HHS Public Access

Author manuscript

Eur Biophys J. Author manuscript; available in PMC 2024 July 01.

Published in final edited form as:

Eur Biophys J. 2023 July ; 52(4-5): 203–213. doi:10.1007/s00249-023-01631-6.

Systematic noise removal from analytical ultracentrifugation data with UltraScan

Saeed Mortezazadeh[1], **Borries Demeler**[1,2]

[1]Department of Chemistry and Biochemistry, University of Lethbridge, Lethbridge, AB T1K3M4, Canada

[2]Department of Chemistry and Biochemistry, University of Montana, Missoula, MT 59812, United States

Abstract

A method for removing time- and radially invariant noise from sedimentation velocity and sedimentation equilibrium experiments performed in an analytical ultracentrifuge is presented. The method averages repeat radial incident light measurements as a function of the photomultiplier response at different wavelengths to remove the majority of the time-invariant noise contributions from intensity data measurements. The results of this method are compared to traditional absorbance data generated with a buffer reference and the Beckman Optima AUC data acquisition program, and with the standard UltraScan refinement workflow. The method avoids the amplification of stochastic noise inherent in the absorbance scan subtraction traditionally employed in sedimentation velocity and equilibrium data. In addition, the collection of intensity data frees up the reference channel for additional samples, doubling the capacity of the instrument. In comparison to absorbance data, the residual mean square deviation of a fitted sedimentation velocity experiment without additional noise correction by UltraScan was improved by a factor of 4.5 when using the new method. This improvement benefits sedimentation equilibrium experiments as well as analytical buoyant density equilibrium experiments where routine time-invariant noise correction calculations cannot be performed.

Keywords

Analytical ultracentrifugation; Noise processing for equilibrium data; Analytical buoyant density equilibrium centrifugation; UltraScan

Introduction

Analytical ultracentrifugation (AUC) is a versatile technique for analyzing a broad range of macromolecules in the solution phase (Hansen et al. 1994; Lebowitz et al. 2002). AUC sedimentation velocity experiments (SVEs) measure the sedimentation and diffusion coefficients of solutes, and their partial concentrations in a mixture. These parameters provide information about the composition of mixtures, the molar mass, and the shape

✉Borries Demeler, demeler@gmail.com.

anisotropy of the solutes in the mixture. The radial and temporal evolution of the concentration profile in an AUC cell is accurately modeled by finite-element solutions of a second-order partial differential equation known as the Lamm equation (Lamm 1929). Experimental data are typically convoluted with three types of noise originating from the optical system. It is essential that all noise contributions are minimized, and if possible, removed, to increase the signal-to-noise ratio (SNR). When using UV–visible absorption optics, the absorbance profile, A, determined in an AUC experiment as a function of time and radius, is proportional to the concentration, and can be calculated for each radial point i and time point j from the incidence intensity, I_0, and the transmitted intensity, I, as shown in Eq. 1

$$A_{i,j} = \log_{10}\left(\frac{I_o}{I}\right)_{i,j}. \quad (1)$$

Experimental noise can be categorized as systematic and stochastic. Systematic noise can sometimes be determined or fitted, and then removed completely. Stochastic noise is random, and cannot be modeled; therefore, stochastic noise sources should be eliminated whenever possible. Typically, two types of systematic noise are observed in an AUC experiment: the first type is time-invariant noise (TIN), which differs for each radial point but has the same offset for each scan throughout time. TIN originates from absorbance variations in the cell windows, for example, from dirt or scratches, from photomultiplier tube (PMT) surface response variations as a function of radius and wavelength, and from dirt on other optical system components (lenses, mirrors). TIN can be thought of as a fingerprint on the lens of a camera. Each image has the same fingerprint flaw, regardless of the time when the image was taken. The second type is radially invariant noise (RIN), which fluctuates with time but is characterized by a constant baseline offset along the radial dimension of the cell for a given scan. Stochastic noise, which ideally has a Gaussian distribution, is independent of time and radius. The observed intensity profile, A_{obs}, can therefore be represented as a sum of the contributions from the absorbance of the molecule(s) (A_{mol}), the TIN contributions from the windows ($A_{TIN,\ windows}$), the PMT, and TIN from the optical system ($A_{TIN,\ optics}$), RIN (A_{RIN}), and stochastic noise (A_{stoch})

$$A_{obs} = A_{mol} + A_{TIN,windows} + A_{TIN,optics} + A_{RIN} + A_{stoch}. \quad (2)$$

RIN and TIN can be readily fitted and eliminated for data from SVEs during data modeling using a robust technique described earlier (Schuck and Demeler 1999), while stochastic noise cannot be removed, and contributes to the residuals of the fit.

Traditionally, in an AUC experiment, the absorbance profile is generated by applying Eq. 1 using the intensity of a buffer scan (I_0) and the intensity of the sample scan (I). This approach also has the benefit of eliminating any time-invariant noise contributions originating in the optical system ($A_{TIN,\ optics}$) that are present in both sample and buffer scan data. It does not, however, remove the TIN contributions originating from the windows ($A_{TIN,\ windows}$), which are different for the sample and reference channel in the AUC. Importantly, the subtraction of a buffer scan from the sample scan decreases the precision of the data by amplifying stochastic noise by a factor of square root of two. In addition,

a second channel is required to hold the buffer, decreasing the rotor capacity by a factor of two. To avoid magnification of stochastic noise, intensity data can be used instead to eliminate the stochastic noise amplification (Demeler 2010), and to free up one channel for another sample. For the conversion from intensity data to absorbance data from SVEs, the reference incidence intensity is typically obtained from the air region at a point above the meniscus, where sample absorbance is absent, but since this value is a constant, TIN, which varies with radius, is not removed by this procedure. To remove TIN numerically from experimental data, the underlying signal from the molecular transport must be time-variant, and fitted simultaneously to solutions of the Lamm equation (Schuck and Demeler 1999; Cao and Demeler 2005, 2008; Brookes et al. 2010). This method, however, cannot be used in sedimentation equilibrium experiments (SEEs) where the concentration flux no longer varies once the system reaches equilibrium, and the approach to equilibrium is not modeled. At equilibrium, the concentration profile of the molecules of interest becomes time-invariant, and can no longer be numerically distinguished from the TIN contained in the absorbance profile (Eq. 2). If TIN is not removed from the experimental data, the fitting results may be compromised and become unreliable. RIN contributions to SEE data can be treated through baseline fitting of SEE data, and are only relevant for time-dependent SVEs. However, stochastic contributions must also be minimized to maximize the observed SNR. These fundamental limitations require the development of novel approaches to address noise processing for equilibrium data. Here, we propose a new method for addressing TIN removal from SEEs by determining the major contribution to TIN, the systematic noise contributions originating in the optical system of the AUC absorbance optics, which represent the same TIN components ($A_{TIN,\ optics}$) that are removed in the traditional absorbance experiments. Our approach determines the time-invariant noise from the optical system by acquiring intensity scans from each wavelength. Intensity data are acquired repeatedly to average intensity recordings over a large number of observations. The stochastic noise contributions from each scan are effectively eliminated by averaging a large number of scans measured under identical conditions. The averaged intensity profile faithfully reflects the time-invariant noise contributions from the optical system, and can be used for the calculation of absorbance profiles using Eq. 1. Open source software was written to extend the UltraScan suite (Demeler and Gorbet 2016; Demeler 2005) with a noise processing module primarily intended for SEEs, in particular for analytical cesium chloride buoyant density gradient equilibrium (ABDE) experiments designed to quantify the ratio of empty and full adeno-associated virus capsids. For SVEs, we also developed an RIN correction module, discussed below.

Methods

To generate instrument-specific TIN correction profiles, an AN60Ti rotor was loaded with a single cell in hole 2, and a counterbalance was loaded into hole 4. Holes 1 and 3 were left empty, and data acquisition profiles were programmed into UltraScan to scan 54 repetitions of each empty hole, resulting in 54 scans/channel × 2 channels/ hole × 2 holes = 216 scans for each wavelength. 100 wavelengths were measured in each run, for a total of 21,600 scans/run. Seven runs were required to collect data for 611 wavelengths, starting from 190 nm and ending at 800 nm. Scans were collected over a radial range

from 5.75 cm to 7.25 cm with a radial resolution of 0.001 cm (10 μm). All data were collected in a Beckman Coulter Optima AUC at the Canadian Center for Hydrodynamics. The rotor speed was chosen such that the scanning interval is minimal. On our instruments, a 14,600 rpm rotor speed provides optimal synchronization between the lamp flash rate and the rotor hole position, yielding intervals as short as 8 s for a complete scan of a single channel. Using the pseudo-absorbance module in UltraScan, intensity scans from identical wavelengths were then aligned and averaged, and used to noise correct experimental data. To generate sedimentation velocity data for testing, a 0.46 ml bovine serum albumin sample was measured at 0.6 AU and at 40,000 rpm at 280 nm and 20 °C in a standard double sector 1.2 cm epon centerpiece (Beckman Coulter, Indianapolis) on an Optima AUC (Beckman Coulter, Indianapolis) at the Canadian Center for Hydrodynamics at the University of Lethbridge. The protein was buffered in Dulbecco phosphate-buffered saline (DPBS), which was also loaded into the reference sector. Absorbance data were obtained by requesting absorbance data from the Optima AUC web application (Beckman Coulter, Indianapolis).

Results and discussion

TIN correction

Figure 1A shows a sample intensity profile of an empty rotor hole at 190 nm, displaying two distinct groups of scans that were measured at 190 nm in a single experiment. Figure 1B displays the residual errors of each radial point of all scans. The residual errors were calculated in two different ways. The upper panel of Fig. 1B shows the difference between the radial intensity of each data point and the intensity of each radial position when averaged over all scans. The average scan is depicted in the black line of Fig. 1A. In the lower panel of Fig. 1B, the residual error obtained after aligning all scans to an arbitrary value is shown. This value is calculated by finding the mean intensity of each scan and subtracting it from each radial intensity value. As can be seen, either set of residuals displays non-random, systematic variations. In addition to intensity variations within each group, a notable jump in light intensity can be observed between the red and blue groups of scans, which is too large to be explained by random, radially invariant noise variations. We visually observed these jumps for most wavelengths measured on our instrument, especially in the lower UV range, and speculated that this jump originates from an adjacent wavelength, caused by a monochromator failure to correctly position the diffraction grating. The Xenon flash lamp installed in the Optima AUC has a strongly variable emission spectrum, especially in the UV range (see Fig. 1D), leading to large intensity jumps for adjacent wavelengths. In addition, the residual errors (Fig. 1B and Fig. 1C) display a different pattern for the two groups, also suggesting that they were the result from measurements at different wavelengths, reflecting variations in sensitivity for different wavelengths across the PMT surface. This difference in scan shapes is also apparent from the residual patterns in the lower panel of Fig. 1B.

The significant difference in the intensity and shape of the scans indicates that the instrument's wavelength resolution is not precise enough to reproducibly scan at the preset wavelength, and that even adjacent wavelengths produce a different response in the PMT, even after adjusting for differences in emission intensity. This problem is most pronounced in the lower UV range, where intensity fluctuations between adjacent wavelengths are

largest (see Fig. 1D). Representative examples of these deviations in the UV (282 nm) and visible range (544 nm) are shown in Fig. 2. Even a relatively small change in the intensity causes unacceptable deviations in the absorbance, adding noise to the data. Therefore, it is important to separate scans that originate from different wavelengths, and to determine which scan belongs to which wavelength. To achieve a uniform intensity profile, each scan must be assessed in terms of absolute light intensity and intensity profile shape. In the first step, an average intensity is determined (Eq. 3) from all scans belonging to each instrument-assigned wavelength and calculating the residuals by subtracting the average intensity, $\bar{I}$, from all points R_i of each individual scan (Fig. 2). The pattern of the residuals can be visually inspected, and clearly identifies different shapes when there are more than one group resulting from a different wavelength. The root-mean-square deviation (RMSD) is then calculated (Eq. 4). This parameter is unique for each wavelength, reflecting the intrinsic noise of the PMT recordings at a particular wavelength. If the RMSD is too high, it either indicates that the intensity of the incident light has degraded because of some defect in the optical system, or that the averaged group of scans stems from multiple wavelengths, especially when the shape of the residuals shows systematic deviations and non-random patterns

$$\bar{I} = \frac{1}{N}\sum_{n=1}^{N} I(R_i) \tag{3}$$

$$RMSD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(I(R_i) - \bar{I}\right)^2}. \tag{4}$$

Wavelength assignment

As indicated above, inaccuracies in the wavelength settings have to be considered when clustering intensity scans for a downstream TIN reference vector. Inaccurate wavelength settings are a common issue on the older Proteomelab XLA instruments, where changing the wavelengths mid-run typically let to variations of ± 3 nm, but in the newer Optima AUC instruments, a 0.5 nm wavelength accuracy is advertised by the vendor, although on some instruments, we have observed larger discrepancies. Regardless, it is necessary to account for wavelength inaccuracies in our algorithm. To address this issue, a clustering algorithm was developed to determine which wavelength a particular set of intensities belongs to. Figure 3A shows two-dimensional plots of the scan intensities as a function of their RMSD values, as well as the corresponding histograms for the first dataset we collected (190 nm). Similar observations were made for most other wavelengths. In this view, three unique groups appear to be present when inspecting the intensity histogram, while the RMSD histogram indicates only two scan groups. Numerous factors can affect the transient state of the optical system, causing intensity fluctuations over time. However, the shape of the intensity average is more important for the determination of the TIN pattern than the intensity amplitude. As a result, clustering of scans into groups of wavelengths by taking into account closely spaced intensity distributions alone can fail to correctly classify which scan belongs to which wavelength. Figure 3B, C shows the probability distribution

function (PDF) of the scans' RMSD and intensity values at wavelength 190 nm and 191 nm. Their superpositions are shown in Fig. 3D. These PDF profiles show similarities and overlaps between neighboring scan groups collected at different wavelength, both in terms of scan shape and intensity. This suggests that it is reasonable to cluster the wavelengths from each scan correctly by inspecting the PDF profiles derived from both the intensity and the averaged scan residuals.

Clustering is an important topic in data science which seeks to divide samples into natural groups based on shared features. Several mathematical algorithms have been developed to deal with the similarities between samples in a data-set. Among these approaches, one-dimensional clustering relies on a probability distribution function and produces robust outcomes. A common way to estimate the probability distribution function is to normalize the histogram of the data. If there are insufficient samples, the resulting histogram will be irregular, regardless of the number of bins. An improved approach uses the kernel density estimation (KDE) method, which is more reliable than the histogram approach for one-dimensional clustering. KDE is a non-parametric statistical method that does not need any parameters pertaining to the underlying distribution, and is able to extract any complicated distribution from a dataset. This approach is based on considering each sample as a smooth function (called the kernel) where the final distribution is obtained by summing all individual functions. Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed samples drawn from an unknown distribution. The PDF is given by Eq. 5

$$\hat{p}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),\ K(x) = e^{\frac{-(x-x_i)^2}{2h^2}}, \tag{5}$$

where K is the kernel function and h is the smoothing bandwidth where all kernels have the same h value. There are several options for choosing the kernel function; the most common is the Gaussian function. Because the resulting estimate is summed over all kernels, the KDE forms a peak around regions of signal density. In contrast, in regions with few observations, only a small number of kernels contribute to the density estimate. For all cases, the bandwidth h selection is the main challenge in KDE. If h is too small, the distribution function will be noisy, and if h is too large, some peaks will be flattened. The correct bandwidth value results in a smooth distribution where all peaks are clearly discernible. The computationally simplest method for bandwidth estimation is the rule of thumb estimator, and can be used for parametric Gaussian functions with $h = 1.06\,\sigma\,n^{-1/5}$, where σ is the standard deviation and n is the number of samples. Silverman's rule of thumb (Silverman 2018) is a modification of this method where the bandwidth is given by $0.9\,Min\,(\sigma, IQR/1.34)n^{-1/5}$, and IQR is the interquartile range. Although this adjustment improves estimation accuracy, this method works better with unimodal distributions. Other methods, such as least-squares cross-validation and plug-in approaches, are a natural choice for bandwidth estimation of unknown multimodal distribution functions. We chose the Sheather–Jones plug-in method for this investigation because of its favorable computing performance compared to the cross-validation approach. Papers (Chen 2017; Sheather 2004) provide good reviews on most of the approaches used in the bandwidth selection procedure. All PDF profiles shown in Fig. 3 exhibit smoother curves than the corresponding histograms.

Hence, the derivative of Eq. 5 can locate the extremes. The minima of the PDF profile define the limits of each cluster scope and the maxima define the peak position. Having specified the cluster regions, two simple assumptions were made to relabel the wavelengths for their corresponding scans: The cluster with the highest PDF value is considered the main cluster. The first assumption is that the main cluster scans belong to the correct wavelength prior to relabeling. The second assumption looks at overlaps between the main cluster of a wavelength with closely spaced non-main clusters of neighboring wavelengths, and assigns them to the main cluster wavelength. Figure 3 illustrates this procedure where there are two clusters at both wavelength 190 nm and 191 nm before relabeling (see Fig. 3B, C). As mentioned earlier, the RMSD profile is used as the primary discriminator for clustering, even when the number of peaks in the RMSD profile varies from the number of peaks in the intensity profile. The first peak at 190 nm and the second peak at 191 nm are considered the main clusters for the respective wavelengths because of their amplitudes. The scans of the non-main cluster at 191 nm were relabeled 190 nm, since their intensities overlap with the main cluster at 190 nm. The intensity profile is used to ensure that transmitted scans are similar to neighboring wavelength scans (see Fig. 3D lower plot). Figure 4 displays the residual errors of calculating the reference scans for several wavelengths before and after the clustering process. As is evident from the left panels in Fig. 4, the residual patterns are far from randomly distributed before clustering, and clearly show both non-random patterns and intensity jumps, but are perfectly random after clustering. The extent of residual asymmetry was greater at shorter wavelengths, because the fluctuation of light intensity in the shorter wavelengths is greater than at longer wavelengths (see Figs. 1D and 2). After clustering, the residual errors were decreased by a factor of 5.1, 3.1, and 2.0 for 190 nm, 282 nm, and 544 nm, respectively. Furthermore, the standard deviation of the residual errors was decreased by the factor of 14.6, 10.1, and 2.7 (Fig. 4). The approach described here produced the desired results for all wavelengths between 190–800 nm.

RIN correction

AUC measurements collected with the UV/visible optical system also include radially invariant noise (RIN). This noise is reflected by a constant offset of the intensity across all radial points in each scan, where this constant offset can vary randomly from scan to scan. RIN primarily originates from fluctuations in the incident light intensity recorded by the optical system. These fluctuations are slow enough that they are not detected during the time it takes to record a scan, but over a longer time period they can become apparent for a series of scans. For SVEs, this offset value can either be directly fitted (Brookes et al. 2010), or by modeling the integral of the signal over time for a sedimenting sample. When integrating the observed scans as a function of time for a radial subsection of the column, a smoothly changing integral profile will be observed, reflecting the solute flux in the cell. This can be modeled faithfully by a polynomial function of degree n, where n depends on the number of sedimenting solutes, their sedimentation and diffusion speeds, and the radial data limits. Regardless of the sedimentation and diffusion coefficients of the sedimenting solutes, during the sedimentation transport, the boundary of each solute moves toward the bottom of the cell, radially dilutes, and eventually pellets at the bottom of the cell. This process will cause a slowly decreasing integral function with each successive scan over the radial subsection (see Fig. 5). The degree n of the polynomial function is chosen such that the RMSD of the

integral data is minimized. This typically occurs with values of n between 5 and 10. Once a suitable polynomial function is found, the residuals of the integral data represent the radially invariant noise offsets, and the residual value can be readily subtracted from each radial point for a given scan. This approach is effective in removing radial-invariant noise from any experimental dataset.

TIN/RIN correction performance comparison

To assess the performance of our method with alternative noise processing algorithms, we measured a protein sample with a buffer reference in the reference channel in sedimentation velocity mode at multiple wavelengths. The resulting data were then analyzed with UltraScan by finite-element fitting using five different approaches: (1) Using the standard refinement workflow for UltraScan published earlier (Demeler 2010), where RIN and TIN contributions present in intensity data are directly fitted using the ASTFEM finite-element method (Cao & Demeler 2005, Brookes et al. 2010). (2) Using absorbance data retrieved from the Beckman web application for the Optima AUC, which generates absorbance data by subtracting reference cell intensities from the sample intensities, without additional noise processing by UltraScan. (3) Same as (2), but using the standard refinement workflow used in (1), including TIN and RIN correction. (4) Using intensity data and our new method, which provides RIN and TIN correction using the procedure described in this work, without additional time and radial-invariant noise correction by UltraScan. (5) Same as (4), but using the standard refinement workflow used in (1), including TIN and RIN correction. The performance of each approach was evaluated by comparing the RMSD obtained in each fit. The RMSD is a metric of the goodness of the fit, and a lower value corresponds to better performance. The results are summarized in Table 1, and visually compared in Fig. 6.

From these measurements, we can conclude that the method presented in this work is far superior to the traditional absorbance/buffer subtraction method, and closely rivals the UltraScan noise correction that fits all systematic noise components directly. When relying solely on buffer subtraction in the traditional absorbance mode, residual TIN resulting from defects or dirt on the cell windows is not removed until fitted, and RIN is poorly corrected, since intensities from buffer and sample channel are performed at different times and therefore encounter fluctuations in the radial offsets. Importantly, the traditional absorbance method convolutes two stochastic vectors, which amplifies stochastic noise by a factor of $\sqrt{2}$, which cannot be avoided, even when fitting TIN and RIN with UltraScan. A dramatic improvement in RMSD is observed already without TIN and RIN fitting with UltraScan, if the RIN correction described in this work is applied, and the TIN is removed with the averaged intensity profile described in this work. As in the absorbance data produced by the Beckman software, TIN originating from the cell windows is not removed until it is fitted with UltraScan, but even without noise fitting in UltraScan, a > 4.5-fold improvement in the RMSD is observed compared to standard absorbance data (Table 1). When RIN and TIN are fitted, the RMSD of these data are within 3% of the optimal data processing using UltraScan's workflow when using raw intensity data. The RMSD for Beckman absorbance data, even after fitting RIN and TIN with UltraScan, is 34% worse, reflective of the $\sqrt{2}$ factor in the stochastic noise. As can be seen in Fig. 6, all UltraScan-fitted residual

traces demonstrate random residual distribution, indicating complete TIN removal only after processing with UltraScan's standard workflow.

These results demonstrate that absorbance data generated on the Optima AUC are inferior compared to intensity data, but removal of PMT-derived TIN by subtracting intensity-averaged reference scans represents a significant improvement in data quality, even if no additional fitting is performed (compare Fig. 6, lower left panel and Fig. 6, upper left panel). Although stochastic noise decreases as the number of intensity scans averaged in our procedure increases, approximately 200 scans are sufficient to achieve a ~ 3% increase in RMSD compared to the optimal data obtained through fitting intensity data directly with standard methods implemented in UltraScan (compare Fig. 6, lower center panel and Fig. 6, right panel). The increase in random noise resulting from the subtraction of the buffer data in the reference channel from the sample channel data is clearly seen in the results from the absorption mode, which remains even after using the fitting method (compare Fig. 6, lower center panel and Fig. 6, upper center panel). The TIN pattern seen in the pseudo-absorbance data before noise fitting originates from the cell windows and cannot be removed when subtracting the averaged intensity profile containing just the PMT contributions. However, the outcome is far better in contrast to the absorption mode, where the TIN noise is substantially larger than the data scaled by the PMT profile.

Conclusion

A method for the removal of time- and radially invariant noise is presented that utilizes a large number of averaged intensity scans from an empty rotor hole, without the interference of cell windows or sample absorbance. A clustering algorithm is presented to identify matching wavelengths for each averaged reference scan. The averaged intensity significantly reduces the associated stochastic noise, avoiding the square root of two amplifications of stochastic noise encountered in absorbance data, and is used to convert intensity data to pseudo-absorbance data, in the process removing time-invariant noise components originating in the optical system. These noise components represent the majority of time-invariant noise in a typical sedimentation experimental setup, and achieve time-invariant noise removal without the requirement for fitting the noise components, as is usually performed in the standard UltraScan sedimentation velocity workflow. This is especially important for sedimentation equilibrium experiments where no methods other than reference channel data subtraction exist to remove time-invariant noise contributions from the optical system. Reference data for converting intensity data to pseudo-absorbance data are generated for each wavelength, and saved in the UltraScan LIMS database along with a time stamp for the correction of experimental data. This means that stored reference scans can be reused for noise correction without the need to use a second channel for the collection of buffer/reference data, doubling instead the sample capacity of the rotor. Reference scans need to be repeated whenever optical components in the analytical ultracentrifuge are exchanged, or the performance properties of the instrument change in any way. The benefit for sedimentation velocity experiments is negligible, as other, more efficient methods exist to remove time- and radially invariant noise numerically (Brookes et al. 2010). Our method effectively eliminates flaws in experimental data that are caused by the instrument, rescuing and significantly improving experimental data using algorithms

that are conveniently provided through GUI software in the UltraScan analysis suite. The advances presented here will be especially useful for multi-wavelength analytical buoyant density equilibrium experiments for the accurate quantification of empty/full capsid ratios for adeno-associated viral vectors, by significantly enhancing the signal-to-noise ratio of these equilibrium data.

Acknowledgements

This research and S.M. were supported by the Biomolecular Interaction Technology Center, University of Delaware, and the Canada 150 Research Chairs program (C150-2017-00015, BD), the Canada Foundation for Innovation (CFI-37589, BD), the National Institutes of Health (1R01GM120600, BD), and the Canadian Natural Science and Engineering Research Council (DG-RGPIN-2019-05637, BD). UltraScan supercomputer calculations were supported through NSF/XSEDE under Grant TGMCB070039N (BD) and University of Texas under Grant TG457201 (BD).

Data availability

The UltraScan software used to analyze the AUC data is open source and freely available from the Github repository (https://github.com/ehb54/ultrascan3), the AUC data itself is available upon request from the corresponding author from the UltraScan LIMS server at the Canadian Center for Hydrodynamics. All raw and edited AUC data are available in the open source OpenAUC format.

References

Brookes E, Cao W, Demeler B (2010) A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J 39(3):405–414. 10.1007/s00249-009-0413-5. (Epub 2009 Feb 27) [PubMed: 19247646]

Cao W, Demeler B (2005) Modeling analytical ultracentrifugation experiments with an adaptive spacetime finite element solution of the Lamm equation. Biophys J 89(3):1589–1602. 10.1529/biophysj.105.061135. (Epub 2005 Jun 24.) [PubMed: 15980162]

Cao W, Demeler B (2008) Modeling analytical ultracentrifugation experiments with an adaptive spacetime finite element solution for multicomponent reacting systems. Biophys J 95(1):54–65. 10.1529/biophysj.107.123950. (Epub 2008 Apr 4.) [PubMed: 18390609]

Chen YC (2017) A tutorial on kernel density estimation and recent advances. Biostat Epidemiol 1(1):161–187

Demeler B (2023) The open source, multi-platform UltraScan software project. Available in source code format, and binaries distributions for Windows, Macintosh and Linux: https://github.com/ehb54

Demeler B (2010) Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Curr Protoc Protein Sci. 10.1002/0471140864.ps0713s60.

Demeler B, Gorbet G (2016) Analytical ultracentrifugation data analysis with UltraScan-III. Ch 8. In: Uchiyama S, Stafford WF, Laue T (eds) Analytical ultracentrifugation: instrumentation, software, and applications. Springer, UK, pp 119–143

Hansen JC, Lebowitz J, Demeler B (1994) Analytical ultracentrifugation of complex macromolecular systems. Biochemistry 33(45):13155–13163 [PubMed: 7947722]

Lamm O (1929) Die Differentialgleichung der Ultrazentrifugierung. Ark Mat Astr Fys 21B:1–4

Lebowitz J, Lewis MS, Schuck P (2002) Modern analytical ultra-centrifugation in protein science: a tutorial review. Protein Sci 11(9):2067–2079. 10.1110/ps.0207702. [PubMed: 12192063]

Schuck P, Demeler B (1999) Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. Biophys J 76(4):2288–2296. 10.1016/S0006-3495(99)77384-4. [PubMed: 10096923]

Sheather SJ (2004) Density estimation. Stat Sci. 10.1214/088342304000000297
Silverman BW (2018) Density estimation for statistics and data analysis. Routledge

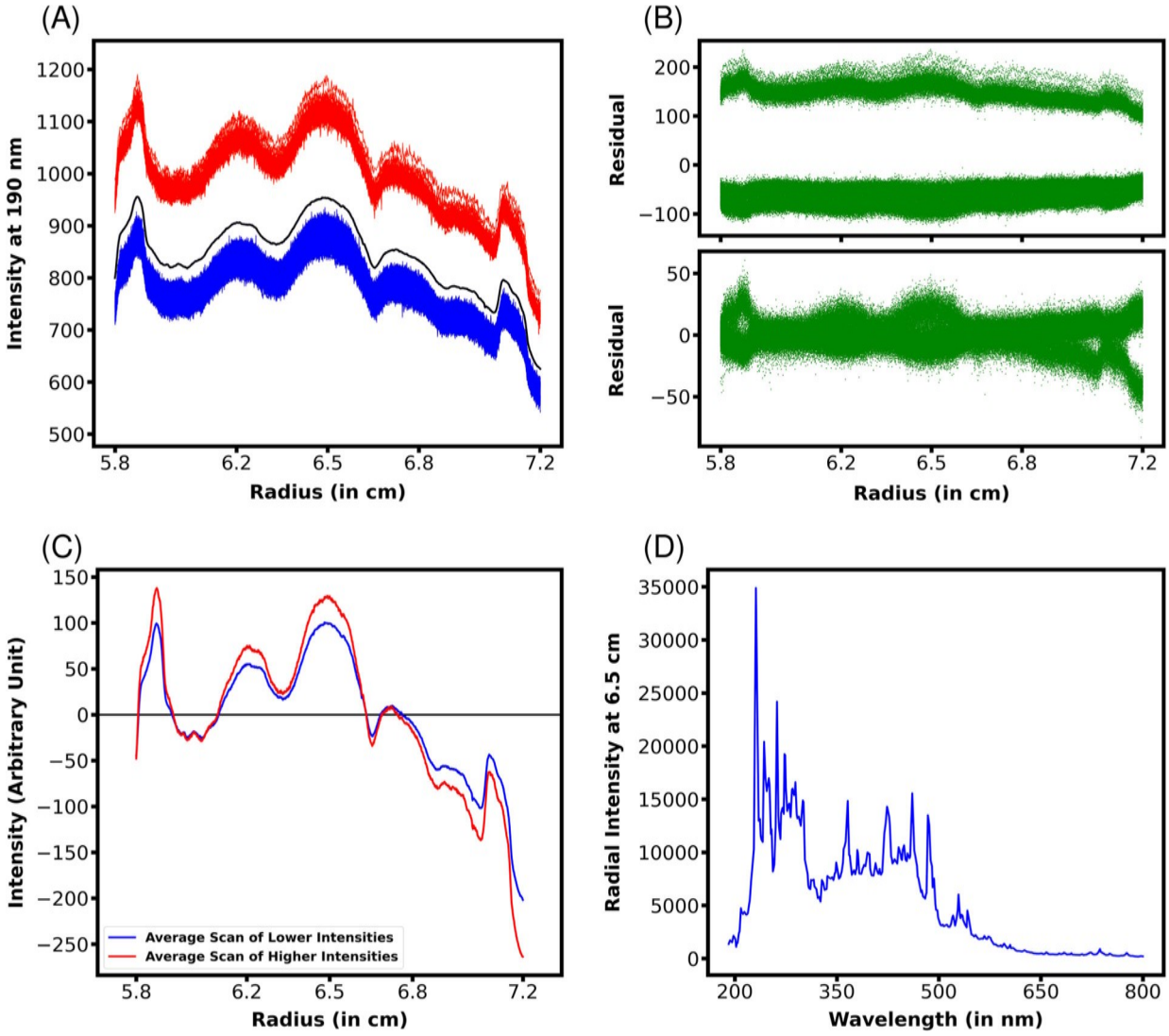

Fig. 1.

A Intensity scans of an empty cell at 190 nm, showing two unique groups (red and blue). The average of all observations is shown in the black line. **B** Upper panel: Residuals when comparing measurements to the intensity average of all scans [black line in (**A**)]. Lower panel: residual errors after subtracting each scan's mean value from each observation. **C** Overlay of the average from the two sets of scans shown in A (the red line corresponds to the red group of scans; the blue line corresponds to the blue group of scans). Both scans are aligned by their mean values. **D** Intensity emission profile of Xenon flash lamp in the Optima AUC

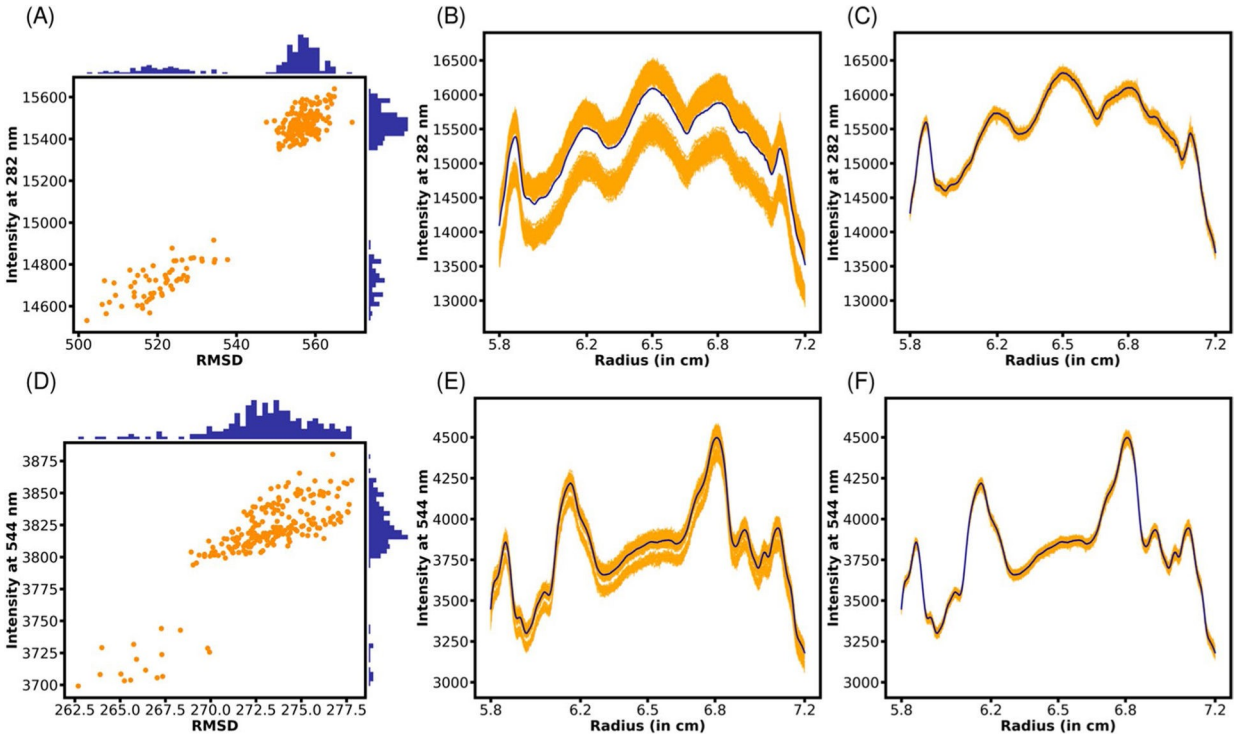

Fig. 2.

Wavelength jumps observed in the Optima AUC. **A** 2D graph of the scan intensities plotted as a function of their RMSD when compared to the average of all scans [blue line in (**B**)] at 282 nm. **B** Radial intensity scans collected at 282 nm before clustering. The blue line denotes the average intensity for each radial point when averaging all scans collected at 282 nm. **C** Same as (**B**), but after clustering and alignment. **D**–**F** Same as (**A**–**C**), but for 544 nm

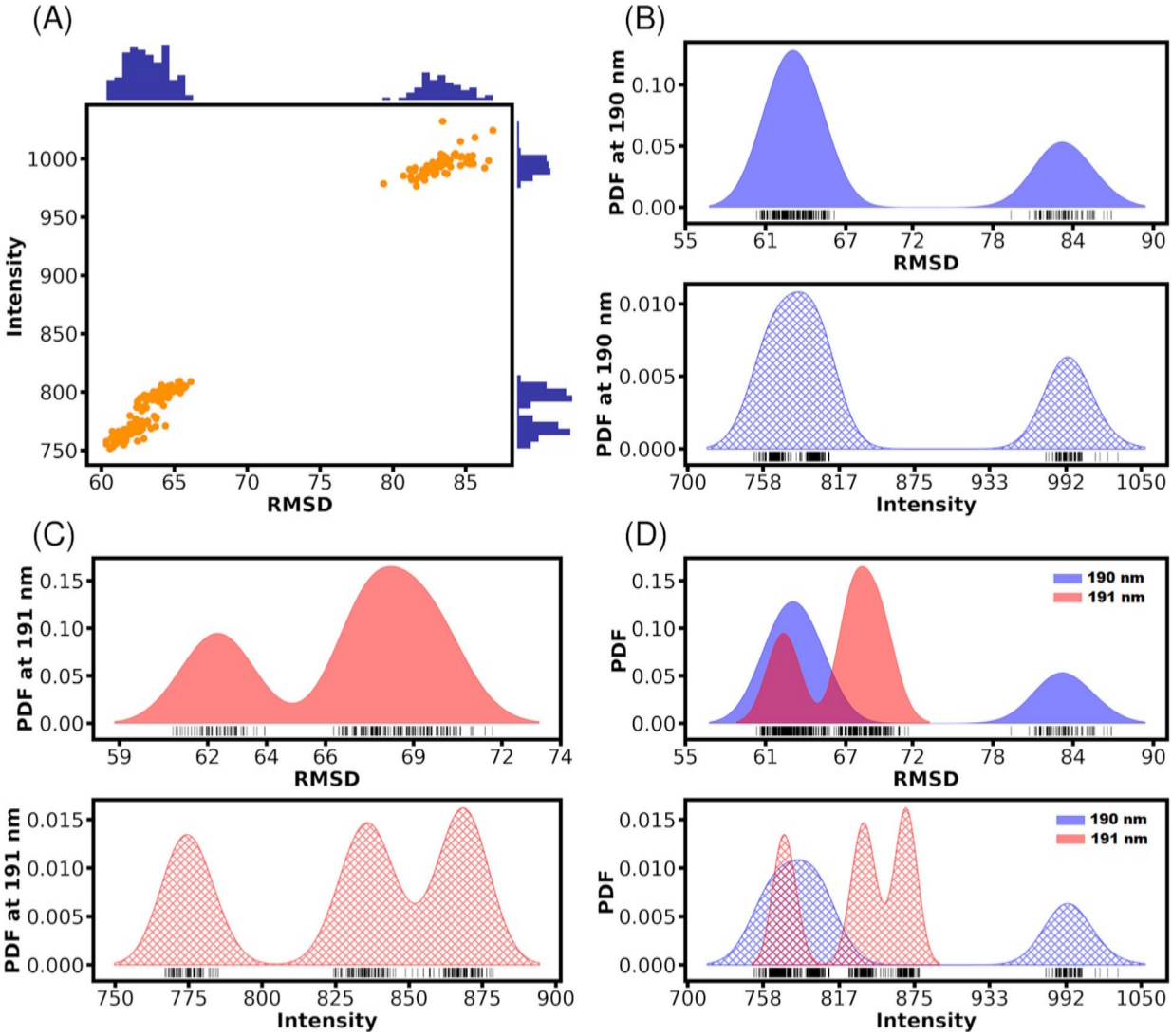

Fig. 3.

A 2D graph of scan intensities as a function of their RMSD values at 190 nm, as well as their histograms. **B**, **C** The probability distribution function of RMSD values (upper solid color image) and intensity values (lower hatch color) at 190 nm and 191 nm, respectively. **D** RMSD and intensity distribution overlaps between 190 nm and 191 nm

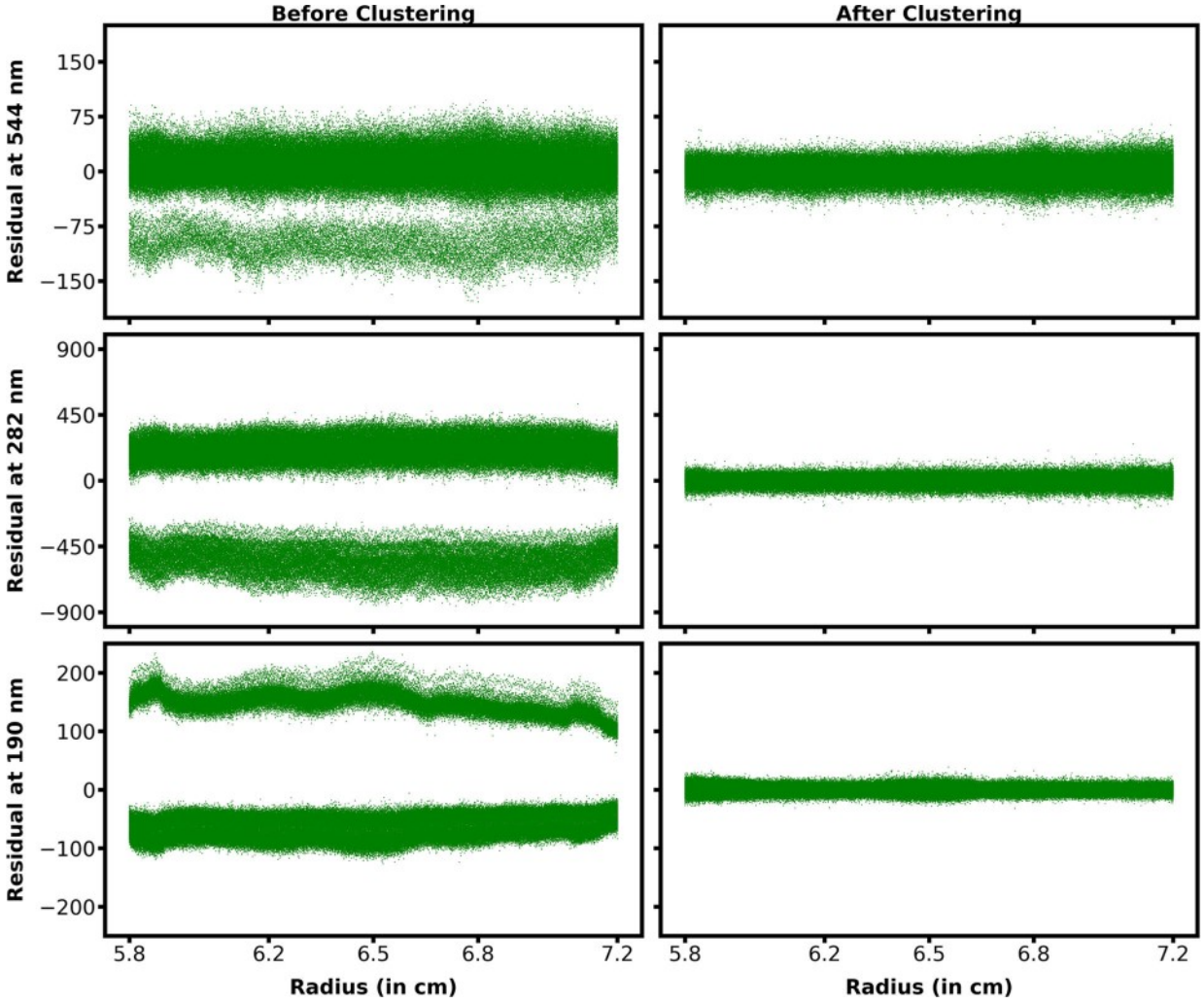

Fig. 4.

Residual errors of the reference scan at wavelengths 190 nm, 282 nm, and 544 nm before and after the clustering process

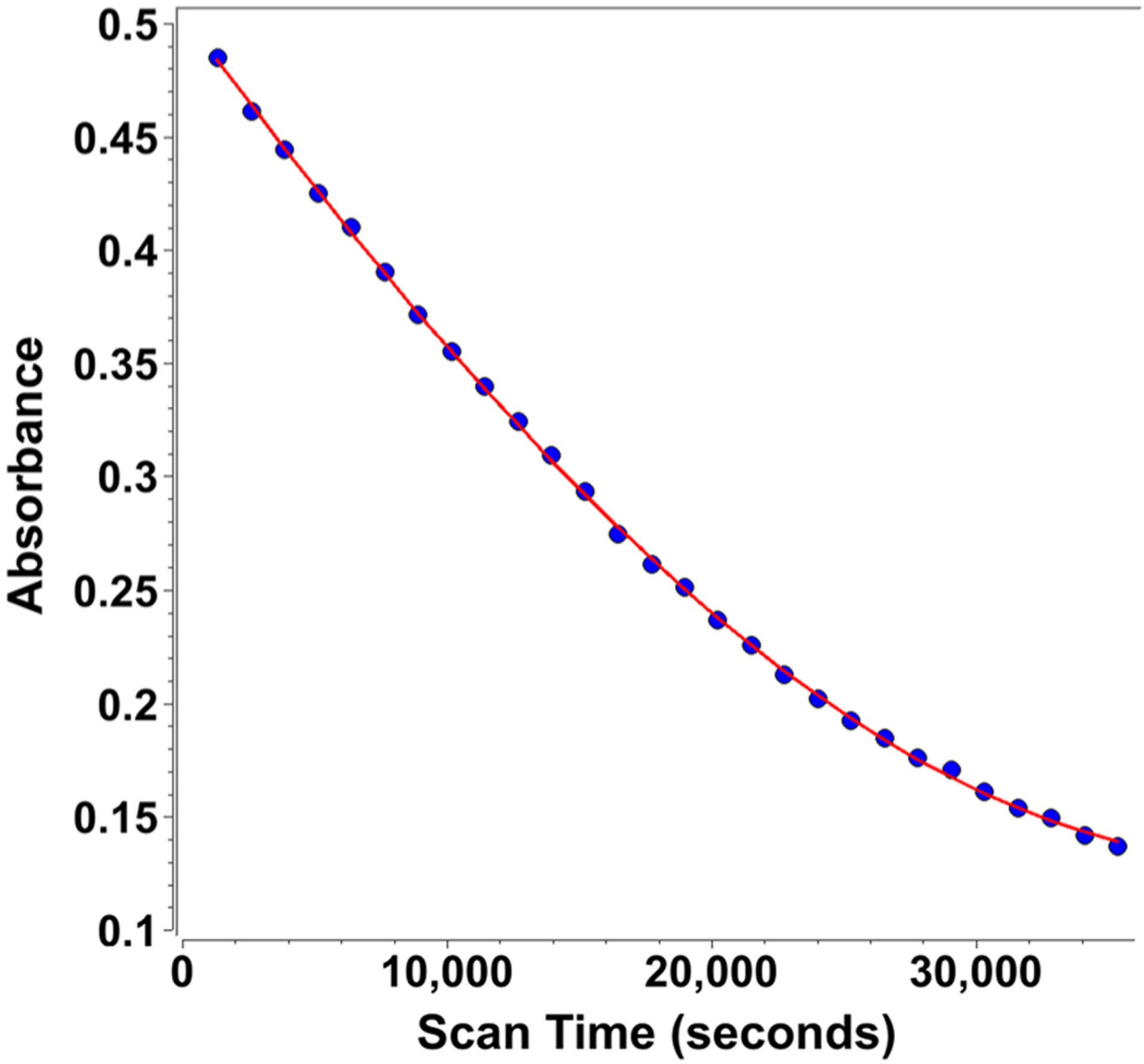

Fig. 5.

Integration of the concentration flux (blue circles) in the AUC cell from an absorbance scan with a polynomial function fit (red line)

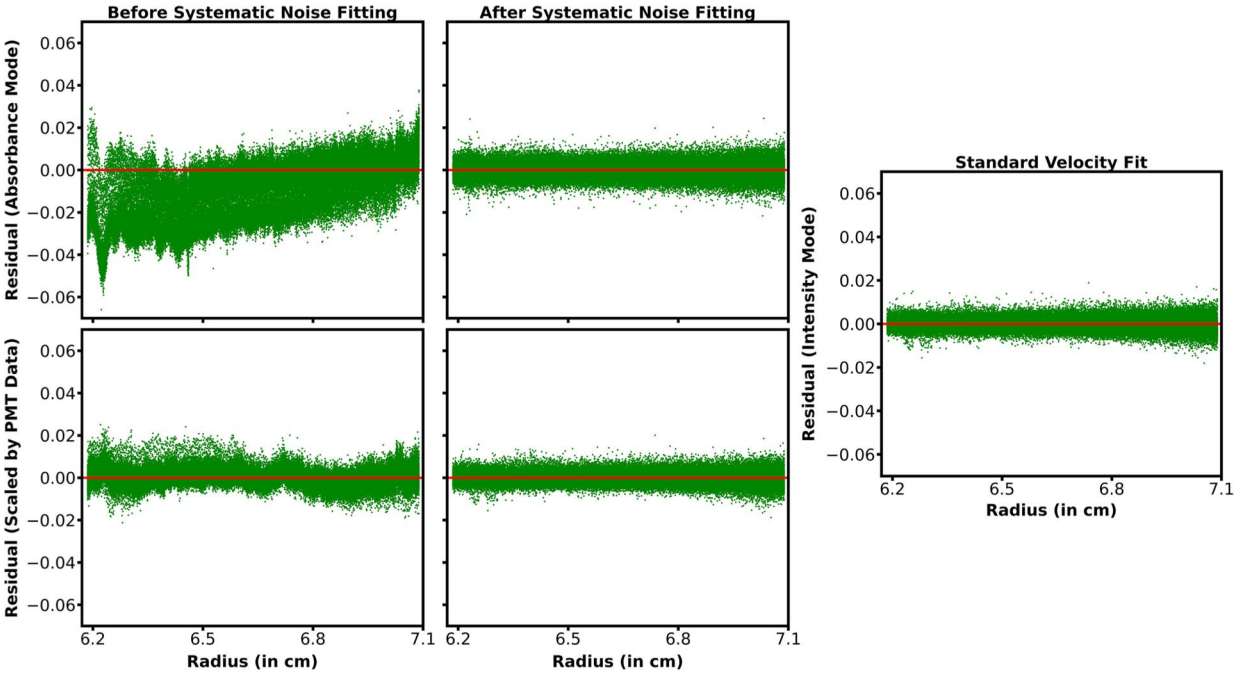

Fig. 6.

The RMSD profiles from an SVE of bovine serum albumin by fitting absorbance data without standard UltraScan RIN and TIN processing (top left panel) and intensity data processing as described in this work, but without standard UltraScan RIN and TIN processing (lower left panel), UltraScan-processed RIN and TIN fitting for absorbance data (center top panel) and intensity data processes as described in this work, but with standard UltraScan RIN and TIN fitting (bottom center panel), and intensity data using standard UltraScan methods for RIN and TIN fitting

Table 1

The RMSD values of fitting for BSA solution in three different modes

	Wavelength (nm)	RMSD (before systematic noise fitting)	RMSD (after systematic noise fitting)
Intensity mode (standard velocity fit)	218	–	0.002964
	219	–	0.002769
	220	–	0.002663
	221	–	0.002637
	222	–	0.002365
			Ave = 0.002680 ± 0.000195
Absorbance mode (Beckman)	218	0.019098	0.003919
	219	0.020463	0.003640
	220	0.019025	0.003602
	221	0.017918	0.003603
	222	0.017574	0.003197
		Ave = 0.018816 ± 0.001018	Ave = 0.003592 ± 0.000230
Pseudo-absorbance (scaled using PMT profiles)	218	0.004186	0.003012
	219	0.004316	0.002818
	220	0.004241	0.002785
	221	0.004155	0.002732
	222	0.003848	0.002480
		Ave = 0.004149 ± 0.000160	Ave = 0.002765 ± 0.000171