel in deuterated 100 mM phosphate buffer, pD = 7.6. (a)–(d) 0.01, 0.05, 0.25, 1 fraction of the total counts. The red line illustrates the fit of each dataset to the broad peak model.

strands ("C_{10}"), present a more complex parameter estimation challenge but still demonstrates good parameter estimation with a reduced number of counts. Above the overlap concentration (~5% w/v), the protein forms an unentangled physical gel held by coiled-coil association.[59] Fig. 6(a)–(d) shows the 1D scattering pattern as the number of counts increases for P4 protein hydrogel fitted to the broad peak model to capture the structure of the protein network. From fitting the total counts scattering curves of P4, the correlation lengths (ξ), which captures the broadness of the peak, is 2.8 nm with 0.14 nm margin of error from 95% confidence interval and 0.04 nm error from MC bootstrapping. The peak component wave vector (q_0) is 0.25 nm^{-1} with 0.026 nm^{-1} margin of error from 95% confidence interval and 0.00298 nm^{-1} error from MC bootstrapping. As shown in Fig. 7(a) and (b), the parameter estimation of P4 requires more counts than the Debye model to achieve parameter estimation within 5% error tolerance. q_0 achieved parameter estimation within 5% of the error tolerance for all fractions of counts. However, ξ requires fraction of counts to 0.25 to produce correlation length within 5% of the error tolerance for ξ. Because the nonlinear model is more complex and has more fitting parameters compared to Debye model, the margin of error of 95% confidence interval associated with ξ is 0.14 nm, which is the same magnitude as the 5% of ξ fitted from the full counts SANS curve. Therefore, for fitting of more complicated models the number of counts or measurement time can be still reduced while achieving good parameter estimation, but the level of reduction is smaller than for simpler models.

The q_0 values obtained from experimental data and bootstrapping replicates shows close agreement starting at 0.1 fraction of counts; however, the mean of ξ from MC bootstrapping constantly exceeds the value obtained from the experimental replicate even at full number of counts as shown in Fig. 7(a) and (b). Two factors may contribute to this discrepancy: (1) the broad peak model cannot capture the physics of the protein hydrogel, and (2) MC bootstrapping provides a biased estimator for ξ. The diagonal entries of the Fisher information matrix represent how precise the estimates are. The smaller the entry is, the harder it is to identify for the corresponding parameter. In this model, all the eigenvalues of the fisher information matrix are nonzero, where the smallest eigenvalue of the fisher information matrix is 3.489. The eigenvalues for the remaining parameters are much greater than the smallest eigenvalue on the order ranging from 10^3 to 10^{13}, indicating good identifiability. This is also

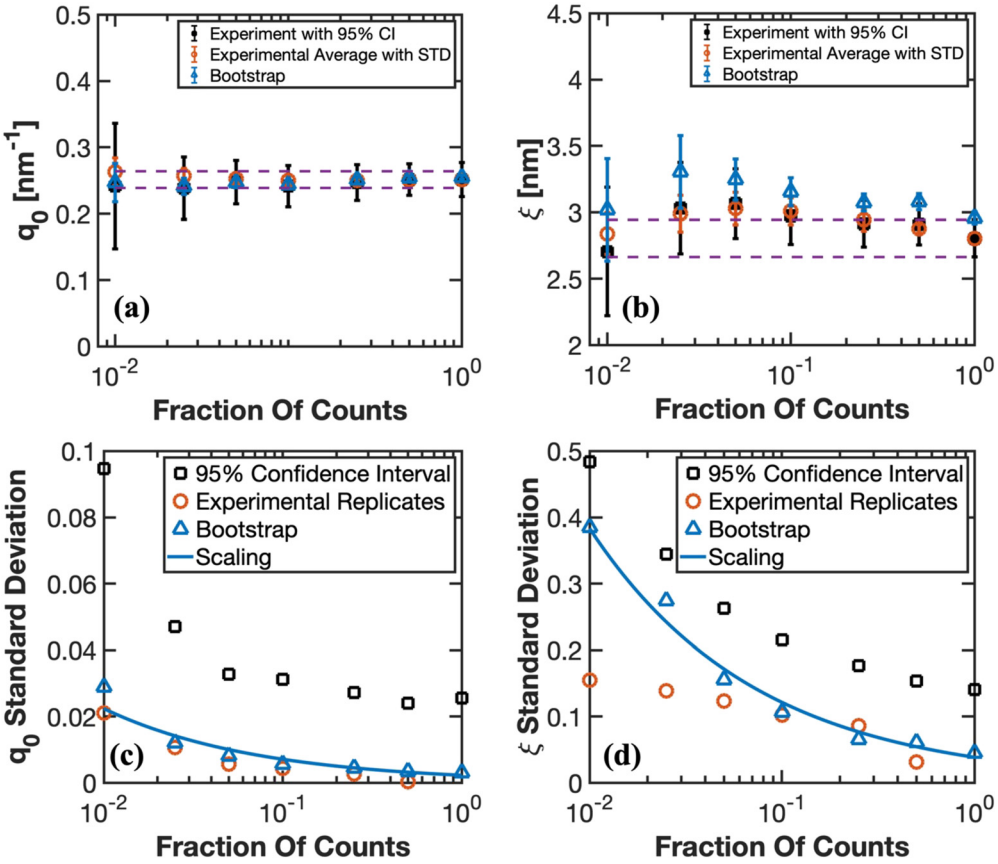

Fig. 7 (a) Peak component wavevector (q_0) as a function of the fraction of total counts. The error bars are 95% confidence intervals from the curve fitting. The dashed line represents 5% of the value extracted from the total counts' dataset. (b) Correlation length (ξ) as a function of the fraction of total counts. The error bars are 95% confidence intervals from the curve fitting. The dashed line represents 5% of the ξ value extracted from the total counts' dataset. (c) Standard deviation of q_0 as a function of the fraction of total counts. (d) Standard deviation of ξ as a function of the fraction of counts. (c) and (d) Black squares represent uncertainty values from margin of error. Blue triangles represent uncertainty values from MC bootstrapping replicates. Orange circles represent uncertainty values from experimental replicates (orange circles). The blue line represents scaling of uncertainty values from MC bootstrapping replicates.

supported by the full convergence of multiple initializations of the parameter estimation. One hundred random initializations are used for assessing the global convergence of the parameter estimation in the broad peak model. The same values of fitted parameters and the weighted least square values were obtained at each fraction of counts as shown in SI. The eigenvector corresponding to the smallest eigenvalue is $[0.0028\ 0.0394\ 0.00064\ 0.9992]^{T}$. This indicates that the hardest identifiable direction is mostly in the ξ parameter that corresponds to the largest entry 0.992.

Literature finds that the broad peak model often fails to generate a good fit to experimental data, which they attribute to the broad peak model's inability to model the network heterogeneity in protein hydrogel systems.[29,60] Our results for the P4 hydrogel systems are broadly consistent with these findings. The distribution of the weighed residuals from fitting the experimental data and bootstrapping data are compared as outlined in the methods section. As shown in Fig. 8, the histogram of weighed residuals from fitting the experimental data has its largest bin to the left of 0. The weighted residual from fitting bootstrapping replicates is less skewed than that of experimental data. Quantitatively, the skewness of the weighted residual distribution, which are 0.47 and 0.27 for experimental data and bootstrapping data, respectively. MC bootstrapping cannot capture the higher degree of skewness that was present in the weighted residual in fitting experimental data. When the model cannot adequately capture the data, the parameter mean from MC bootstrapping can fail to agree even at high fraction of counts.

The weighted least square estimator is a biased estimator for the broad peak model, which can also cause disagreement of parameter values from fitting experimental data and bootstrapping data at full number of counts. The weighted least square estimator is an unbiased estimator for linear models when the residuals follow a Gaussian distribution.[61] However, as shown in Table 1, from the bias testing, the expectation value $E[\xi]$ is 17% higher than the ξ value used for generating the simulated SANS curve, which is consistent with the bias observed in Fig. 7(b). For all the fitting parameters in P4, there are various degrees of bias associated with the weighted nonlinear least square estimator, and even Gaussian noise based on standard deviation ΔI from SANS reduction file can deviate the fitting

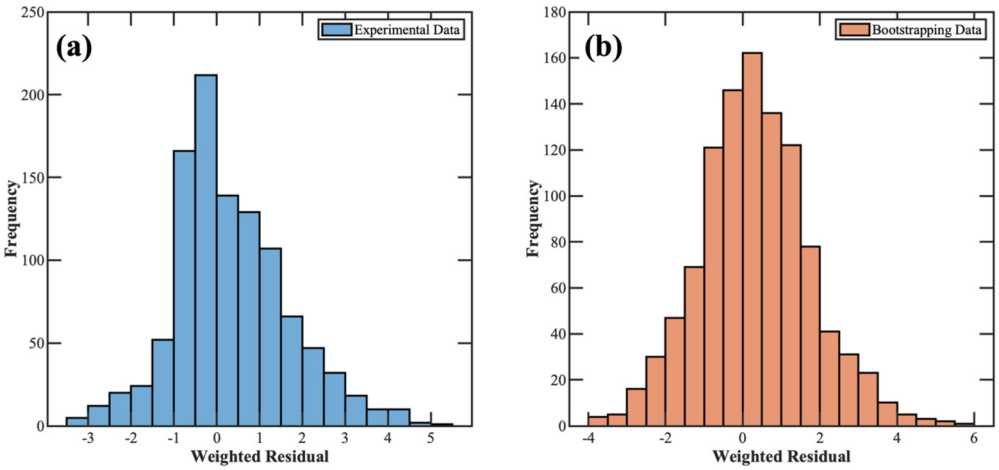

Fig. 8 Histogram comparison from plotting the weighted residual from fitting the broad peak model to (a) P4 experimental data from full counts and (b) one MC bootstrapping replicate simulated from the same P4 experimental data.

parameters from the ground truth. Bias from parameter estimation is intrinsic to the estimator and the model. Under the covariance analysis of broad peak model in Fig. S14 and Table S1, all parameters from the model will be biased to different degrees with unsuited estimator. When the most biased parameter A is fixed to the value obtained at full counts, the absolute values of bias in the other parameters are reduced to various degrees. The largest bias reduction occurs in the parameter q_0, which the bias decreases from 36.42% to 17.71%. Table S2 summarizes the degrees of bias calculated using eqn (22) for all parameters. When certain parameters can be known from either literature, complementary experimental characterization or simulations, the effect of bias in the weighted nonlinear least squares estimator can be greatly reduced. When the estimator is biased for the best model for certain SANS data, error quantification should include the degree of bias and the standard deviation of parameter values from fitting simulated SANS data with Gaussian noise.

For both q_0 and ξ, MC bootstrapping matches better with experimental uncertainties than margin of error from 95% confidence interval despite the biased estimator as shown in Fig. 7(c) and (d). Similar to 16 mM PEG solution and Pluronic F-127, both MC bootstrapping and margin of error in general predict parameter uncertainties for q_0 and ξ that are higher than experimental uncertainties. The P4 protein hydrogel's parameters show similar scaling of uncertainties when fitted to the broad peak model even though the estimators are biased. The standard deviation of q_0 is higher than the scaling prediction at 0.01 fraction of counts, as indicated by the 238% and 0.35% higher variances at low q and high q, respectively, in intensity for the 0.01 fraction of counts. For other fractions of counts, the scaling $\frac{1}{\sqrt{a}}$ matches well with the standard deviation of q_0 obtained from MC bootstrapping. The parameter scaling of ξ matches well with bootstrapping uncertainties at all fractions of counts. Therefore, MC bootstrapping remains an effective means for estimating minimum experimental times to determine a parameter within a given error tolerance even for models for which the estimator is significantly biased.

Table 1 Comparison between true parameter X_{true} from fitting the P4 experimental data and the expectation value from fitting simulated data $E[X]$ to the broad peak model. The bias is the percent difference between X_{true} and $E[X]$

Parameter X	X_{true}	$E[X]$	Bias (%)
A	2.00×10^{-5}	9.04×10^{-5}	350
n	2.46	2.17	−11.5
C	0.45	0.36	−19.62
m_0	1.87	1.66	−11.17
q_0	0.025	0.0343	36.42
ξ	28.04	32.79	16.94

This accelerated SANS algorithm can present significant time savings at the beamline when the structural model is known *a priori*. MC bootstrapping takes less than 20 seconds on average on a laptop to compute. For the three samples used for the study, 16 mM PEG solution was measured for 52 min, P4 hydrogel was measured 29 min, and Pluronic F-127 was measured for 10 min to reach the desired neutron counts. Since measurement time scales linearly with neutron counts, measurement time can be reduced to 0.52 min for 16 mM PEG solution, 15 min for P4 hydrogel, and 2.5 min for Pluronic F-127 while obtaining good parameter estimation. In the case of contrast matching, the algorithm can still be applied for optimizing beamtime if the structural model is known, but longer counting times may be required to account for incoherent background.

This work builds on advances in literature by extending the focus from SANS intensity profile data reconstruction to model-based parameter estimation. Chen and co-workers demonstrated that Gaussian process regression (GPR) can reconstruct smooth, noise-reduced scattering profiles from sparse data without a model for the scattering function, effectively decreasing the number of neutrons required to produce smooth data.[24] The workflow described here addresses the next step in the

analysis pipeline: extracting physical parameters with quantified uncertainties, testing for bias, and informing real-time experimental decisions.

Model differentiation

In addition to parameter estimation, SANS can be used to test different structural hypotheses about a sample by comparing the goodness of fit for candidate structural models when the model is unavailable *a priori* to the experiments. To test whether a reduced number of counts can reliably perform model differentiation, competing SANS models are fitted to the SANS scattering data, and the AIC and BIC for each model are computed at different fractions of counts. The best model is selected based on the lowest AIC or BIC value at each fraction of counts. The protein gel P4 and the Pluronic F-127 micelles are used for model differentiation because those classes of materials can be fit to various SANS models.

AIC and BIC can differentiate the best model even at reduced fractions of counts from the competing models. The candidate models for the P4 protein gel are the broad peak model,[32] the fine scale polymer gel,[40,62] the Gauss–Lorentz gel model,[63] the Debye–Bueche model,[64] and the Ornstein–Zernike and squared Lorentz model.[41] All models can be used to describe polymer network systems depending on the different nanostructures. As shown in Fig. 9, the magnitude of AIC and BIC for the same model are the same across different numbers of counts because the maximum likelihood estimates dominate the value of AIC and BIC in comparison to the number of parameters in the models. The Debye–Bueche and the Ornstein–Zernike and squared Lorentz models have consistently higher AIC values regardless of the fraction of counts, which indicates that the models are unsuited to describe the scattering intensities of P4 protein hydrogel as shown in Fig. 10(d), (e) and SI, Fig. S10, S11. The constant AIC originates from the large residual sum of squares from the lower q region where the scattering intensities of the models plateau but the P4 scattering intensity is decreasing.

The broad peak model, the fine scale polymer gel model, and the Gauss Lorentz gel model all fitted similarly well for the lower number of counts from 0.01 to 0.025 fractions of counts. The broad peak model has the lowest AIC and BIC values among all the models. As the fraction of counts increases to 0.05 and above, the broad peak model yields significantly lower AIC and BIC values. At 0.25 fraction of counts, the AIC and BIC fine scale polymer gel model and the Gauss Lorentz gel model starts to reach relatively constant values. However, the AIC or BIC values for the broad peak model continue to decrease. The fitting results at full number of counts for each model are shown in Fig. 10. From inspecting the fitting curves, the broad peak model can capture the low q and mid q region better than both the fine scale polymer gel model and the Gauss Lorentz model, which is consistent with the intuitive notion of goodness of fit. The best model from model differentiation agrees with the choice of model from previous published SANS fitting model for P4 protein hydrogel.[29] From comparing the AIC and BIC values, the broad peak model is the most suited model for P4 even at very low number of counts.

MC bootstrapping for model differentiation suggests that identifying the best model requires sufficient neutron counts, as shown in Fig. S7(b), which differs from experimental data for model differentiation. This discrepancy arises from the larger ΔI used in MC bootstrapping at lower fraction of counts, evidenced by the 238% and 0.35% higher variances of ΔI than Var[I] at low q and high q, respectively, in intensity for the 0.01 fraction of counts. High ΔI values used for simulating MC bootstrapping replicates introduces large noise to simulated data, increasing the values of AIC and BIC in the broad peak model slightly above the fine-scale polymer gel model at lower fraction of counts. As the fraction of counts increases and ΔI decreases, the SANS curve from MC bootstrapping becomes smoother and ΔI value approaches the true standard deviation in intensity, the SANS curve from bootstrapping smooths, and the broad peak model becomes the best fit.

The analysis of model differentiation on Pluronic F-127 has similar result as P4 protein hydrogel, which shows that AIC or

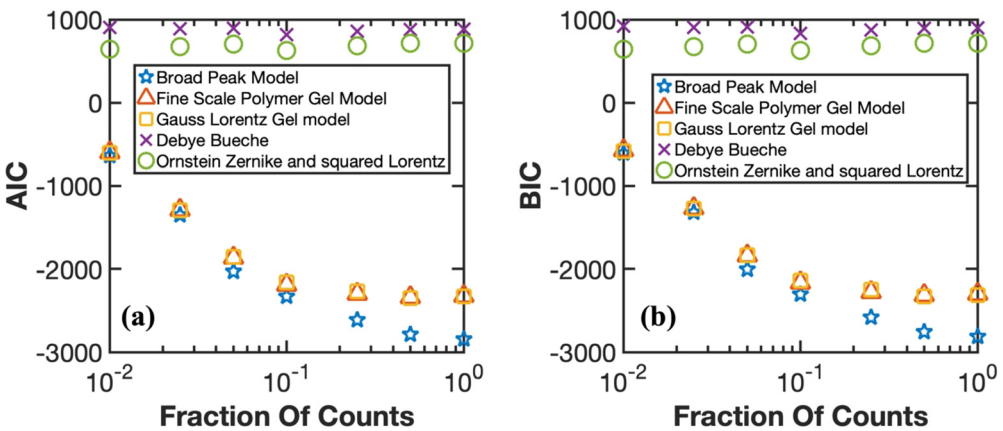

Fig. 9 P4 protein hydrogel model differentiation. (a) AIC as a function of fraction of counts for the candidate models. (b) BIC as a function of fraction of counts for the candidate models.

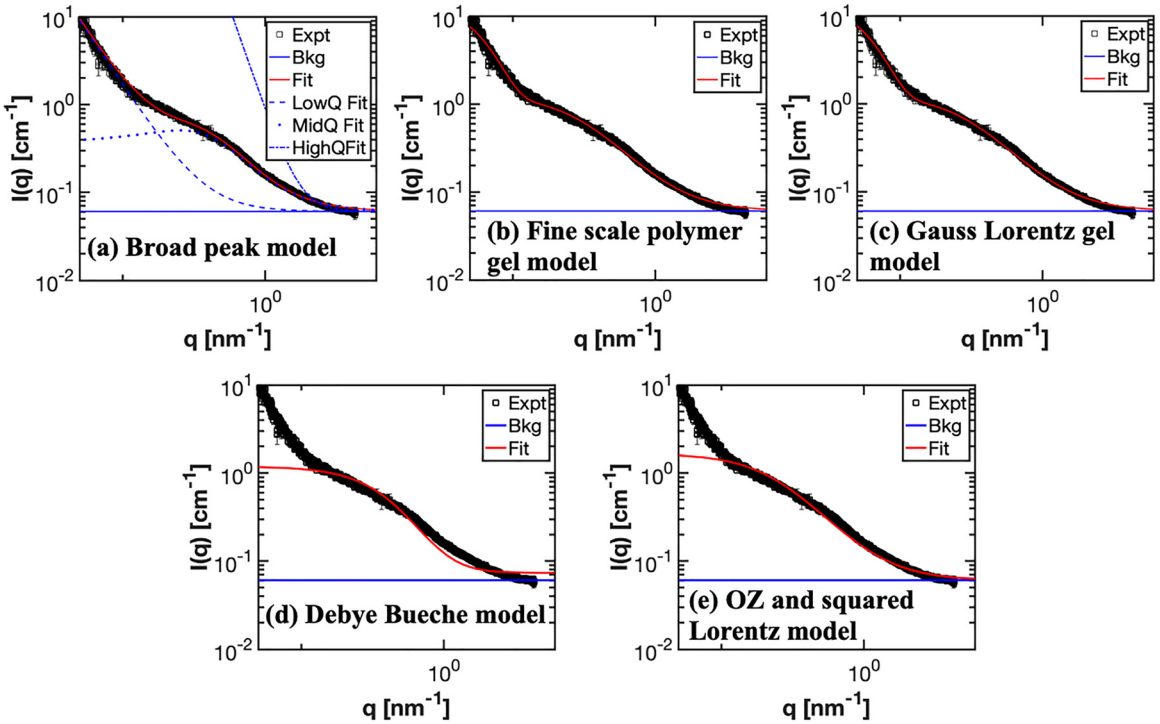

Fig. 10 SANS fitting results of P4 protein hydrogel for full number of counts. (a) Broad peak model. (b) Fine scale polymer gel model. (c) Gauss Lorentz gel model. (d) Debye Bueche model. (e) Ornstein Zernike and squared Lorentz model.

BIC can differentiate models even at a low fraction of counts as shown in Fig. 11. The candidate models for the Pluronic F-127 are the spherical micelle model with polydispersity,[44,48] the sphere model,[65] and the fuzzy sphere model.[66] Since the sample is dilute, the SANS data were fitted to only the form factor in this model differentiation. All models have spherical shapes, and the SANS data should identify micelle formation between spherical shape models with the SANS scattering pattern. The various scattering models for fitting experimental data are shown in Fig. S12 and S13 in the SI. The spherical micelle model with polydispersity has the lowest AIC or BIC across all fractions of counts. Starting at 0.025 fraction of counts, the slope of the AIC or BIC of the spherical micelle model is much steeper than those of the competing models. In this case, the AIC or BIC of the competing models remains relatively constant regardless of the fraction of counts. The AIC or BIC of the spherical micelle model with polydispersity also reaches relatively constant at 0.1 fraction of counts. In this example, AIC and BIC successfully identified micelle formation at 0.1 fraction of counts from spherical model candidates at reduced number of counts, indicating that relatively few counts are required in order to identify the most suitable model.

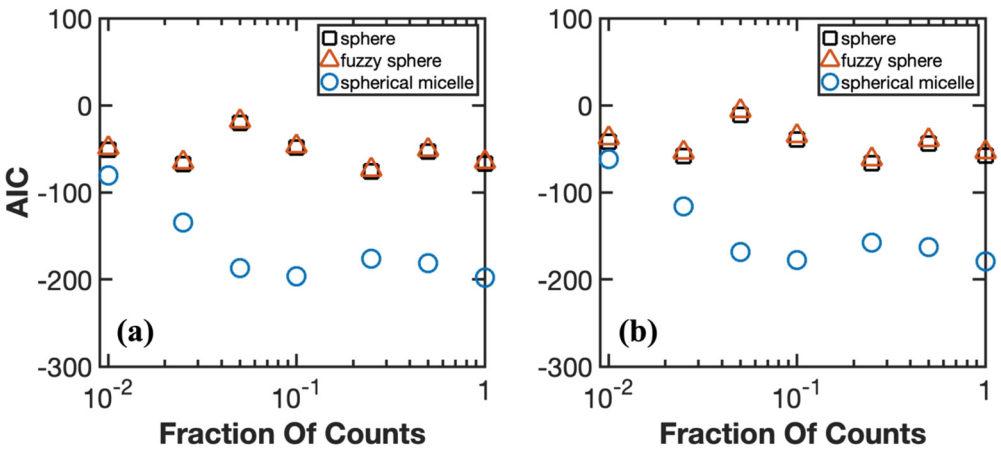

Fig. 11 Pluronic F-127 model differentiation. (a) AIC as a function of fraction of counts for the candidate models. (b) BIC as a function of fraction of counts for the candidate models.

Model differentiation complements, rather than replaces, model development—it provides a systematic method to assess newly proposed scattering functions alongside established models using optimized neutron counts. This approach enables broader validation of novel models across diverse experimental systems, ultimately accelerating the refinement and adoption of physical models in the SANS community.

From the above analysis, SANS users can reliably identify best structural models even at relatively low fraction of counts. An accelerated model differentiation algorithm is proposed to determine optimal neutron counts for data acquisition. Users first select suitable scattering model functions that can capture the targeted nanostructure. SANS measurements with relatively few counts (for example, 5k, 10k, 20k counts) are then acquired, and weighed nonlinear least squares is used to fit each of the candidate model functions on-line, with corresponding AIC and BIC values. The computation of AIC or BIC values for each model allows users to compare both the values and their rate of decrease with increasing neutron counts. The best model is chosen, and data collection for the purpose of model differentiation is stopped, based on the following heuristic: the best model should both minimize AIC or BIC among candidate models for a fixed counts and should have the steepest decrease rate, so that the absolute difference between the best model's AIC and other candidate AIC values are increasing. If the decrease rate is insufficient, additional counts should be acquired until the best model shows a significant AIC or BIC reduction and competing models plateau in AIC or BIC values.

Conclusion

This work proposes an accelerated SANS workflow that uses short SANS acquisitions to predict the optimal number of counts for parameter estimation and model differentiation. For parameter estimation, a short initial acquisition may be used to provide estimates on parameters and uncertainties using MC bootstrapping. Based on this estimate and the scaling of parameter uncertainties $\frac{1}{\sqrt{a}}$, where a is the fraction of counts, the minimum number of counts required to achieve a given error in the parameter may be determined at the beamline. Three representative polymer materials were used to demonstrate that parameter estimation can be achieved with reduced number of counts to within a targeted accuracy. Low error for structural parameter estimation can be obtained using 0.01, 0.25, and 0.5 fraction of the full total counts for the polymer in solution, the spherical micelle, and the associative protein hydrogel, respectively. By leveraging a method that estimates the minimum number of counts necessary to achieve a desired level of uncertainty in parameter estimation, this work addresses the critical challenge of balancing experimental throughput with data quality in SANS experiments.

A key finding of the work was the examination of the bias associated with estimator used in SANS model fitting. The weighted least squares estimator was found to introduce bias when applied to the broad peak model used for the P4 protein hydrogel. Fixing the most biased parameters can substantially reduce the fitting bias in the other parameters. The findings highlight the importance of carefully selecting appropriate estimators depending on the complexity of the model and reducing biases by determining values of certain parameters from either literature or complimentary characterizations or simulations.

For model differentiation, AIC and BIC can reliably differentiate between competing structural models even with reduced number of counts. P4 protein hydrogel and Pluronic F-127 micelles were used for model differentiation task because they can be described by multiple competing models. For both the P4 protein hydrogel and Pluronic F-127 micelles, the broad peak model and spherical micelle model, respectively, were identified as the best-fit models across various fractions of counts. The AIC and BIC values for the best models are consistently the lowest and decrease more rapidly as the count fractions increased, making them distinguishable even with less neutron counts. The ability to differentiate between competing models under such conditions is particularly valuable for high-throughput screening of material libraries, where rapid and accurate structural characterization is essential.

Conflicts of interest

There are no conflicts to declare.

Data availability

Supplementary information is available. See DOI: **https://doi.org/10.1039/d4sm01350f**.

The data and code for analysis is available on **https://github.com/olsenlabmit/Accelerated-SANS-Code.git**.

Acknowledgements

We thank the department of energy (Grant: DE-SC0007106) for supporting this research and the support of Oak Ridge National Laboratory (ORNL) for the neutron scattering facilities in this work. This research used resources at the High Flux Isotope Reactor, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. We acknowledge Dr Lilin He for the experimental assistance and data analysis at ORNL. We thank Dr Ameya Rao and Dr Haley Beech in assistance of the material preparation, Dr Helen Yao for the curve fitting MATLAB code to selected models, and Dr Andrew Salmon for the advice on statistical analysis.

References

1 R.-J. Roe, *Methods of X-ray and neutron scattering in polymer science*, Oxford University Press on Demand, 2000.
2 M. E. Helgeson, S. E. Moran, H. Z. An and P. S. Doyle, Mesoporous organohydrogels from thermogelling photocrosslinkable nanoemulsions, *Nat. Mater.*, 2012, **11**(4), 344–352.

3 E. J. Yearley, I. E. Zarraga, S. J. Shire, T. M. Scherer, Y. Gokarn, N. J. Wagner and Y. Liu, Small-Angle Neutron Scattering Characterization of Monoclonal Antibody Conformations and Interactions at High Concentrations, *Biophys. J.*, 2013, **105**(3), 720–731.
4 G. Baym, Direct Calculation of Electronic Properties of Metals from Neutron Scattering Data, *Phys. Rev.*, 1964, **135**(6A), A1691–A1692.
5 A. C. Wright, A. G. Clare, D. I. Grimley and R. N. Sinclair, Neutron scattering studies of network glasses, *J. Non-Cryst. Solids*, 1989, **112**(1–3), 33–47.
6 A. J. Allen, Characterization of ceramics by x-ray and neutron small-angle scattering, *J. Am. Ceram. Soc.*, 2005, **88**(6), 1367–1381.
7 Y. B. Melnichenko and G. D. Wignall, Small-angle neutron scattering in materials science: Recent practical applications, *J. Appl. Phys.*, 2007, **102**(2), 3.
8 P.-G. De Gennes and P.-G. Gennes, *Scaling concepts in polymer physics*, Cornell University Press, 1979.
9 M. Doi, S. F. Edwards and S. F. Edwards, *The theory of polymer dynamics*, Oxford University Press, 1988, vol. 73.
10 C. M. Jeffries, J. Ilavsky, A. Martel, S. Hinrichs, A. Meyer, J. S. Pedersen, A. V. Sokolova and D. I. Svergun, Small-angle X-ray and neutron scattering, *Nat. Rev. Methods Primers*, 2021, **1**(1), 1–39.
11 L. Li, J. Jakowski, C. Do and K. Hong, Deuteration and Polymers: Rich History with Great Potential, *Macromolecules*, 2021, **54**(8), 3555–3584.
12 M.-C. Chang, Y. Wei, W.-R. Chen and C. Do, Deep learning-based super-resolution for small-angle neutron scattering data: attempt to accelerate experimental workflow, *MRS Commun.*, 2020, **10**(1), 11–17.
13 P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature*, 2016, **533**(7601), 73–76.
14 Z. Chen, N. Andrejevic, N. C. Drucker, T. Nguyen, R. P. Xian, T. Smidt, Y. Wang, R. Ernstorfer, D. A. Tennant, M. Chan and M. Li, Machine learning on neutron and x-ray scattering and spectroscopies, *Chem. Phys. Rev.*, 2021, **2**(3), 031301.
15 Z. Wang, Z. Sun, H. Yin, X. Liu, J. Wang, H. Zhao, C. H. Pang, T. Wu, S. Li, Z. Yin and X.-F. Yu, Data-Driven Materials Innovation and Applications, *Adv. Mater.*, 2022, **34**(36), 2104113.
16 R. Hoogenboom, M. A. Meier and U. S. Schubert, Combinatorial methods, automated synthesis and high-throughput screening in polymer research: past and present, *Macromol. Rapid Commun.*, 2003, **24**(1), 15–32.
17 M. Trobe and M. D. Burke, The molecular industrial revolution: automated synthesis of small molecules, *Angew. Chem., Int. Ed.*, 2018, **57**(16), 4192–4214.
18 C. Zhang, M. W. Bates, Z. Geng, A. E. Levi, D. Vigil, S. M. Barbon, T. Loman, K. T. Delaney, G. H. Fredrickson, C. M. Bates, A. K. Whittaker and C. J. Hawker, Rapid Generation of Block Copolymer Libraries Using Automated Chromatographic Separation, *J. Am. Chem. Soc.*, 2020, **142**(21), 9843–9849.
19 W. T. Heller, J. Hetrick, J. Bilheux, J. M. B. Calvo, W.-R. Chen, L. DeBeer-Schmitt, C. Do, M. Doucet, M. R. Fitzsimmons, W. F. Godoy, G. E. Granroth, S. Hahn, L. He, F. Islam, J. Lin, K. C. Littrell, M. McDonnell, J. McGaha, P. F. Peterson, S. V. Pingali, S. Qian, A. T. Savici, Y. Shang, C. B. Stanley, V. S. Urban, R. E. Whitfield, C. Zhang, W. Zhou, J. J. Billings, M. J. Cuneo, R. M. F. Leal, T. Wang and B. Wu, drtsans: The data reduction toolkit for small-angle neutron scattering at Oak Ridge National Laboratory, *SoftwareX*, 2022, **19**, 101101.
20 S. Qian, W. Heller, W.-R. Chen, A. Christianson, C. Do, Y. Wang, J. Y. Y. Lin, T. Huegle, C. Jiang, C. Boone, C. Hart and V. Graves, CENTAUR—The small- and wide-angle neutron scattering diffractometer/spectrometer for the Second Target Station of the Spallation Neutron Source, *Rev. Sci. Instrum.*, 2022, **93**(7), 075104.
21 M. Steinhart and J. Pleštic, Possible improvements in the precision and accuracy of small-angle X-ray scattering measurements, *J. Appl. Crystallogr.*, 1993, **26**(4), 591–601.
22 K. Saito, M. Yano, H. Hino, T. Shoji, A. Asahara, H. Morita, C. Mitsumata, J. Kohlbrecher and K. Ono, Accelerating small-angle scattering experiments on anisotropic samples using kernel density estimation, *Sci. Rep.*, 2019, **9**(1), 1526.
23 T. Kanazawa, A. Asahara and H. Morita, Accelerating small-angle scattering experiments with simulation-based machine learning, *J. Phys.: Mater.*, 2020, **3**(1), 015001.
24 C.-H. Tung, S. Yip, G.-R. Huang, L. Porcar, Y. Shinohara, B. G. Sumpter, L. Ding, C. Do and W.-R. Chen, Unlocking hidden information in sparse small-angle neutron scattering measurements, *J. Colloid Interface Sci.*, 2025, **692**, 137554.
25 C.-H. Tung, L. Ding, Y. Shinohara, G.-R. Huang, J.-M. Carrillo, W.-R. Chen and C. Do, A convergence metric for counting statistics in time-resolved small angle neutron scattering, *J. Chem. Phys.*, 2025, **163**(7), 074107.
26 H. K. Beech, J. A. Johnson and B. D. Olsen, Conformation of Network Strands in Polymer Gels, *ACS Macro Lett.*, 2023, **12**(3), 325–330.
27 M. Zhong, R. Wang, K. Kawamoto, B. D. Olsen and J. A. Johnson, Quantifying the impact of molecular defects on polymer network elasticity, *Science*, 2016, **353**(6305), 1264–1268.
28 M. J. Glassman, J. Chan and B. D. Olsen, Reinforcement of Shear Thinning Protein Hydrogels by Responsive Block Copolymer Self-Assembly, *Adv. Funct. Mater.*, 2013, **23**(9), 1182–1193.
29 A. Rao, H. Yao and B. D. Olsen, Bridging dynamic regimes of segmental relaxation and center-of-mass diffusion in associative protein hydrogels, *Phys. Rev. Res.*, 2020, **2**(4), 043369.
30 W. T. Heller, M. Cuneo, L. Debeer-Schmitt, C. Do, L. He, L. Heroux, K. Littrell, S. V. Pingali, S. Qian, C. Stanley, V. S. Urban, B. Wu and W. Bras, The suite of small-angle neutron scattering instruments at Oak Ridge National Laboratory, *J. Appl. Crystallogr.*, 2018, **51**(2), 242–248.

31 P. Debye, Molecular-weight determination by light scattering, *J. Phys. Chem.*, 1947, **51**(1), 18–32.
32 F. Horkay and B. Hammouda, Small-angle neutron scattering from typical synthetic and biopolymer solutions, *Colloid Polym. Sci.*, 2008, **286**(6), 611–620.
33 C. E. R. Edwards, D. J. Mai, S. Tang and B. D. Olsen, Molecular anisotropy and rearrangement as mechanisms of toughness and extensibility in entangled physical gels, *Phys. Rev. Mater.*, 2020, **4**(1), 015602.
34 B. Hammouda, D. L. Ho and S. Kline, Insight into clustering in poly (ethylene oxide) solutions, *Macromolecules*, 2004, **37**(18), 6932–6937.
35 M. J. A. Hore, B. Hammouda, Y. Li and H. Cheng, Co-Nonsolvency of Poly(n-isopropylacrylamide) in Deuterated Water/Ethanol Mixtures, *Macromolecules*, 2013, **46**(19), 7894–7901.
36 R. A. Hule, R. P. Nagarkar, B. Hammouda, J. P. Schneider and D. J. Pochan, Dependence of Self-Assembled Peptide Hydrogel Network Structure on Local Fibril Nanostructure, *Macromolecules*, 2009, **42**(18), 7137–7145.
37 E. M. Saffer, M. A. Lackey, D. M. Griffin, S. Kishore, G. N. Tew and S. R. Bhatia, SANS study of highly resilient poly(ethylene glycol) hydrogels, *Soft Matter*, 2014, **10**(12), 1905–1916.
38 T. Matsunaga, T. Sakai, Y. Akagi, U.-I. Chung and M. Shibayama, SANS and SLS Studies on Tetra-Arm PEG Gels in As-Prepared and Swollen States, *Macromolecules*, 2009, **42**(16), 6245–6252.
39 T. Kanaya, M. Ohkura, K. Kaji, M. Furusaka and M. Misawa, Structure of Poly(vinyl alcohol) Gels Studied by Wide- and Small-Angle Neutron Scattering, *Macromolecules*, 1994, **27**(20), 5609–5615.
40 S. Mallam, F. Horkay, A. M. Hecht, A. R. Rennie and E. Geissler, Microscopic and macroscopic thermodynamic observations in swollen poly(dimethylsiloxane) networks, *Macromolecules*, 1991, **24**(2), 543–548.
41 M. Shibayama, K. Isono, S. Okabe, T. Karino and M. Nagao, SANS study on pressure-induced phase separation of poly (N-isopropylacrylamide) aqueous solutions and gels, *Macromolecules*, 2004, **37**(8), 2909–2918.
42 I. W. Hamley, *Block copolymers in solution: fundamentals and applications*, John Wiley & Sons, 2005.
43 T.-H. Kim, Y.-S. Han, J.-D. Jang and B.-S. Seong, SANS study on self-assembled structures of Pluronic F127 triblock copolymer induced by additives and temperatureThis article will form part of a virtual special issue of the journal, presenting some highlights of the 15th International Small-Angle Scattering Conference (SAS2012). This special issue will be available in early 2014, *J. Appl. Crystallogr.*, 2014, **47**(1), 53–59.
44 J. S. Pedersen, Form factors of block copolymer micelles with spherical, ellipsoidal and cylindrical cores, *J. Appl. Crystallogr.*, 2000, **33**(3–1), 637–640.
45 J. S. Pedersen and M. C. Gerstenberg, Scattering Form Factor of Block Copolymer Micelles, *Macromolecules*, 1996, **29**(4), 1363–1365.
46 J. S. Pedersen, Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting, *Adv. Colloid Interface Sci.*, 1997, **70**, 171–210.
47 Y. Lin and P. Alexandridis, Temperature-Dependent Adsorption of Pluronic F127 Block Copolymers onto Carbon Black Particles Dispersed in Aqueous Media, *J. Phys. Chem. B*, 2002, **106**(42), 10834–10844.
48 S. Manet, A. Lecchi, M. Impéror-Clerc, V. Zholobenko, D. Durand, C. L. P. Oliveira, J. S. Pedersen, I. Grillo, F. Meneau and C. Rochas, Structure of Micelles of a Nonionic Block Copolymer Determined by SANS and SAXS, *J. Phys. Chem. B*, 2011, **115**(39), 11318–11329.
49 G. A. F. Seber and C. J. Wild, *Nonlinear Regression*, John Wiley & Sons, New York, 1989.
50 R. N. Pérez, J. Amaro and E. R. Arriola, Bootstrapping the statistical uncertainties of NN scattering data, *Phys. Lett. B*, 2014, **738**, 155–159.
51 H. Yao and B. D. Olsen, SANS quantification of bound water in water-soluble polymers across multiple concentration regimes, *Soft Matter*, 2021, **17**(21), 5303–5318.
52 F. Nielsen, A simple approximation method for the Fisher–Rao distance between multivariate normal distributions, *Entropy*, 2023, **25**(4), 654.
53 K. P. Burnham and D. R. Anderson, Multimodel Inference, *Sociol. Methods Res.*, 2016, **33**(2), 261–304.
54 T. Hastie, R. Tibshirani, J. H. Friedman and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2009, vol. 2.
55 S. Movassaghian, O. M. Merkel and V. P. Torchilin, Applications of polymer micelles for imaging and drug delivery, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 2015, **7**(5), 691–707.
56 K. S. Kim, W. Park, J. Hu, Y. H. Bae and K. Na, A cancer-recognizable MRI contrast agents using pH-responsive polymeric micelle, *Biomaterials*, 2014, **35**(1), 337–343.
57 H. V. Sureka, A. C. Obermeyer, R. J. Flores and B. D. Olsen, Catalytic biosensors from complex coacervate core micelle (C3M) thin films, *ACS Appl. Mater. Interfaces*, 2019, **11**(35), 32354–32365.
58 D. Gokhale, I. Chen and P. S. Doyle, Micelle-laden hydrogel microparticles for the removal of hydrophobic micropollutants from water, *ACS Appl. Polym. Mater.*, 2022, **4**(1), 746–754.
59 V. N. Malashkevich, R. A. Kammerer, V. P. Efimov, T. Schulthess and J. Engel, The Crystal Structure of a Five-Stranded Coiled Coil in COMP: A Prototype Ion Channel?, *Science*, 1996, **274**(5288), 761–765.
60 A. Rao and B. D. Olsen, Structural and dynamic heterogeneity in associative networks formed by artificially engineered protein polymers, *Soft Matter*, 2023, **19**(33), 6314–6328.
61 N. H. Bingham and J. M. Fry, *Regression: Linear models in statistics*, Springer Science & Business Media, 2010.
62 M. Shibayama, T. Tanaka and C. C. Han, Small-angle neutron scattering study on weakly charged temperature sensitive polymer gels, *J. Chem. Phys.*, 1992, **97**(9), 6842–6854.

63 G. Evmenenko, E. Theunissen, K. Mortensen and H. Reynaers, SANS study of surfactant ordering in κ-carrageenan/cetylpyridinium chloride complexes, *Polymer*, 2001, **42**(7), 2907–2913.
64 P. Debye and A. Bueche, Scattering by an inhomogeneous solid, *J. Appl. Phys.*, 1949, **20**(6), 518–525.
65 A. Guinier, G. Fournet and K. L. Yudowitch, *Small-angle scattering of X-rays*, 1955.
66 M. Stieger, J. S. Pedersen, P. Lindner and W. Richtering, Are Thermoresponsive Microgels Model Systems for Concentrated Colloidal Suspensions? A Rheology and Small-Angle Neutron Scattering Study, *Langmuir*, 2004, **20**(17), 7283–7292.