SasView Tutorials

Simultaneous 1D Data Fitting in SasView Version 6

www.sasview.org

Preamble

SasView was originally developed by the University of Tennessee as part of the Distributed Data Analysis of Neutron Scattering Experiments (DANSE) project funded by the US National Science Foundation (NSF), but is currently being developed as an Open Source project hosted on GitHub and managed by a consortium of scattering facilities. Participating facilities include (in alphabetical order): the Australian National Science & Technology Centre for Neutron Scattering, the Diamond Light Source, the European Spallation Source, the Federal Institute for Materials Research and Testing, the Institut Laue Langevin, the ISIS Pulsed Neutron & Muon Source, the National Institute of Standards & Technology Center for Neutron Research, the Oak Ridge National Laboratory Neutron Sciences Directorate, and the Technical University Delft Reactor Institute.

If you make use of SasView

If you use SasView to do productive scientific research that leads to a publication, we ask that you acknowledge use of the program with the following text:

> *This work benefited from the use of the SasView application, originally developed under NSF Award DMR-0520547. SasView also contains code developed with funding from the EU Horizon 2020 programme under the SINE2020 project Grant No 654000.*

Contributors to this Tutorial

Steve King (stephen.king@stfc.ac.uk)
Piotr Rozyczko (piotr.rozyczko@esss.se)
Wojciech Potrzebowski (wojciech.potrzebowski@esss.se)
Paul Butler (butlerpd@udel.edu)
Miguel Gonzalez (gonzalezm@ill.fr)

Last revised: 22 August 2024

Learning Objective

This tutorial will demonstrate how to simultaneously fit 1D ('intensity' versus Q) datasets in SasView. Batch fitting of multiple 1D datasets, and the fitting of 2D datasets, are considered in separate tutorials.

It is assumed that the reader has some familiarity with the purpose and principles of data fitting. If not, these Wikipedia articles provide an overview:

- https://en.wikipedia.org/wiki/Curve_fitting
- https://en.wikipedia.org/wiki/Mathematical_optimization

It is also assumed that the reader is familiar with using SasView to fit individual datasets (Single Fitting Mode) as covered in the tutorial ***basic_1d_fitting_in_sasview***.

The program interface shown in this tutorial is SasView Version 6.0.0 running on a Windows platform but, apart from a few small differences in look and functionality, this tutorial is generally applicable to any version of SasView Version 6.x on any platform. However, there are separate tutorials for SasView 5.x and for using the old program interface released with SasView 4.x.

Glossary

a priori information	Known facts about the system whose datasets are being modelled that can guide the selection of model or model parameters.
Chi-square (X^2, 'Chi2')	A statistical test of how well a chosen model fits the data with a given set of model parameters. In SasView this means $X^2 = \sum ((I(Q)_{meas} - I(Q)_{calc})^2 / E(Q)^2)$ where I is the scattering intensity and E is the error on the intensity value. Clearly, as $X^2 \rightarrow 0$, the better the model fit is. NB: The SasView interface actually reports a variation of chi-square called the reduced chi-square, sometimes referred to as the 'goodness-of-fit' $X^2_{Reduced} = \sum ((I(Q)_{meas} - I(Q)_{calc})^2 / E(Q)^2) / (N_{pts} - N_{params})$ where N_{pts} is the number of data points in the dataset and N_{params} is the number of model parameters being optimised (which may be less than the total number of parameters in the model!). As $X_R^2 \rightarrow 1$, the better the model fit is.
Compute/Plot [button]	Perform a direct calculation of the model with the current parameters but without any optimisation

Correlated parameters	To obtain the best solutions the optimiser needs to explore the widest possible parameter space. That is best achieved if every parameter can be considered independent of every other parameter. However, there will be instances where this is not the case and one or more parameters may be correlated. In such cases the best approach is to fix the values of some of correlated parameters using *a priori* information and then optimise the remaining values. A particularly common instance of correlated parameters encountered when model-fitting SAS data is when the components of the forward scattering intensity $I(0)=\phi V(\Delta\rho)^2$ where Φ is the volume fraction of scatterers, V is the volume of one scatterer (and so dependent on size parameters), and Δρ is the contrast (difference in SLDs), are separated out as individual parameters in a model. Correlated parameters often manifest themselves in the SasView interface as having very large uncertainties.
Fit [button]	Perform a calculation of the model optimising the selected parameters
Model-fitting	The process of finding a good yet physically-realistic mathematical description ('solution') for a dataset or collection of datasets. The procedure employed to achieve this is called optimisation.
OpenCL	A low-level software framework that allows calculations to be distributed between any compatible processors (eg, GPUs as well as CPUs) on the host computer. OpenCL can speed up demanding model-fits if suitable hardware is available.
OpenMP	A software framework that permits shared-memory multi-processing (ie, parallelisation) of calculations. OpenMP can speed up demanding model-fits if suitable hardware is available.
Optimiser	The mathematical algorithm used to perform the model-fitting.
Polydispersity	Where one or more model parameters have a distribution of values. SasView allows for 2 types of polydispersity: *Size* — where, for example, the radii and/or lengths of the scatterers have a distribution of values *This will apply in most instances*

	Angular	where the scatterers are anisometric in shape (eg, cylinders) and exhibit preferred orientations with respect to Q, for example, in under shear or in a magnetic field
Reduced Chi2	See Chi-square	
Residual	The difference between the measured and calculated values of a function at a given point. In SasView this means: $R(Q) = I(Q)_{meas} - I(Q)_{calc}$ where *I* is the scattering intensity. NB: The SasView interface actually reports a variation of the residual called the normalised residual $R(Q)_{Normalised} = (I(Q)_{meas} - I(Q)_{calc})/E(Q)$ where *E* is the error on the intensity value. A normalised residual can be thought of as the number of standard deviations between the measured value and the calculated value. Thus, for a good fit, 68% of the values will lie within -1 < $R(Q)_{Normalised}$ < +1, and 95% within +/-2. Residuals larger than +/-3 indicate significant problems with either the input data or the choice of model or model parameters.	
SLD	Abbreviation for Scattering Length Density, a measure of the ability of a molecule to scatter. Strictly speaking, SLD is a SANS quantity, so if fitting SAXS data use electron density values in their place. SLD values (neutron and X-ray) can be calculated with the SLD Calculator Tool in SasView.	
Smearing	Sometimes the instrumental geometry used to acquire the experimental data has an impact on the clarity of features in the scattering curve. For example, peaks or fringes might be slightly broadened. This is known as Q-resolution smearing. To compensate for this effect SasView can add a resolution contribution into a model calculation (which by definition will be exact) to make it more representative of what has been measured experimentally. SasView provides 4 smearing options:	

	None	no smearing correction is applied
	Use dQ Data	the measured Q-resolution for each data point is used to apply a smearing correction *This is the default if dQ data is present*
	Custom Pinhole	SasView will apply a smearing correction suitable for data measured on 'pinhole geometry' instruments (eg, most SAXS/SANS instruments)
	Custom Slit	SasView will apply a smearing correction suitable for data measured on 'slit geometry' instruments (eg, most USAXS/USANS instruments)
Theory	The name SasView gives to a model calculation.	
Uncertainties	Every experimental measurement, including the measurement of *I(Q)*, is subject to some degree of error (which will, ideally, be included in the dataset). Similarly, the parameters returned by optimisation will have some associated range of uncertainty. Parameters with uncertainties that are more than 95% of the parameter value should be viewed with deep suspicion. NB: Unhelpfully, but for reasons of space, the SasView interface actually labels parameter uncertainties as errors.	
Weighting	An optimiser can be instructed to pay less or more attention to data points in a dataset by changing the weighting of those data points. SasView provides 4 weighting options:	
	No Weighting	all data points will be weighted equally
	Use dI Data	the data points will be inversely weighted according to their measured intensity errors (ie, less prominence will be given to data points with large errors) *This is the default if dI data is present*
	Use sqrt |(*I Data*)|	the data points will be inversely weighted according to the square root of their intensity values
	Use |(*I Data*)|	the data points will be inversely weighted according to their intensity values

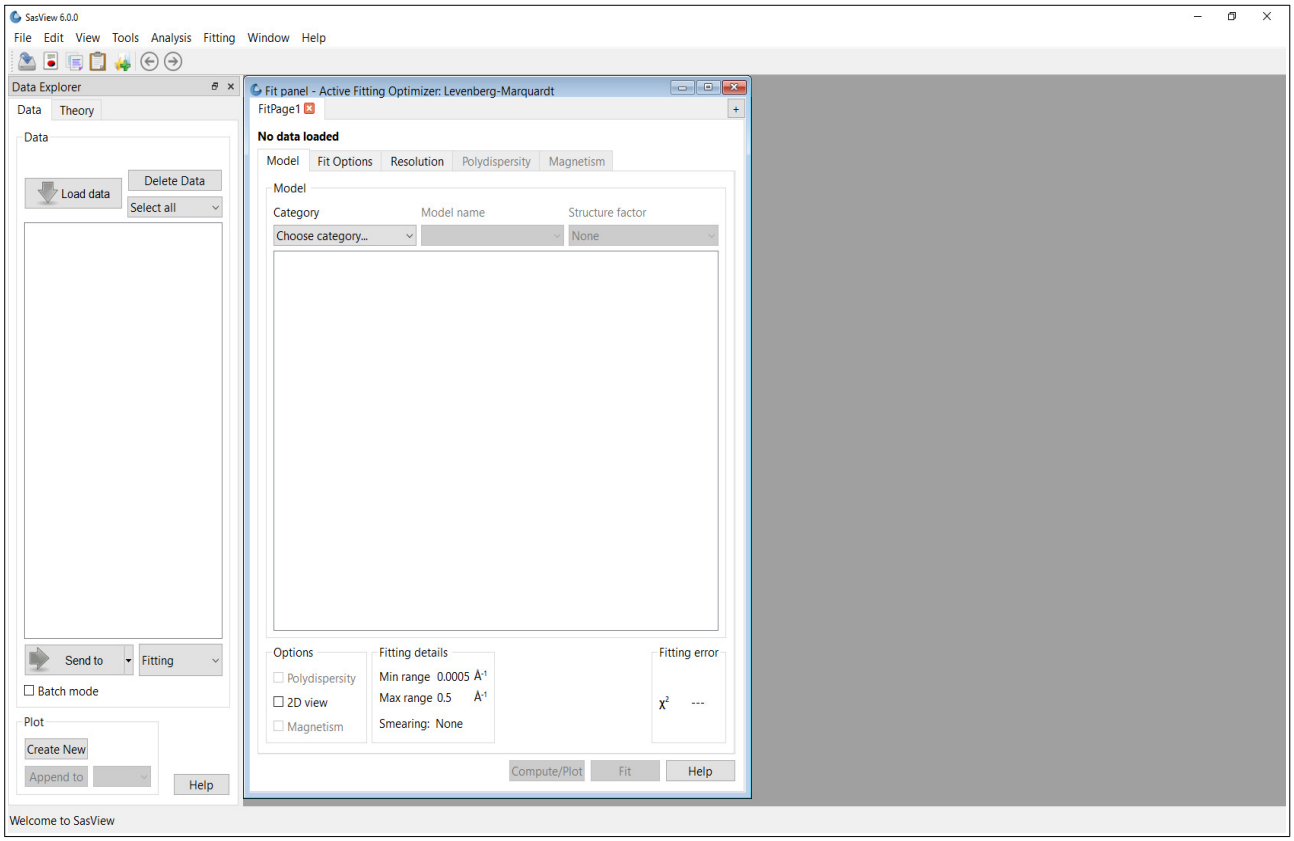
Running SasView

Windows

Either select SasView from '**Start**'> '**All Programs**' or, if you asked the installer to create one, double-click on the SasView desktop icon.

Mac OS

Go in to your '**Applications**' folder and select SasView.



Forward

Simultaneous Fitting is a hugely useful extension of Single Fitting that allows for some relatively extensive constraints between parameters in several FitPage's (eg, to constrain all parameters in a series of datasets at different contrasts to be the same except for the solvent SLD parameter, or to fix the radius of the sphere in one FitPage to be the same as the radius of the cylinder in a second FitPage, etc), or between parameters in a single FitPage (eg, to constrain the length to be twice that of the radius, etc).

***Simultaneous Fitting may also be used to co-fit SANS and SAXS data** from the same sample. However, this is not quite as straightforward as it sounds, because SAXS data is typically over-binned (i.e. contains more data points), the intensity uncertainties on each data point are typically far, far, smaller (because of the greater brilliance of X-ray sources), and the instrumental resolution may be significantly different, compared to the SANS data. The first two factors, in particular, bias the fit optimisation in favour of the SAXS data. Analogous concerns will apply to the simultaneous fitting of, say, SANS and USANS (or SAXS and USAXS) data and, possibly, SANS and SESANS data.*

Different use cases may therefore require different bias compensation strategies.

One conceptually simple compensation strategy is to change the weighting one or more of the datasets have in the optimisation process. This can be done by computing a 'significance factor' (or 'baseline weighting') for each dataset and then multiplying that by some 'adjustment factor'. The problem with this approach, of course, lies in deciding on a suitable 'significance factor' and how much to vary it by!

As an example, a 'significance factor' for dataset d having N data points might be defined as:

$$SF_d = \sum_{i=1}^{N} \mathrm{E}_i(\sigma_i)$$

where σ is the relative error (dI / I) on a data point and E_i might then be:

σ *; easy to understand*

σ^2 *; analogous to the variance in statistics*

$1/\sigma$ *; useful with systematic uncertainty*

$1/\sigma^2$ *; the usual form for weighting in statistics*

$1/\sqrt{\sigma^2}$ *; analogous to a Normal distribution*

An added complication is that the effectiveness of a chosen compensation scheme may be impacted by ones choice of Fitting Opitmizer.

SasView version 6.x has the option to implement a compensation strategy with a significance factor using $1/\sigma^2$.

Example

This example demonstrates a simultaneous model fit with constraints to three datasets, including the use of the DREAM optimiser and compensating for size polydispersity and instrument resolution.

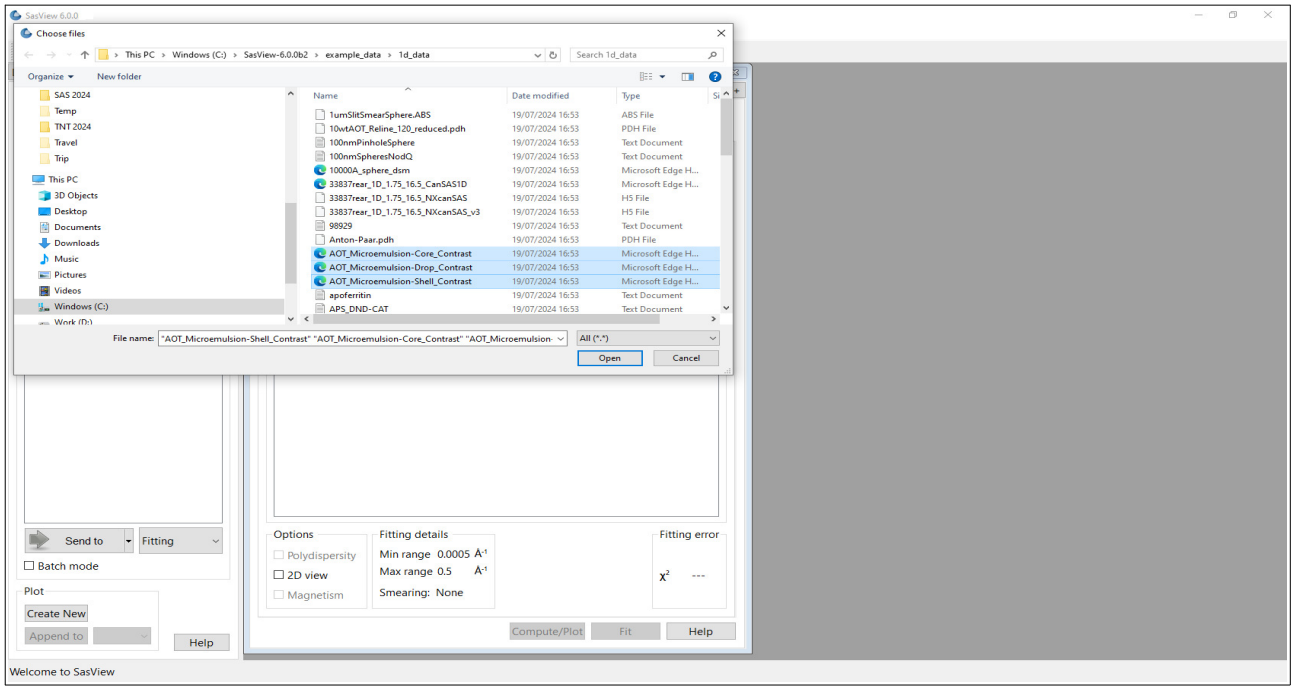
Simultaneous Fitting is an extension of the Single Fitting that allows for some relatively extensive constraints between several FitPage's (eg, to constrain all parameters in a series of datasets at different contrasts to be the same except for the solvent SLD parameter, or to fix the radius of the sphere in one FitPage to be the same as the radius of the cylinder in a second FitPage, etc), or between parameters in a single FitPage (eg, to constrain the length to be twice that of the radius, etc).

Tip: If you need to fit a custom Plugin Model, you must ensure that model is available to SasView before proceeding.

In the Data Explorer panel, click the Load Data button, and navigate to the **\example_data\1d_data** folder in the SasView installation directory.

Select the datasets **AOT_Microemulsion-Core_Contrast.xml**, **AOT_Microemulsion-Drop_Contrast.xml** and **AOT_Microemulsion-Shell_Contrast.xml**, and click the Open button.

These data are from a water-in-cyclohexane microemulsion system stabilised by a layer of the surfactant Aerosol OT® (AOT) located at the water/oil interface. The three datasets correspond to three different contrast conditions highlighting three different structural components: the 'core' (just the water droplets), 'drop' (the water droplets plus the surfactant layer), and the 'shell' (just the surfactant layer). The different contrasts (differences in SLD values) are achieved by changing which components are deuterated.

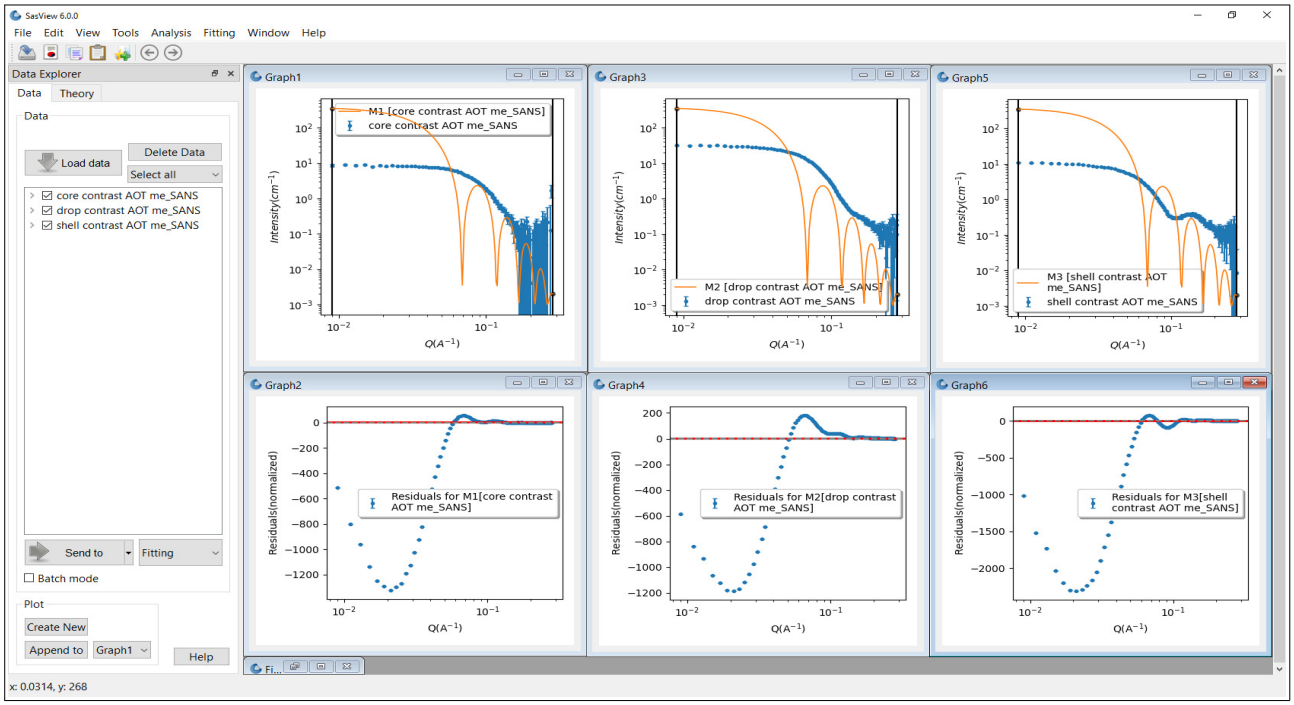


At the bottom of the Data Explorer panel, ensure that the analysis drop-down box says *Fitting* and click the Send to button. As the load will have left all three datasets checked in the Data Explorer panel, this should result in the creation of three FitPage's.

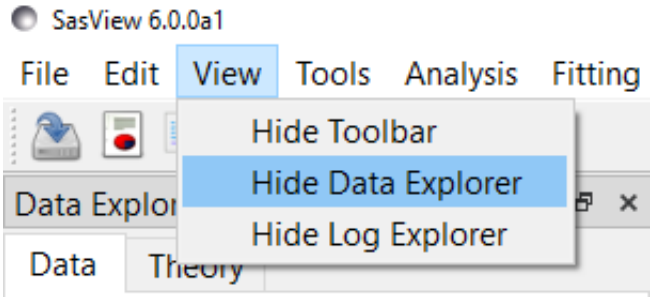
Tip: If the multiple datasets are in one file, load that file, Unselect all Data (at the top of the Data Explorer panel), reselect just those datasets to be fitted, and then Send to Fitting.

Go to each FitPage in turn and select the **core_shell_sphere** model.

Then click on **Compute/Plot** on each page to see the graphs of the three theory curves added to dataset points.



You may wish to hide the Data Explorer window to maximise the available work space.



Enter the following SLD values in the appropriate FitPage's:

Contrast / SLD	*sld_core*	*sld_shell*	*sld_solvent*
'core'	$+6.39\times10^{-6}$ Å^{-2}	$+0.62\times10^{-6}$ Å^{-2}	-0.28×10^{-6} Å^{-2}
'drop'	-0.55×10^{-6} Å^{-2}	$+0.62\times10^{-6}$ Å^{-2}	$+6.68\times10^{-6}$ Å^{-2}
'shell'	$+6.39\times10^{-6}$ Å^{-2}	$+0.62\times10^{-6}$ Å^{-2}	$+6.68\times10^{-6}$ Å^{-2}

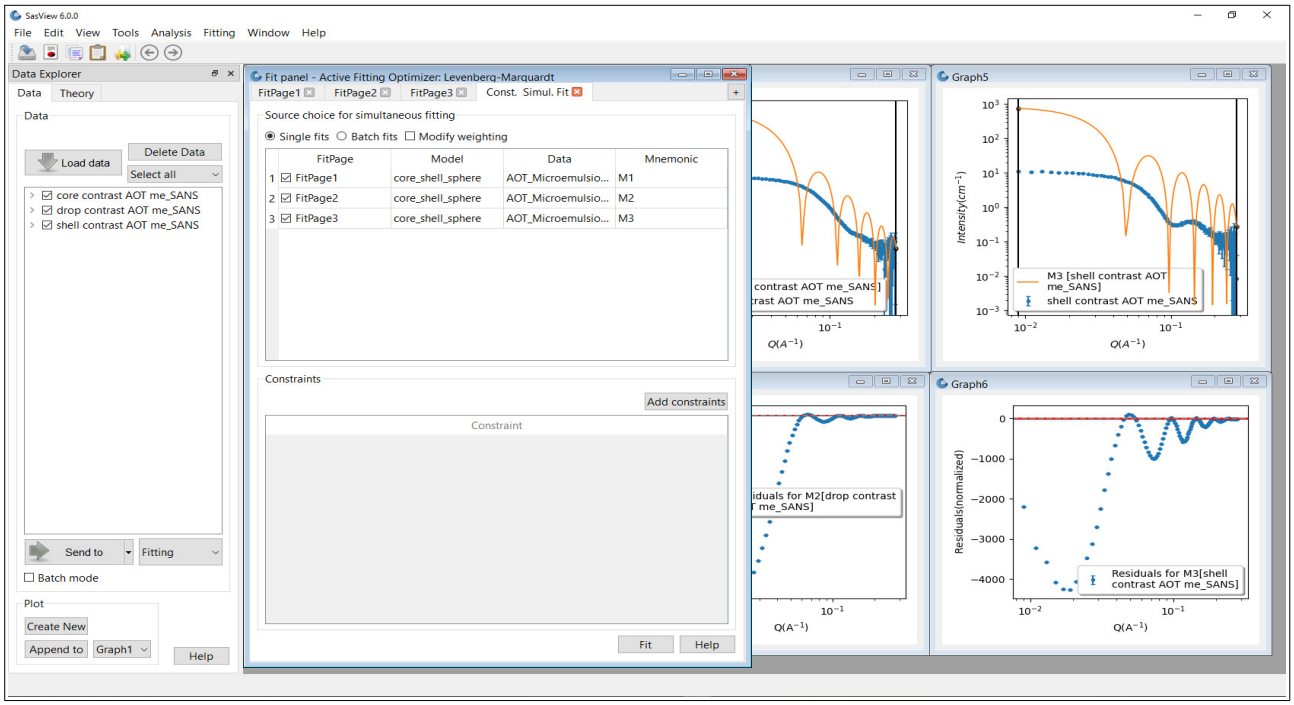
The objective of this simultaneous fit will be to find common values for the *radius* (of the droplet cores) and *thickness* (of the surfactant layer) parameters that provide a good solution to all three datasets.

We shall also assume that the person who made the samples was meticulous and that the volume fraction of droplets in each sample is the same!

As when fitting a single dataset, we still have to tell SasView which parameters we want it to optimise. So go to each FitPage and check the *scale*, *background*, *radius* and *thickness* parameters.

The next step is to tell SasView that the *radius* parameters for the 'core', 'drop' and 'shell' contrast datasets are equal, and then do the same for the *thickness* and *scale* parameters. This is called introducing constraints.

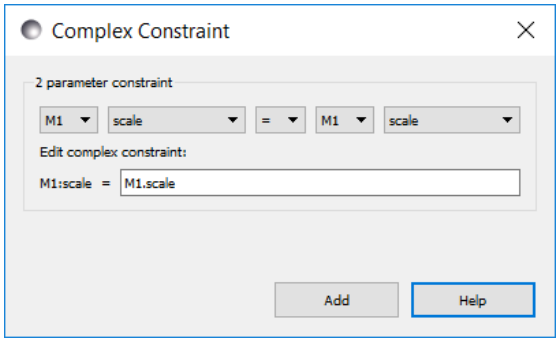
Go to the Menu Bar and click Fitting followed by Constrained or Simultaneous Fit. In the Constrained & Simultaneous Fit page that appears, make sure that the boxes next to the three FitPage names are selected (because we want to construct constraints with all three theories).



Tip: To change the weighting of datasets during optimization, check the Modify weighting box. The 'Weighting' boxes that appear are the 'adjustment factors'; see the *Forward*.

For more information on using this functionality, please refer to the separate tutorial '*Modifying Weights in SasView*'.

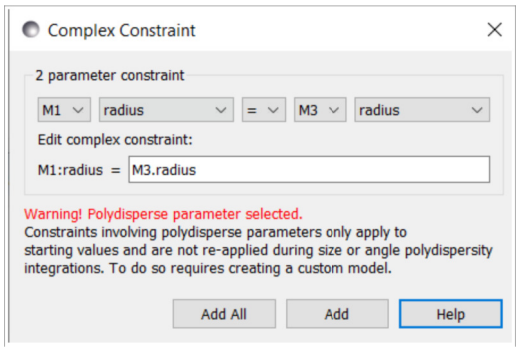
Then, in the section of the page called Constraints, click the Add constraints button. A new Complex Constraint dialog appears.



Constraints can either be added individually or *en masse* (an Add All button will appear when one of the theory drop-downs is changed). Use the drop-downs to create the constraint:

M1.radius = M3.radius

and then click Add.



Repeat the process so as to construct the constraints:

M1.thickness = M3.thickness

M1.scale = M3.scale

M2.radius = M3.radius

M2.thickness = M3.thickness

M2.scale = M3.scale

Close the Complex Constraint dialog box.

Aside: Although these constraints equate 'core' and 'drop' parameters to their 'shell' counterparts, this choice is entirely arbitrary. The choice of which dataset goes on which side of the constraint equality should not influence the final solution.

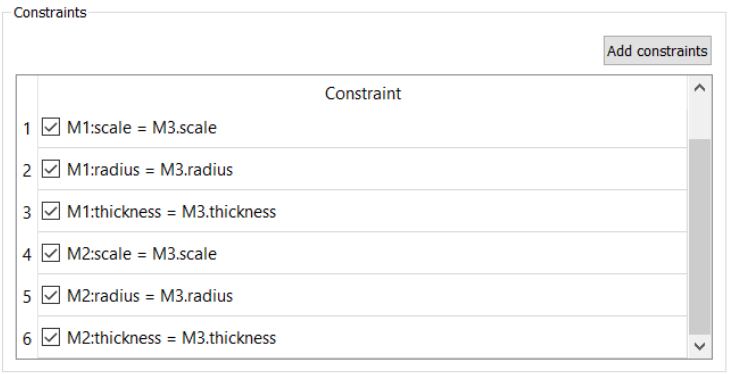
Tip: Do not chain constraints, for example, by doing:

M1.radius = M3.radius

M2.radius = M1.radius

This could lead to unpredictable results.

As they are created the constraints are listed in the Constraint table in the lower part of the Const. Simul. Fit page.



Note that these constraints represent the simplest set of constraint conditions but, if required, more complex constraints could be created. For example, if it were known that the volume fraction of droplets in the 'drop' dataset were 10% lower than in the 'shell' dataset then instead of the constraint

M2.scale = M3.scale

one could have

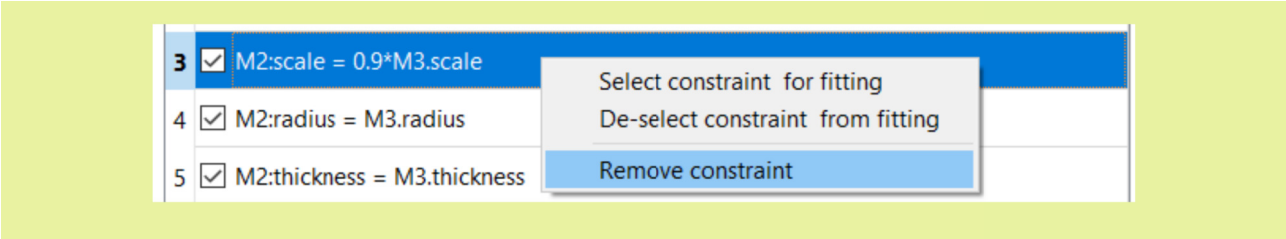
M2.scale = 0.9 * M3.scale

and so on.

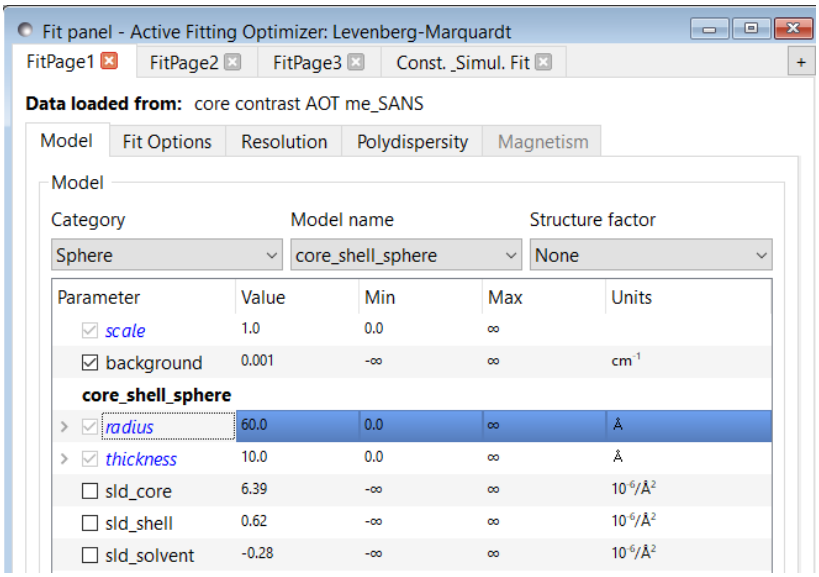
Note: unlike earlier versions of SasView, SasView 6.x also allows you to add constraints involving polydispersity parameters. This is demonstrated later.

Tip: Double-clicking on a constraint allows you to edit it directly:

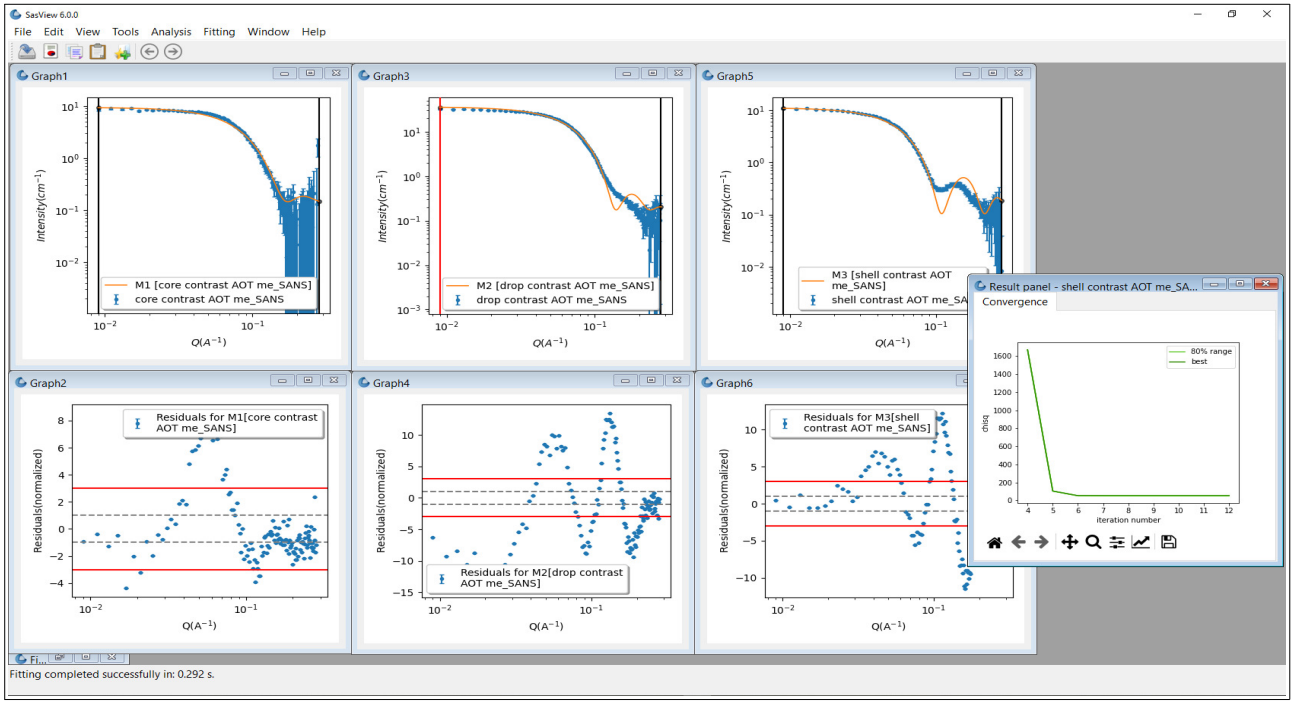
And right-clicking on a constraint allows you to select/deselect and remove it.



Now click on FitPage1 or FitPage2. Notice that the constrained parameters are shown in *blue italic* font.



When ready, click the Fit button on the Const. Simult. Fit tab. <u>DO NOT click the Fit buttons on the individual FitPages</u>.

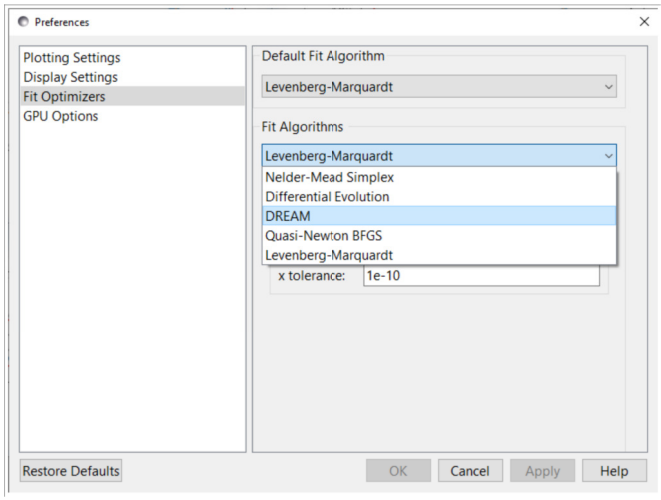


Fitting proceeds and the plots and FitPage parameters update. A Results Panel will also appear.

The fits returned leave room for improvement, particularly in the case of the 'drop' contrast (theory M2). This might be because the default optimiser is struggling to handle this multiple dataset problem. To test this, we can change the optimiser.

Tip: In a constrained fit the number of degrees of freedom are reduced by the constraints. This will impact the *Reduced Chi2* value shown on a specific FitPage.

Go to the Menu Bar and click Fitting followed by Fit Algorithms.



The default optimiser, called the Levenberg-Marquardt, is a fast optimiser, but is prone to finding what are called 'local solutions' in parameter-space, rather than the true 'global solution'. More robust optimisers will be more likely to find the 'global solution' but will be slower. The most robust optimiser in SasView is the one called DREAM.

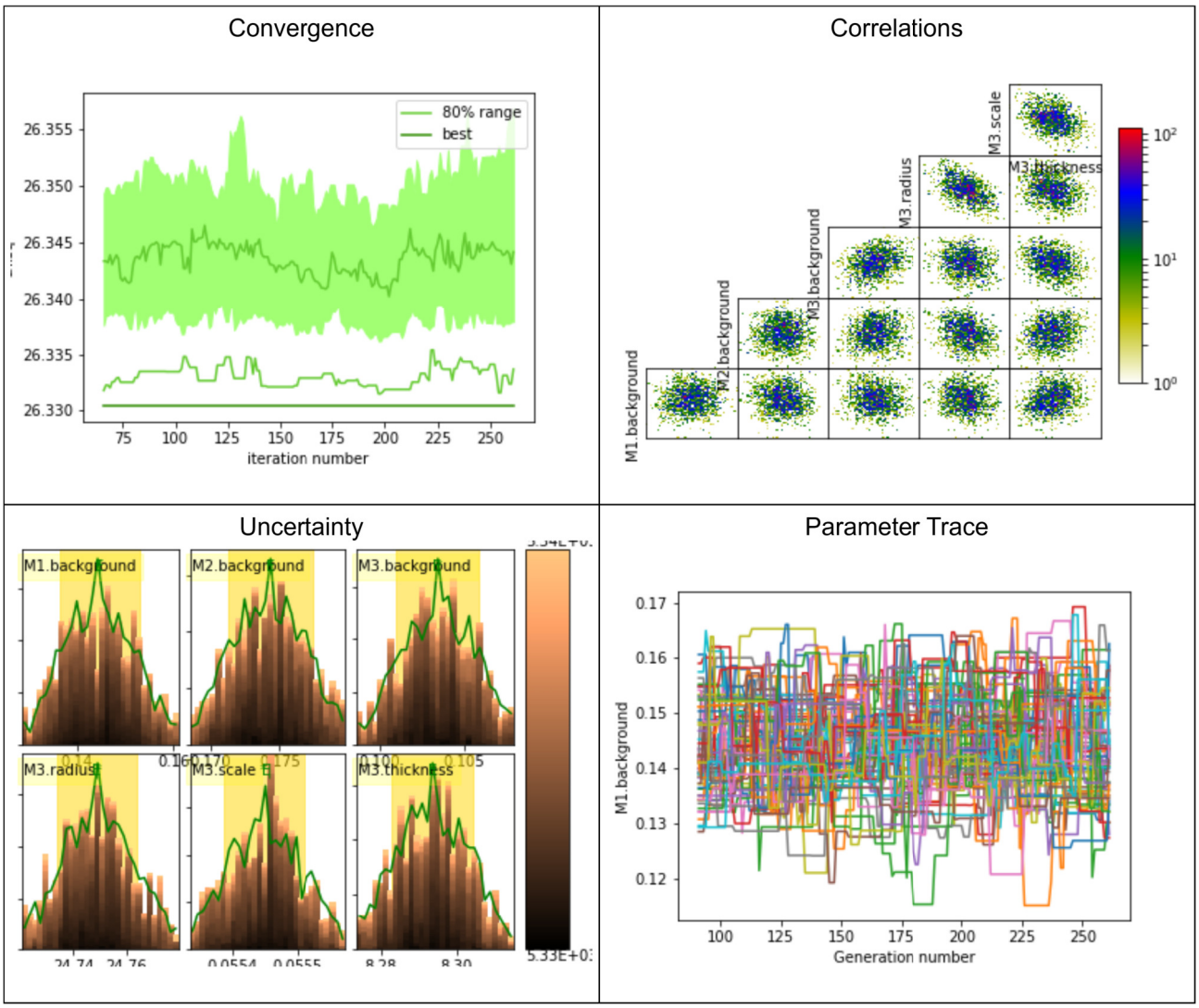
Select DREAM from the drop-down list of algorithms and click Apply and Cancel.

Tip: Detailed information about the different optimisers available in SasView can be found in the Help Documentation.

Return to the Const & Simul Fit page and click Fit. When the fit has completed, go to the Menu Bar and click Fitting and then select Fit Results (it might be hidden).

The Result Panel now contains additional tabs providing detailed insight into the quality of the fit. The tabs are:

- *Convergence:* how X^2 has evolved during the fit
- *Correlation:* 'maps' depicting the relationships between parameters
- *Uncertainty:* the distribution of best fit values for each parameter
- *Parameter Trace:* the relative variation of each parameter



These plots indicate that DREAM has found a robust solution, but it is also evident from the theory fits to the experimental data that the model is still misrepresenting the data. Something else needs to be incorporated.

On each FitPage, check the Polydispersity box, then go into the Polydispersity tab and set a *Distribution of radius* parameter (*PD[ratio]*) of 0.15 with a **lognormal** distribution, and check the box so it will optimise.

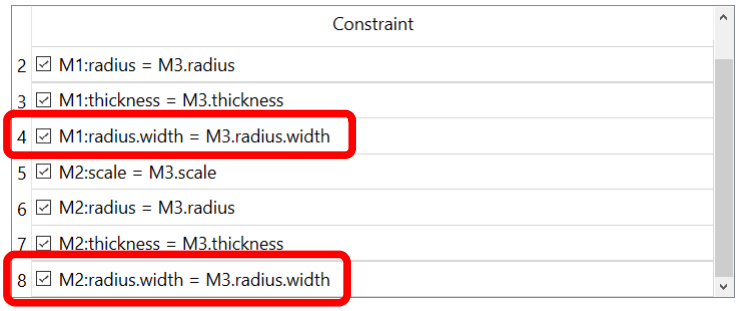
Polydispersity and Orientational Distribution

Parameter	PD[ratio]	Error	Min	Max	Npts	Nsigs	Functior
☑ Distribution of radius	0.15		0.0	1.0	80	8	lognormal

Tip: Select the polydispersity distribution *function* type before setting the *PD[ratio]* value.

Three additional plots will appear for each of the *radius* polydispersity distributions. Minimise them.

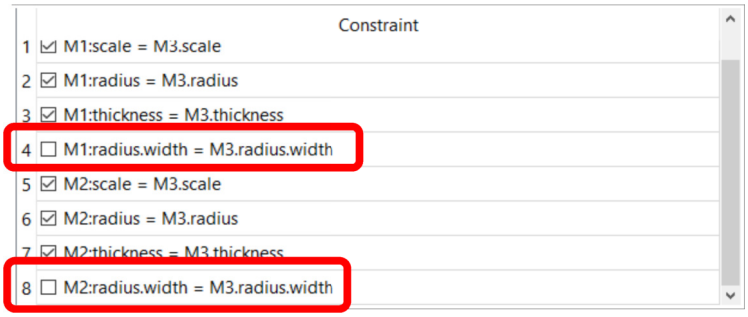
Return to the Const & Simul Fit page and create constraints to tell SasView that the *Distribution of radius* parameters for the 'core' and 'drop' contrast datasets are the same as the *Distribution of radius* parameter for the 'shell' contrast dataset.



Click Fit.

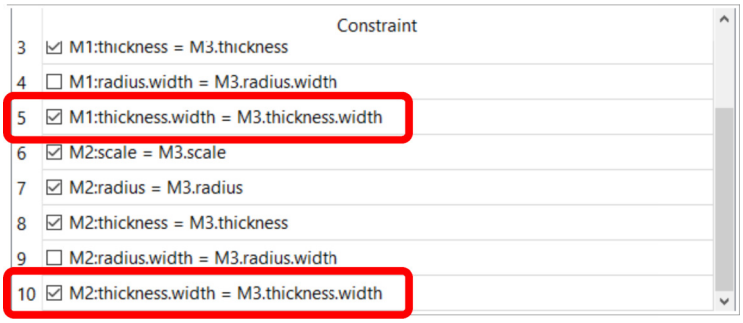
When the fit completes the *radius* distributions reappear. Check that they have reached a positive baseline to both sides of their peak (they should have) and then minimise them again.

In the Const & Simul Fit page uncheck the two *radius.width* constraints to deactivate them.



Then, in the Polydispersity tabs of each of the FitPages. uncheck the *Distribution of radius* parameters (but otherwise leave them untouched) and now set a *Distribution of thickness* parameter of 0.2 with a **gaussian** distribution and check it so it will optimise. Minimise any *thickness* distribution plots as they appear.

Return to the Const & Simul Fit page and this time create constraints to tell SasView that the *Distribution of thickness* parameters for the 'core' and 'drop' contrast datasets are the same as the *Distribution of thickness* parameter for the 'shell' contrast dataset.



Click Fit.

WARNING: This fit may take ~20 mins to complete!!!

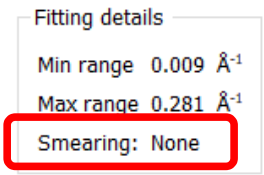
Polydispersity has now been added to each of the *radius* and *thickness* parameters and, **unlike in SasView 5.x or earlier**, each of those polydispersities was constrained to be the same for each dataset.

However, optimising the polydispersities slows the optimisation considerably, which is why we unchecked the *Distribution of radius* parameters above. Now that we also have estimates for the *Distribution of thickness* parameters we can disable them too.

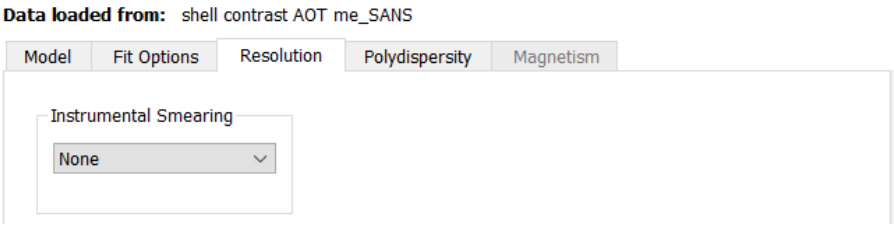
In the Const & Simul Fit page uncheck the two *thickness.width* constraints to deactivate those constraints. Then, in the Polydispersity tabs of each of the FitPages. uncheck the *Distribution of thickness* parameters (but otherwise leave them untouched).

The fits are clearly improving but could be improved further.

None of the three datasets contain any instrumental resolution information. This is actually flagged at the bottom of each FitPage

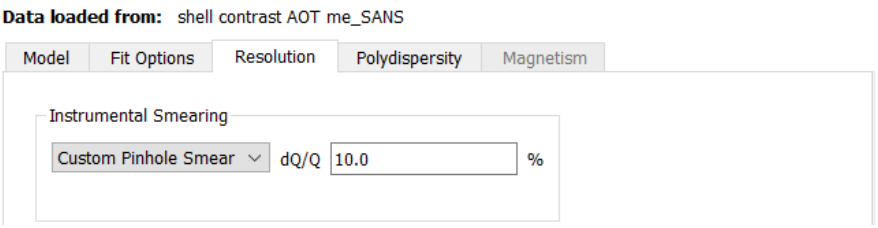


and in the accompanying *Resolution* tab



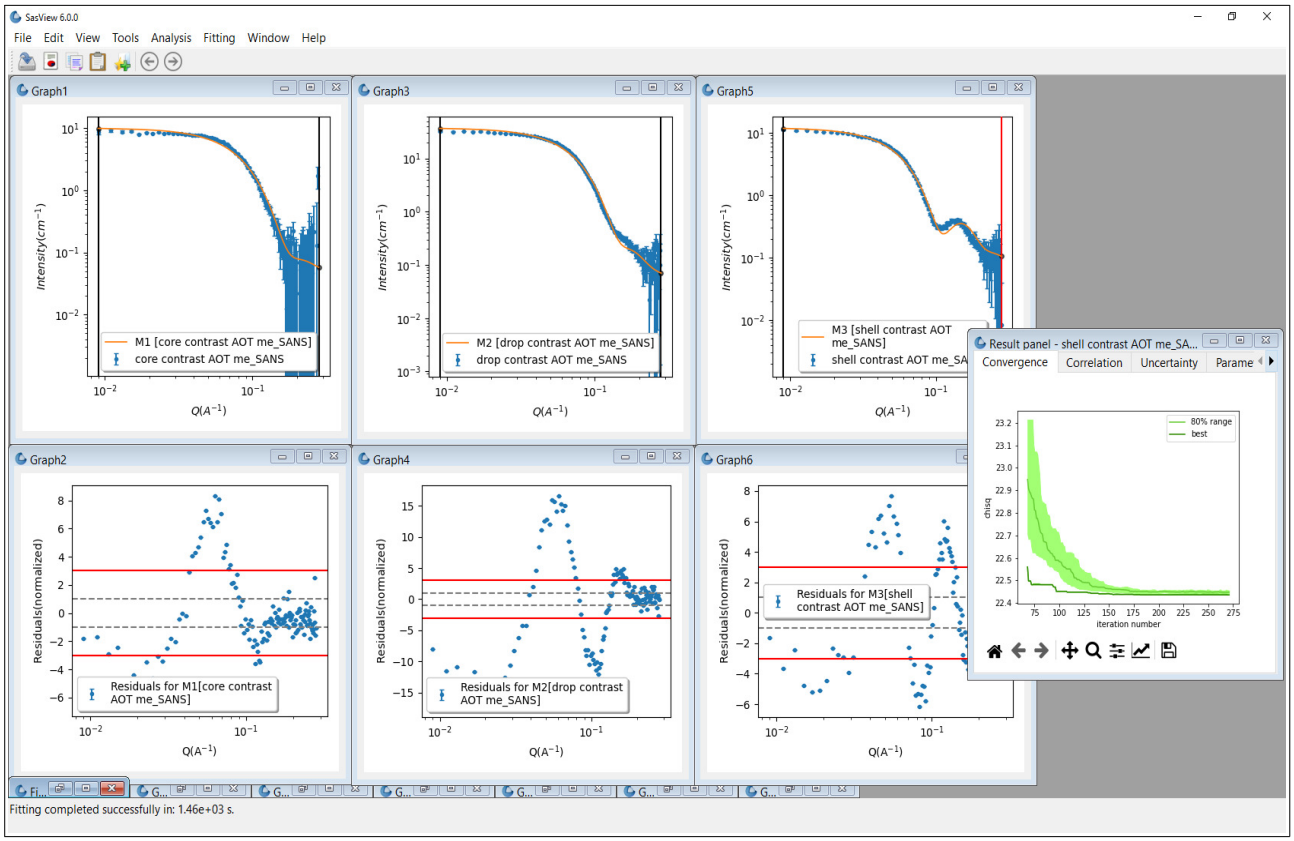
However, these data will have been subject to smearing by instrumental resolution factors. In such cases, SasView can estimate and apply a generic compensation. The test data help documentation (https://www.sasview.org/docs/user/sasview/test/testdata_help.html) tells us that the data were collected on the LOQ instrument at the ISIS Neutron Source. This is a classic pinhole-collimated SANS instrument.

Go to the *Resolution* tab in each FitPage, select *Custom Pinhole Smear* from the dropdown and set dQ/Q = 10% (this will likely under-estimate the resolution at low Q-values and over-estimate it at higher Q-values).



Return to the Const & Simul Fit page click Fit.

This fit could take even longer to complete than the preceding one, so pay attention to the parameter values in the *Log Explorer.* If they have stopped changing, click *Stop fit*. Note that you will need to let the fit run for at least a few minutes to see any change!



The resulting solution suggests the microemulsion droplets have a water core of *radius* ~23.5 +/- 0.02 Å (with ~12% Log-Normal polydispersity) and a surfactant shell layer ~7.3 +/- 0.01 Å thick (with ~42% Gaussian polydispersity).

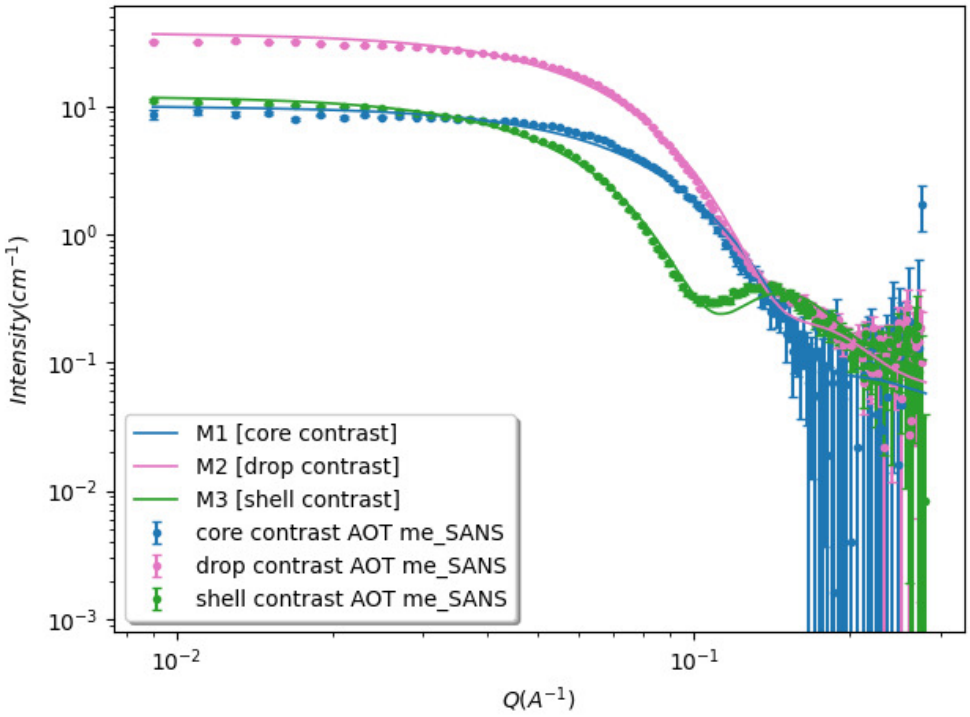
These values are consistent with literature reports on similar systems, for example, https://doi.org/10.1016/0921-4526(92)90822-A and https://doi.org/10.1021/la00035a016, but the actual droplet sizes do vary with the water/oil ratio. However, as can be seen, the fitting is still far from perfect.

Can you improve it further?!

Summary of Fitting Results

	'core'	'drop'	'shell'
X^2	6.96	49.9	10.1
scale	0.056 ± 0.0001	0.056 ± 0.0001	0.056 ± 0.0001
background (cm^{-1})	0.046 ± 0.012	0.051 ± 0.006	0.060 ± 0.003
radius (Å)	23.5 ± 0.02	23.5 ± 0.02	23.5 ± 0.02
radius distribution [a]	0.13	0.13	0.13
thickness (Å)	7.3 ± 0.01	7.3 ± 0.01	7.3 ± 0.01
thickness distribution [b]	0.42	0.42	0.42

Notes: a) LogNormal distribution; b) Gaussian distribution



Further Information

For further information, please consult the

SasView Tutorial Series

or

http://www.sasview.org

or email

help@sasview.org