



Soft Matter

A User-friendly Graphical User Interface for Dynamic Light Scattering Data Analysis

Journal:	<i>Soft Matter</i>
Manuscript ID	SM-ART-04-2023-000469.R1
Article Type:	Paper
Date Submitted by the Author:	25-May-2023
Complete List of Authors:	Salazar, Matthew; UCLA, Chemical and Biomolecular Engineering Srivastav, Harsh; UCLA, Chemical and Biomolecular Engineering Srivastava, Abhishek; Cornell University, Sibley School of Mechanical and Aerospace Engineering Srivastava, Samanvaya; UCLA, Chemical and Biomolecular Engineering

SCHOLARONE™
Manuscripts

A User-friendly Graphical User Interface for Dynamic Light Scattering Data Analysis

Matthew Salazar,^{1,†} Harsh Srivastav,^{1,‡} Abhishek Srivastava,^{2,¬} Samanvaya Srivastava^{1,3,4,*}

Abstract

Dynamic light scattering (DLS) is a commonly used analytical tool for characterizing the size distribution of colloids in a dispersion or a solution. Typically, the intensity of a scattering produced from the sample at a fixed angle from an incident laser beam is recorded as a function of time and converted into time autocorrelation data, which can be inverted to estimate the distribution of colloid diffusivity to estimate the colloid size distribution. For polydisperse samples, this inversion problem, being a Fredholm integral equation of the first kind, is ill-posed and is typically handled using cumulant expansions or regularization methods. Here, we introduce a user-friendly graphical user interface (GUI) for analyzing the measured scattering intensity time autocorrelation data using both the cumulant expansion method and regularization methods, with the latter implemented using various commonly employed algorithms, including NNLS, CONTIN, REPES, and DYNALS. The GUI provides default values for all the parameters associated with the data analysis, but, importantly, allows the user to modulate any and all of the fit parameters, offering extreme flexibility. Additionally, the GUI also enables a comparison of the size distributions generated by various algorithms and an evaluation of their performance. We present the fit results obtained from the GUI for model monomodal and bimodal dispersions to highlight the strengths, limitations, and scope of applicability of these algorithms for analyzing time autocorrelation data from DLS.

1. Introduction

Dynamic light scattering (DLS) is a non-destructive spectroscopic technique widely used for determining the size distribution of suspended colloids in dispersions or solutions.^{1–14} Typical applications for this method include measuring the size distributions of proteins, micelles, polymers, and nanoparticles.^{4–13} In practice, a laser beam is directed towards a sample solution/dispersion and the intensity of the scattered light is measured by a photon detector set at a *fixed* angle from the path of the laser and a fixed distance from the scattering sample.^{8,10} Owing to the continuous Brownian motion of the scatterers in the sample solutions/dispersions, the intensity of the scattered light fluctuates over time. The detector measures the fluctuating light intensity as a function of time, which is converted into time autocorrelation data and subsequently analyzed to obtain a distribution of the scatterer's diffusion coefficient, enabling an estimation of the distribution of the scatterer's solvodynamic (typically hydrodynamic) size.^{8,10}

Several methods have been developed for extracting the scatterer size distribution data from the initial intensity-time measurements and can be divided into two broad categories: method of cumulants and regularization methods.^{1,3,8,10–12} The difference between these two groups of methods is how they approach the problem of extracting the diffusion coefficient distribution. The method of cumulants,^{2,15} the simpler method, attempts to fit the modified correlation data to a cumulant generating function with one to four terms. The cumulants of the function provide information on the distribution of the diffusion coefficient such as the mean, variance, skew, and

kurtosis.¹⁵ This method works well for predicting the monomodal polydisperse size distributions because the algorithm inherently assumes a single mean scatterer size but struggles with multimodal size distributions.

Regularization methods, in contrast, seek to fit a diffusion coefficient distribution to the intensity autocorrelation data without employing a predefined function for the size distribution.^{1,16,17} The simplest regularization method is the non-negative least squares method (NNLS),¹⁸ which attempts to find the diffusion coefficient distribution that results in the smallest absolute difference between the autocorrelation data and the corresponding fit while enforcing a constraint that the diffusion coefficient distribution, and as such the size distribution, cannot contain a negative value. Algorithms such as CONTIN,^{17,19,20} REPES,¹⁶ and DYNALS have built upon the NNLS algorithm by introducing side constraints to adjust the curvature of the size distribution. Regularization methods are very flexible algorithms because they do not need a predefined function but can run into the issue of producing wide distributions and, as such, require some background knowledge of the system to narrow down the range of diffusion coefficients (or particle sizes).

Typically, commercial DLS instruments and the software packages that accompany them provide inflexible data analysis options and do not allow for straightforward ways to compare the size distribution predictions from multiple algorithms or the ability to tune or constrain the model parameters, with the latter being especially relevant for regularization algorithms. This “one method fits all data” approach has limited the

¹ Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA 90095

[†] Present address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA

[‡] Present address: Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA 94720

² Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14853

[¬] Present address: Corporate Research Laboratory, 3M Company, St. Paul, MN 55144

³ California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA 90095

⁴ Institute for Carbon Management, University of California, Los Angeles, Los Angeles, CA 90095

* Email: samsri@ucla.edu

accuracy and applicability of DLS as it underestimates the versatility of all the models and fitting algorithm and their respective advantages. Here, we aim to address these software shortcomings by developing a user-friendly graphical user interface (GUI) to analyze the time autocorrelation data from DLS experiments using multiple algorithms, comparisons between their particle size distribution predictions, and performance evaluation for identifying the optimal algorithm for fitting any user generated DLS autocorrelation data.

This paper is organized as follows – we begin by describing the mathematical underpinnings of the two families of algorithms for obtaining the size distributions from DLS autocorrelation data: the method of cumulants and regularization methods. We then briefly overview the experimental methodology and the GUI that has been developed. Then, we present the results and a discussion on the fits obtained from DLS autocorrelation data for model monomodal or bimodal dispersions. Lastly, we summarize and discuss the efficacy of various algorithms and limitations of DLS to obtain particle size distributions.

2. Theory

The DLS setup comprises an incident laser (wavelength λ) at a solution/dispersion (solvent viscosity η and refractive index n) maintained at temperature T and detection of the scattered light using a photon-counting detector stationed at a certain fixed angle θ from the incident beam. Scatterers in a solution/dispersion elastically scatter the incident laser and some of the scattered photons are captured by the photodetector. Since the scatterers undergo continuous Brownian motion, the flux of the photons captured at the detector (i.e., intensity I) varies with time t . The number of photons measured by the photodetector in a typical DLS experiment is used to calculate a normalized time autocorrelation function of the scattered light intensity g_2 as:¹

$$g_2(q, \tau) = \frac{\langle I(q, t)I(q, t + \tau) \rangle}{\langle I(q, t) \rangle \langle I(q, t + \tau) \rangle} \quad (1)$$

Here τ is the correlation time delay and $\langle \rangle$ represent an average of intensities over time t . The wave vector q holds information on the experimental conditions and is defined as:

$$q = \frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad (2)$$

The autocorrelation of the intensity can be related to the field-field time autocorrelation function, $g_1(q, \tau)$ as:

$$g_2(q, \tau) = B + \beta |g_1(q, \tau)|^2 \quad (3)$$

Here the baseline B and β are the values of $g_2(q, \tau)$ as τ approaches infinity and zero, respectively. In general, the field-field time autocorrelation function, $g_1(q, \tau)$, can be expressed as a Laplace transformation:

$$g_1(q, \tau) = \int_0^\infty G(\Gamma) \exp(-\Gamma\tau) d\Gamma \quad (4)$$

Here Γ is the decay rate and $G(\Gamma)$ is the decay rate distribution function normalized to unity, i.e., $\int_0^\infty G(\Gamma) d\Gamma = 1$. With these

relationships as the foundation, both the cumulant and regularization algorithms attempt to extract $G(\Gamma)$. Γ , in turn, is related to the diffusion coefficient, D , as:

$$\Gamma = Dq^2 \quad (5)$$

Once the diffusion coefficient distribution is known, the distribution of the scatterer hydrodynamics radius, R , is estimated using the Stokes-Einstein equation as:

$$R = k_B T / 6\pi\eta D \quad (6)$$

Here k_B is the Boltzmann's constant.

2.1. Method of Cumulants

The cumulant method describes the exponential of the decay rate using a Taylor series expansion around the exponential term of the average decay rate $\bar{\Gamma}$ as:

$$\exp(-\Gamma\tau) = \exp(-\bar{\Gamma}\tau) \left[1 - (\Gamma - \bar{\Gamma})\tau + \frac{(\Gamma - \bar{\Gamma})^2}{2!} \tau^2 - \frac{(\Gamma - \bar{\Gamma})^3}{3!} \tau^3 \dots \right] \quad (7)$$

Substituting the above expression in Eq. 4 results in an approximate field-field correlation function:

$$\ln(g_1(q, \tau)) \cong -k_1\tau + \frac{k_2}{2!} \tau^2 + \frac{k_3}{3!} \tau^3 + \dots \quad (8)$$

Substituting Eq. 8 into Eq. 3 and rearranging yields the following relationship between coefficients:

$$g_2 - B = \beta g_1^2 = \beta \exp\left(-k_1\tau + \frac{k_2}{2!} \tau^2 + \frac{k_3}{3!} \tau^3 + \dots\right) \quad (9)$$

The k_i terms are the cumulants and describe the distribution function; k_1 is the average, k_2 is the variance, k_3 is the skew, and k_4 is the kurtosis of the distribution.

2.2. Regularization Methods

Regularization methods do not attempt to fit a predefined function and require little manipulation of Eq. 4 to find a decay rate distribution. We note, however, that the integral equation in Eq. 4 is a Fredholm integral equation of the first kind. It is an ill-posed problem, such that small errors or perturbations in the g_1 can result in large fluctuations in $G(\Gamma)$. A unique solution to the problem does not exist. For this reason, regularization approaches generally invoke a side constraint to smoothen the size distribution function while achieving minimal deviation of the fit from the experimental data. Discretizing the field correlation function, g_1 , through a Riemann sum yields:

$$g_1(q, \tau_m) = \sum_{n=1}^N \tilde{G}(\Gamma_n) e^{-\Gamma_n \tau_m} \quad (10)$$

Where $\tilde{G}(\Gamma_n)$ is the distribution function of the decay rate and is normalized to unity, i.e., $\sum_{n=1}^N \tilde{G}(\Gamma_n) d\Gamma_n = 1$. The above equation can be rewritten using vector notation as:

$$g_m = A_{m,n} x_n \quad (11)$$

g_m contains the field correlation data, x_n contains the decay rate distribution, $\tilde{G}(\Gamma_n)$, and $A_{m,n}$ is a transfer matrix containing

the decay rate and time delay relationship, $e^{-\Gamma_n \tau_m}$. Using the matrix notation recasts this problem as a minimization problem for the decay rate distribution with the following objective function:

$$V = \|A_{m,n}x_n - g_m\| \quad (12)$$

However, because A is not necessarily a square matrix and there are many possible solutions, a side constraint is typically added to make the solution numerically tractable. Although this constraint varies based on the specific algorithm, the fitting problem can be generalized to take the following form:

$$V = \|A_{m,n}x_n - g_m\|^2 + \alpha^2 \|\omega - \Omega x_n\|^2 \quad (13)$$

Here α is the regularization parameter that dictates the relative weight of the side constraint, ω is a term that stores any prior information of the distribution and Ω is a term that contains some constraint on the decay rate distribution solution x_n . The regularization parameter is chosen by the user and involves a trade-off between reducing the normalized error of the fit and the restrictions on the fit distribution. As such, an L -curve algorithm was developed to provide the user with the optimal regularization parameter.²¹ It is typically assumed that the user would have no prior information about the system and ω is set to zero. The value of the other terms for each of the four regularization methods used by the GUI is summarized in Table 1.

Algorithm	α	Ω
NNLS	0	0
CONTIN	variable	Second Derivative of x_n
REPES	variable	Second Derivative of x_n
DYNALS	0	First Derivative of x_n

Table 1. The values of α and Ω for the regularization methods implemented in the GUI in this work.

3. Experimental Methods and GUI

DLS autocorrelation data from three standard aqueous dispersions of monodisperse spherical silica nanoparticles (NanoXact Silica Nanospheres, nanoComposix) were utilized for the comparison of the algorithms. The diameters of the silica nanoparticles, as provided by the vendor (determined by transmission electron microscopy), are $47 \text{ nm} \pm 3 \text{ nm}$, $118.5 \text{ nm} \pm 5.7 \text{ nm}$, and $194 \text{ nm} \pm 16 \text{ nm}$. The hydrodynamic radii of the particles are also provided by the vendor as 27 nm , 68.8 , and 106 nm , respectively. We will refer to these particles as 25 nm , 65 nm , and 100 nm silica standards in this study. The DLS autocorrelation data were obtained on a BI-200SM goniometer containing a red laser diode with a wavelength of 637 nm and a TurboCorr digital correlator (Brookhaven Instruments, Holtsville, NY). The detector was set at an angle of 90° from the laser, and all experiments were run at 25°C . The refractive

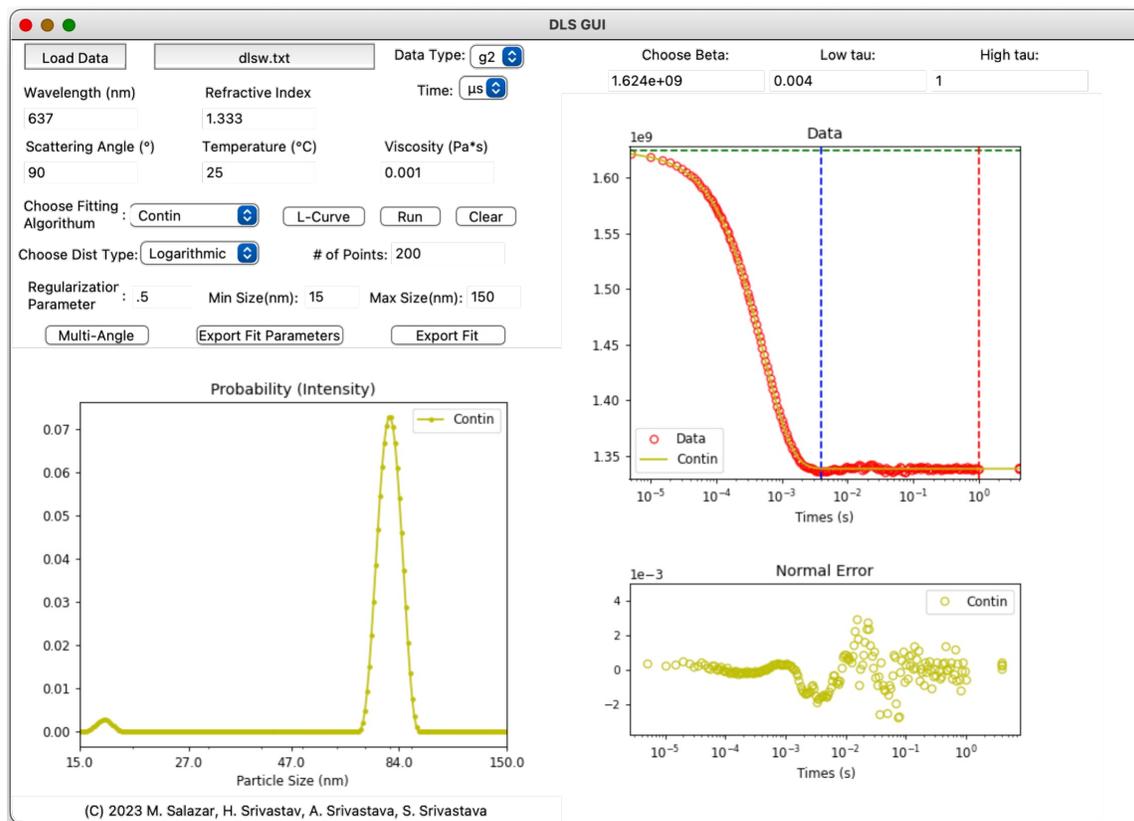


Figure 1. An image of the graphical user interface (GUI) for DLS data analysis. The upper right plot displays the autocorrelation data and the fit to the data. The dotted blue and red lines show the bounds used to find the baseline for the calculations. The dotted green line represents the beta coefficient used for the calculations. The lower right plot displays the normalized error for the fits. The lower left plot displays the size distribution resulting from the selected algorithm.

index and the viscosity of water were 1.333 and 1 cP, respectively. The 100 nm silica standard was used to compare the performance of the algorithm in describing monomodal distributions. Different molar mixtures of 100 nm and 25 nm silica standards were used to investigate the ability of the algorithm to model bimodal distributions. For the CONTIN algorithm, a mixture of 25 nm and 65 nm silica standards was analyzed to investigate the ability of the algorithm to model a bimodal distribution and differentiate between the two similar particle sizes.

A graphic user interface (GUI) was developed in Python to analyze the correlation function data and is available for download as an executable file from these websites.^{22,23} A screenshot of the GUI is shown in **Figure 1**. Once the data file is loaded into the GUI, it provides an estimate of the value of β and a range for the baseline (B) estimation. If needed, these parameters can be adjusted by the user by changing the values in the input boxes on the top right, and an updated plot will be displayed upon clicking the update button. Below the load data button (on the top left), the GUI also has input boxes to allow users to provide the parameters used in the data analysis. A drop-down menu for the selection of the data analysis algorithm allows a selection from four cumulant (linear, quadratic, cubic, and quartic) and four regularization (CONTIN, NNLS, DYNALS, and REPES) methods. Once an algorithm is selected, the GUI provides an estimate of the range of the size distributions for the subsequent analysis. If needed, the user may adjust this range from prior knowledge of the system. For the CONTIN and REPES algorithms, the L -curve button allows for the estimation of the optimal regularization parameter using the L -curve method. Once satisfactory parameters are added, the data analysis updates the correlation function plot wherein a fit to the experimental data is displayed, along with updated plots depicting the size distribution and the normalized error between the experimental data and the model fits. The analysis of the data can be extracted to a text file by selecting the Export Fit button. For creating a size distribution corresponding to the cumulant expansions, the parameters from the fits were included in Gram-Charlier expansion^{24,25} on a Gaussian distribution to provide a probability distribution function of the decay rate.

4. Results and Discussion

We present the results from the analysis of the DLS autocorrelation data for monomodal and polymodal dispersions with commonly used algorithms using the GUI we have developed to highlight its versatility and ease of use. We will demonstrate how the GUI allows a comparative study of the various algorithms with respect to their accuracy and suitability for monomodal and bimodal dispersions. Volume-based size distributions are reported; intensity- and number-based size distributions can be estimated in the GUI if needed.

4.1. Particle Dispersions with Monomodal Size Distributions

The autocorrelation function data for a 100 nm silica standard was imported into the GUI and analyzed using the quadratic

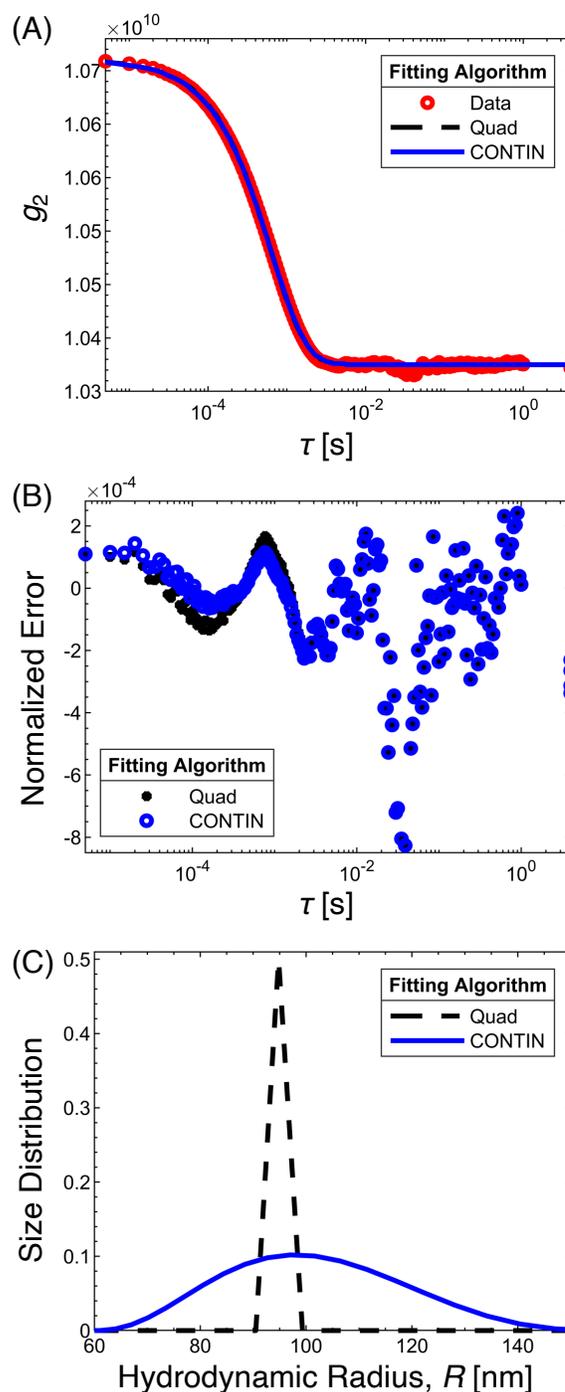


Figure 2. Analysis of DLS data from an aqueous dispersion of 100 nm particles. (A) The experimental correlation data and fits using the quadratic Taylor expansion and CONTIN algorithms ($\alpha = 0.78$). (B) The normalized error of the fits. (C) The size distributions computed by the two algorithms.

cumulant expansion and CONTIN algorithms. As shown in **Figure 2A**, both algorithms were able to fit the correlation function data very well; the normalized error of both fits was less than 10^{-3} , with the quadratic algorithm having a markedly larger error at early delay times (τ) as compared to the CONTIN algorithm (**Figure 2B**). Even with the slight

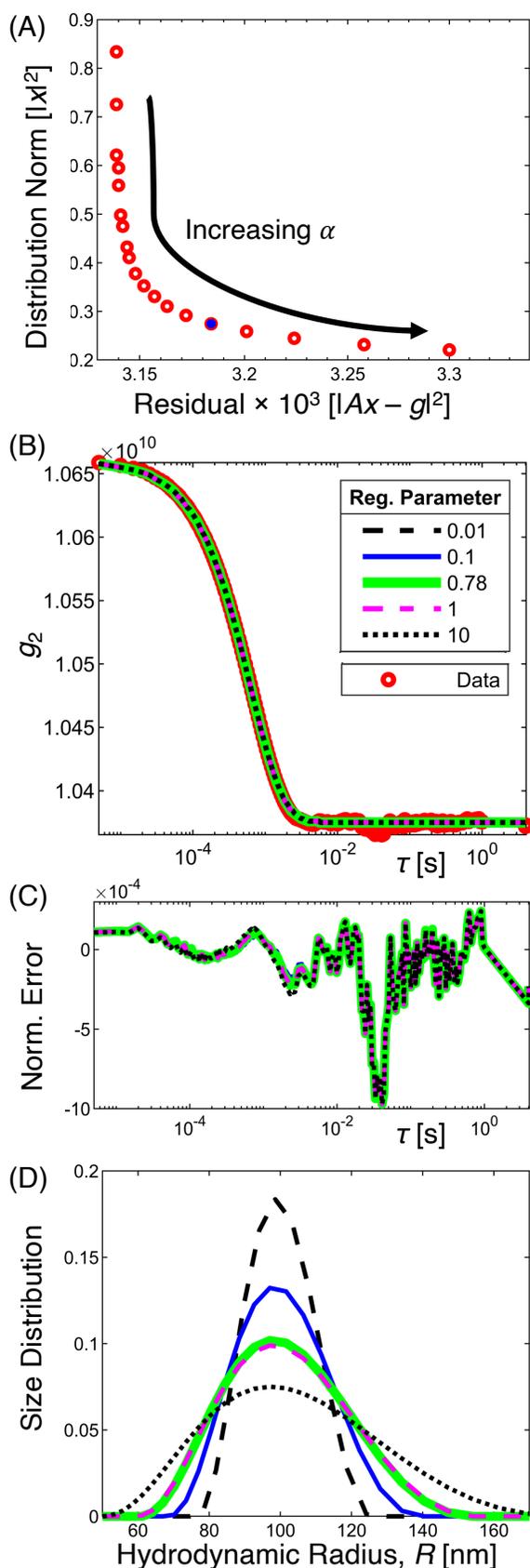


Figure 3. The influence of the regularization parameter on the fits using the CONTIN algorithm to the DLS data from an aqueous dispersion of 100 nm standards. (A) The L -curve showing the variation of the norm of the distribution and the residual with varying α . The optimal regularization parameter $\alpha = 0.78$ corresponds to the filled data point. (B) The experimental autocorrelation data and fits from CONTIN algorithm with varying α . (C) The corresponding normalized error between the data and the fits, and (D) the corresponding particle size distributions generated by the fits.

differences in the normalized error, both algorithms accurately modeled the mean particle size with the peaks occurring at a hydrodynamic radius of about 100 nm (Figure 2C). However, the two algorithms provided different widths of the particle size distributions.

This difference in size distributions can be attributed to how each of these two algorithms approach the problem of calculating a decay rate distribution. The cumulant algorithms attempt to find parameters to fit into a modified Gaussian distribution. For the case of quadratic expansion, the algorithm attempts to fit a Gaussian distribution with a mean and a standard deviation. For the monodispersed sample, the standard deviation was found to be two orders of magnitude smaller than the mean, resulting in an apparent sharp peak. In contrast, the broad distribution predicted by the CONTIN algorithm is the result of both the regularization method and the side constraint. The CONTIN method attempts to fit the distribution of the decay function directly to the data through a transformation matrix using least-squares regression and does not have a predefined equation, allowing the algorithm flexibility in fitting a decay rate distribution while still converging rather swiftly (Figure S1). This results in the CONTIN algorithm providing a fit with smaller errors compared to the cumulant algorithms but can result in an artificially wider size distribution.

The breadth of the size distribution predicted by the CONTIN method is also dependent on the weight of the side constraint dictated by the regularization parameter α ; changing α may result in different size distributions corresponding to the same autocorrelation data set. The larger the regularization parameter, the larger the effect of the side constraint; the CONTIN side constraints favor smaller concavity resulting in wider distributions at higher regularization values. As demonstrated in Figure 3A, as the regularization parameter decreases, so does the error, at the expense of a wider distribution. We note that fits to the autocorrelation data corresponding to different α values, varying across three orders of magnitude, appear to be very similar (Figure 3B), and so do the normalized error values as a function of the delay time τ (Figure 3C). The predicted size distributions, however, become markedly broader with increasing α (Figure 3D).

Thus, finding the optimal regularization parameter requires an L -curve criterion which was developed to simultaneously incorporate the residual norm and the distribution norm. The

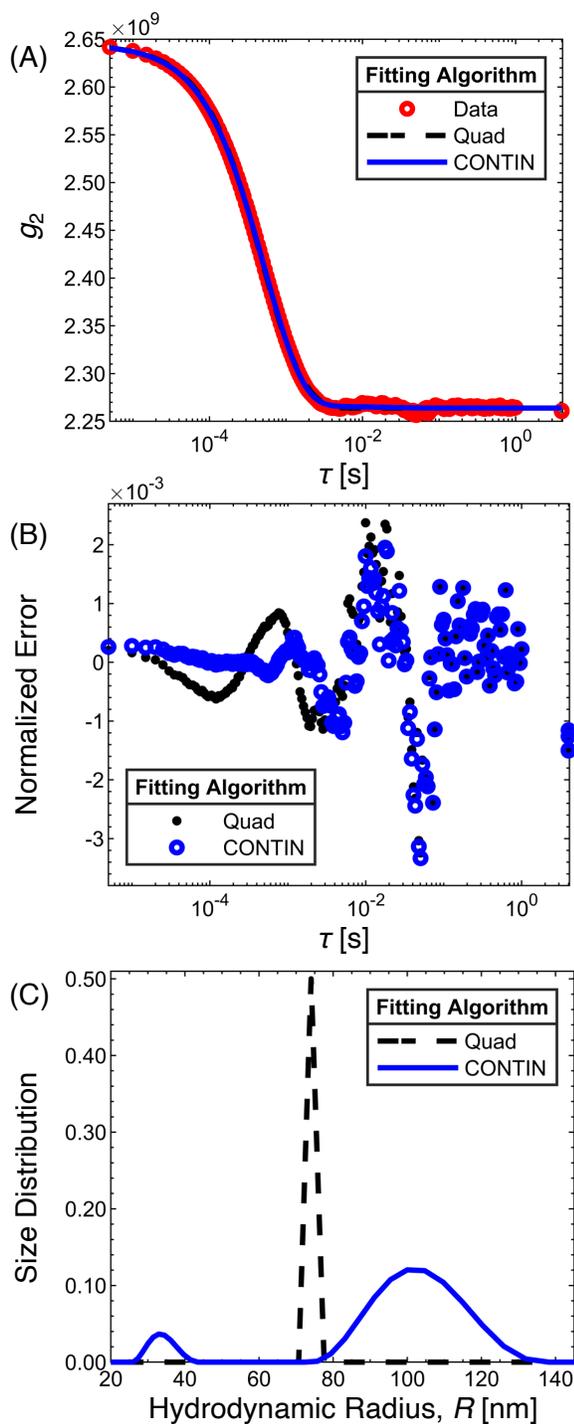


Figure 4. Analysis of DLS data from a bimodal 5:1 aqueous dispersion of 25 nm and 100 nm standards. (A) The experimental correlation data and fits using the quadratic Taylor expansion and CONTIN algorithms ($\alpha = 0.11$). (B) The normalized error of the fits. (C) The size distributions computed by the two algorithms.

norm of the distribution and residual values are calculated for a range of regularization parameters. These points are then scaled to a square plane as described by Castellanos et al.²¹ and two fitting lines were drawn: one through the points with the

lowest residual values and another through those with the lowest distribution norms. The regularization parameter that lies closest to the intersection point of these two lines provides the best combination of normalized error and distribution norm and is deemed to be the optimal regularization parameter. We note that Scotti et al.¹ showed recently that the CONTIN algorithm combined with the L -curve criteria could provide suitable performance to determine size distributions for colloidal dispersions. As **Figure S2** demonstrates, the REPES algorithm demonstrates a similar trade-off as the CONTIN. As such, for the remainder of this paper, the fit results shown correspond to an optimal regularization parameter that was found using the L -curve criterion.

4.2. Particle Dispersions with Polymodal Size Distributions

Both the cumulant and regularization methods work well for systems with monomodal scatterer size distributions, but to be useful for experimental datasets, their performance, and accuracy of prediction should be investigated for systems with polymodal size distributions as well. To this end, autocorrelation data from a 5:1 dispersion (by number concentration of particles) of 25 nm and 100 nm silica standards was analyzed using different methods. The cumulant methods were not expected to perform well for such systems, given the inherent assumption of a monomodal distribution of the decay rate that feeds into them. At the same time, regularization methods that do not make any initial assumptions about the decay rate distributions are expected to perform better in obtaining the expected size distribution.

Interestingly, the fits obtained from the cumulant methods (quadratic cumulant expansion) and the CONTIN regularization method do not appear to be very different (**Figure 4A**). The magnitude of the normalized error was also less than 10^{-3} (**Figure 4B**), although systematic deviations between the data and the fits were observed for the cumulant methods, indicating the inability of the quadratic cumulant method to capture the decay of the g_2 with τ precisely.

An inspection of the predicted size distributions from the two methods (**Figure 4C**) clearly demonstrates the difference in the ability of the cumulant and the regularization methods to describe polymodal solutions/dispersions. While the CONTIN algorithm predicted a bimodal size distribution with two populations, each centered around 30 nm and 100 nm radii, respectively. The quadratic cumulant expansion method, contrastingly, predicted a monomodal size distribution with a ~ 75 nm mean radius.

These differences can also be explained by the algorithms' approach to fitting the autocorrelation data. The Gaussian distribution that the cumulant expansion algorithm is attempting to fit inherently assumes there is only one average particle size. As such, the algorithm tries to find parameters that will best fit the data but will have no physical significance. In contrast, the CONTIN algorithm was able to capture the bimodal distribution with the two peaks being close to the

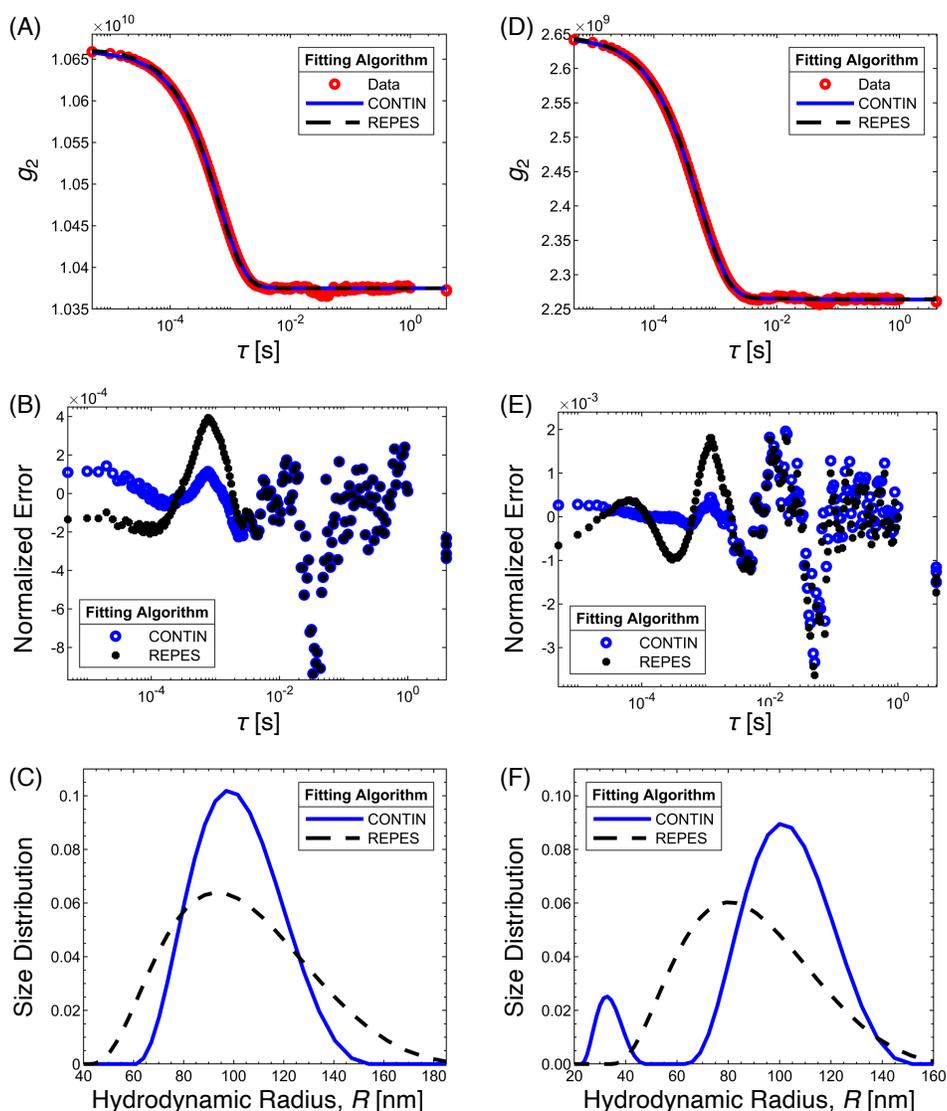


Figure 5. Comparing the performance of CONTIN and REPES algorithms (A, D) The experimental correlation data and fits using the CONTIN and REPES algorithms for (A) an aqueous dispersion of 100 nm standards ($\alpha = 0.78$ for CONTIN and 14 for REPES) and (D) bimodal 5:1 aqueous dispersion of 25 nm and 100 nm standards ($\alpha = 0.11$ for CONTIN and 14 for REPES). (B, E) The corresponding normalized errors from the fits. (C, F) The corresponding size distributions computed by the two algorithms.

average size of the individual particles because it possesses the flexibility of fitting an arbitrary decay rate distribution instead of a predefined function to the autocorrelation data.

This case study also illustrates the importance of choosing the appropriate β value, especially for systems with significant scattering contributions in the low- τ range. A suboptimal β value can result in slight deviations between the autocorrelation data and the fit, resulting in size distribution predictions with systematic offsets. An example of such a case is presented in **Figure S3**, where the same data as shown in Figure 4 was analyzed but with a slightly lower than optimal β value, leading to an overestimation of the size of the smaller particle population.

4.3. Comparing Regularization Methods

The GUI we developed also enabled a comparison of the different regularization methods. As noted in Section 2, NNLS is the most primitive regularization method with no side constraints, while the DYNALS algorithm is a special case of the REPES algorithm with the regularization parameter α set to 0.5. Therefore, we chose to restrict the comparison among the regularization methods in their ability to describe model dispersions with monomodal and bimodal size distributions to the CONTIN and the REPES algorithms only, realizing that NNLS and DYNALS algorithms will not perform better than either of the chosen algorithms.

Considering first the dispersion with a monomodal size distribution, the autocorrelation data are satisfactorily described by both CONTIN and REPES algorithms (**Figure**

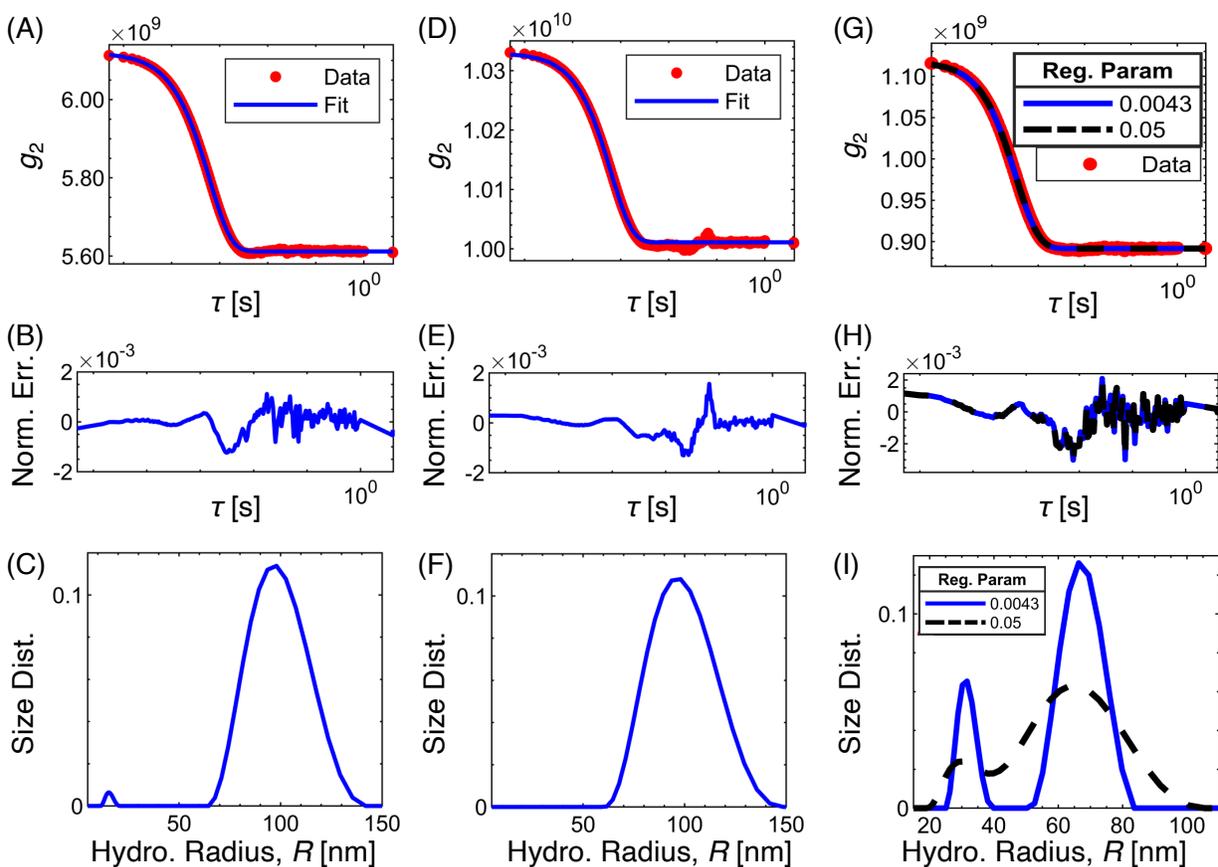


Figure 6. Testing the performance of CONTIN algorithm (A, D, G) The experimental correlation data and fits using the CONTIN algorithm for (A) a bimodal 1:1 aqueous dispersion of 25 nm and 100 nm standards ($\alpha = 0.3$), (D) a bimodal 1:5 aqueous dispersion of 25 nm and 100 nm standards ($\alpha = 0.78$), and (G) a bimodal 5:1 aqueous dispersion of 25 nm and 65 nm standards. (B, E, H) The corresponding normalized errors from the respective fits. (C, F, I) The corresponding size distributions. In (G-I), fits, errors, and size distributions to the optimal alpha ($\alpha=0.043$) (solid blue) and a typical alpha ($\alpha=0.5$) (dashed black) are shown.

5A). However, systemic errors begin to appear in the fits obtained by the REPES algorithm, although they are still smaller than the errors produced in the large τ range (Figure 5B). The differences between the size distribution predictions become apparent in Figure 5C, where REPES was found to predict a slightly smaller mean size and broader size distribution as compared to the CONTIN algorithm.

These differences became even more significant for dispersions with bimodal size distributions, wherein it became clear that not all the regularization methods predict a bimodal distribution. Even though the fits to the autocorrelation data appeared similar (Figure 5D), the normalized error clearly shows the inability of the REPES algorithm to fit the autocorrelation data for $\tau \leq 10^{-2}$ s (Figure 5E). Correspondingly, the REPES algorithm was unable to detect the bimodal size distribution and predicted a monomodal size distribution instead, with a mean size similar to that predicted by the quadratic cumulant method (Figure 4B).

The disparity between the size distributions obtained using the CONTIN and REPES regularization methods can be attributed to the REPES algorithm employing a minimization function

that utilizes the experimental g_1 (instead of CONTIN employing g_2) and not normalizing the sum of the distribution to one in the minimization function but instead normalizing the distribution after the minimization. The REPES algorithm (and the DYNALS algorithm) is more effective in handling noise in the data compared to the CONTIN algorithm¹⁶. As such, peaks with low intensity will be treated as noise and will not appear in the final distribution. In the case of the bimodal sample analyzed, the REPES algorithm treats the weak intensity 25 nm peak as noise and generates a monodispersed fit similar to what the cumulants expansion methods would provide. In contrast, the CONTIN algorithm is less aggressive in terms of noise cancellation; therefore, it can pick up the smaller intensity peaks at the expense of higher noise in the predicted size distributions and yield, in this case correctly, a bimodal distribution.

4.4. Pushing the Limits of the DLS Technique and the CONTIN Algorithm

A limitation of the DLS method is the disproportionate contribution of larger particles towards the scattering intensity, masking the scattering contributions of smaller particles. The

Rayleigh scattering intensity of the scattered light is proportional to the sixth power of the radius of the scatterers; thus, larger particles will dominate the scattering intensity.¹¹ We note that the scattering intensity is also directly proportional to the number density of the scatterers. Thus, scattering signals from smaller particles can be increased by increasing their concentration in the sample if only the tentative estimates of the scatterer sizes are known *a priori*.

Earlier, we demonstrated that CONTIN could detect small particles in a bidisperse mixture (size ratio 4:1, concentration ratio 5:1) when the smaller particles were present in significant excess. Here, we push the algorithm further to predict size distributions in mixtures with smaller relative concentrations of the smaller particles. In **Figures 6A-F**, we demonstrate the results from such fits obtained from our GUI for two bidisperse mixtures (size ratio 4:1) with concentration ratios of 1:1 (**Figures 6A-C**) and 1:5 (**Figures 6D-F**), respectively. Clearly, good fits to the autocorrelation data were produced (**Figures 6A and 6D**, respectively), and the errors produced were small (mostly less than 10^{-3} , **Figures 6B and 6E**). Indeed, the CONTIN algorithm was able to predict a bimodal size distribution for the 1:1 dispersion of 25 nm and 100 nm particles (**Figure 6C**), although the intensity of the 25 nm peak is much smaller here as compared to the 5:1 dispersion (**Figure 5F**), appeared at a smaller size than expected, and could easily be mistaken for noise. However, as shown in **Figure 6F**, the CONTIN algorithm failed to predict the bimodal distribution for the 1:5 dispersion of 25 and 100 nm particles. The algorithm instead predicted a monomodal distribution with a peak corresponding to the larger particle size. This, again, can be attributed to the small contribution to the scattering intensity from the smaller particles.

Another limitation of the CONTIN algorithm can be its inability to differentiate between particles of similar-size populations. Having demonstrated that, by choosing the optimal regularization parameter α , the CONTIN algorithm can enable the detection of bidisperse mixtures with a size ratio of 1:4, we endeavored to push the limits of the algorithm by testing its ability to differentiate particle populations with a size ratio of 5:12 by employing a 5:1 dispersion of 25 and 65 nm standards. As is evident in **Figures 6G-I**, the algorithm was indeed able to capture the bidisperse size distribution when working with the optimal α , providing satisfactory fits (**Figure 6G**) with minimal errors (**Figure 6H**) and predicting two populations centered around 30 nm and 65 nm, respectively (**Figure 6I**). In contrast, when the algorithm operates at a non-optimal α , it is unable to capture the size distribution correctly. This difference in the algorithm performance highlights the importance of combining prior knowledge of the samples with the use of optimal α values, especially for samples with multiple scatterer populations with similar sizes.

5. Conclusion

In this work, a GUI and the underlying numerical calculation engine were developed to provide users the ability to

implement, compare, and evaluate algorithms (including cumulant methods and regularization methods such as NNLS, CONTIN, REPES, and DYNALS) available for analyzing time autocorrelation data obtained from DLS experiments. The cumulant and regularization methods (CONTIN and REPES) were both shown to accurately describe the average particle size of colloidal dispersions containing monomodal size distributions. For dispersions with bimodal size distributions, however, the cumulant method was shown to be ineffective, while the CONTIN algorithm performed satisfactorily in capturing the size distribution. The CONTIN algorithm was further tested in bimodal dispersions with a size ratio of 1:4 and concentration ratios of 5:1, 1:1, and 1:5. While the algorithm performed well in the first two cases, it failed to capture the bimodal size distribution for the 1:5 concentration dispersion, which can be attributed to the significantly small contribution of the smaller particles towards the overall scattering from the sample. Lastly, we pushed the CONTIN algorithm to detect bimodal size distribution in a 5:1 concentration dispersion of particles with a size ratio of 1:2.4. The CONTIN algorithm, supported by the optimal regularization parameter estimated using the *L*-curve criterion, was able to capture the bimodal distribution in this case as well. Our analysis highlights that, when coupled with the appropriate data analysis algorithms, DLS can serve as a versatile and robust analytical technique in diverse applications. However, appropriate care must be exercised in choosing the parameters, especially for the regularization approaches, to obtain physically relevant results. The GUI we have developed here provides a powerful software tool for comparing the analysis from multiple autocorrelation data fitting algorithms and aids in the selection of the appropriate parameters to constrain and support the data fitting procedure, enabling real-time data analysis, and supporting the optimization of experimental protocols.

Conflicts of interest

There are no conflicts to declare.

Funding Sources

This research was supported by National Science Foundation under Grant No. DMR-2048285.

Acknowledgments

We acknowledge helpful discussions with Prof. Vivek Sharma and Dr. Jeffrey Ting.

References

- 1 A. Scotti, W. Liu, J. S. Hyatt, E. S. Herman, H. S. Choi, J. W. Kim, L. A. Lyon, U. Gasser and A. Fernandez-Nieves, *J Chem Phys*, 2015, **142**, 234905.
- 2 A. G. Mailer, P. S. Clegg and P. N. Pusey, *Journal of Physics: Condensed Matter*, 2015, **27**, 145102.
- 3 O. Glatter, in *Scattering Methods and their Application in Colloid and Interface Science*, ed. O. Glatter, Elsevier, 2018, pp. 223–263.

- 4 P. Rao, Z. Yu, H. Han, Y. Xu and L. Ke, in *Evaluation Technologies for Food Quality*, Elsevier, 2019, pp. 535–557.
- 5 R. Pecora, *Journal of Nanoparticle Research*, 2000, **2**, 123–131.
- 6 R. Pecora, *Static and Dynamic Light Scattering in Medicine and Biology*, 1993, **1884**, 2.
- 7 T. Kourti, J. F. MacGregor, A. E. Hamielec, D. F. Nicoli and V. B. Elings, *Polymeric Materials Science and Engineering, Proceedings of the ACS Division of Polymeric Materials Science and Engineering*, 1988, **59**, 160–164.
- 8 B. J. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, New York, NY, 1976.
- 9 J. Yang and K. Mattison, *China Particuology*, 2003, **1**, 219–222.
- 10 J. Stetefeld, S. A. McKenna and T. R. Patel, *Biophys Rev*, 2016, **8**, 409–427.
- 11 S. Falke and C. Betzel, in *Bioanalysis*, eds. A. Pereira, P. Tavares and P. Limão-Vieira, Springer, 2019, vol. 8, pp. 173–193.
- 12 S. Bhattacharjee, *Journal of Controlled Release*, 2016, **235**, 337–351.
- 13 O. Glatter, in *Scattering Methods and their Application in Colloid and Interface Science*, Elsevier, 2018, pp. 265–297.
- 14 J. R. Vega, L. M. Gugliotta, V. D. G. Gonzalez and G. R. Meira, *J Colloid Interface Sci*, 2003, **261**, 74–81.
- 15 D. E. Koppel, *J Chem Phys*, 1972, **57**, 4814–4820.
- 16 J. Jakeš, *Collect Czechoslov Chem Commun*, 1995, **60**, 1781–1797.
- 17 S. Provencher, *Comput Phys Commun*, 1982, **27**, 213–227.
- 18 I. D. Morrison, E. F. Grabowski and C. A. Herb, *Langmuir*, 1985, **1**, 496–501.
- 19 S. W. Provencher, *Comput Phys Commun*, 1982, **27**, 229–242.
- 20 S. W. Provencher, *Die Makromolekulare Chemie*, 1979, **180**, 201–209.
- 21 J. L. Castellanos, S. Gómez and V. Guerra, *Applied Numerical Mathematics*, 2002, **43**, 359–373.
- 22 M. Salazar, H. Srivastav and S. Srivastava, *Scattering Data Analysis Software*, <https://sites.google.com/view/srivastava-lab/software>, (accessed 29 January 2023).
- 23 S. Srivastava, DLS GUI, <https://osf.io/6b3zf/>, (accessed 23 May 2023). DOI: 10.17605/OSF.IO/6B3ZF
- 24 E. Jondeau and M. Rockinger, *J Econ Dyn Control*, 2001, **25**, 1457–1483.
- 25 S. Seabold and J. Perktold, *Proceedings of the 9th Python in Science Conference*.