±
3

SasView Tutorials

Modifying Weights 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

Miguel Gonzalez (gonzalezm@ill.fr)
Steve King (stephen.king@stfc.ac.uk)

Last revised: 22 August 2024

Acknowledgements

The option to modify the weights of each data set was originally coded by Iestyn Cadwallader-Jones, during an internship at the ILL in 2021. Useful discussions about the limitations and the dangers of this approach with Paul Kienzle, Lionel Porcar, Paul Butler, Wojciech Potrzebowski, Stephen King, Najet Mahmoudi, Sylvain Prevost, and Tim Snow are gratefully acknowledged.

As a pre-requisite to this tutorial is is assumed that the reader is familiar with using SasView to simultaneously fit datasets as covered in the tutorial ***simultaneous_1d_fitting_in_sasview***.

The program interface shown in this tutorial is a pre-release development version of 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.

The functionality described in this tutorial is not available in earlier versions of SasView.

Learning Objectives

This tutorial shows how to modify the weights of the data sets when performing a simultaneous fit in order to “enhance” or “decrease” the influence of one or several of them and “drive” the fit to a desired answer.

Warning

Ideally, this option **should never be used!** The fitting algorithm will use all the data points, and even if a particular data set has more points or smaller errors and therefore has a larger influence in determining the total χ^2, the information provided by the additional data sets should be enough to allow the fitting algorithm to find the best solution. The problem arises when there is not a unique solution that works equally well for all data sets. This could be an indication of systematic errors, ill-defined resolution functions, problems with sample preparation, etc. In this case, if one of the sets dominates the fitting (usually because it has a much larger number of points and/or much smaller relative errors, which happens often when combining neutron and x-ray data), the solution found will often be close to the optimal solution for that set, while the other data sets are mostly ignored. In such cases, it is recommended to check first for all possible causes of discrepancies between the data sets and ensure that all the sets to be treated simultaneously can be fitted with a unique model and a compatible set of parameters. Only then, if one has good reasons to think that the influence of a given set should be enhanced in order to provide additional physical insight for one parameter, should the option to modify the weights be used with the needed care.

To do

It would be great to adding an example using real experimental data. But it should be relatively simple and instructive. Any ideas?

Data sets

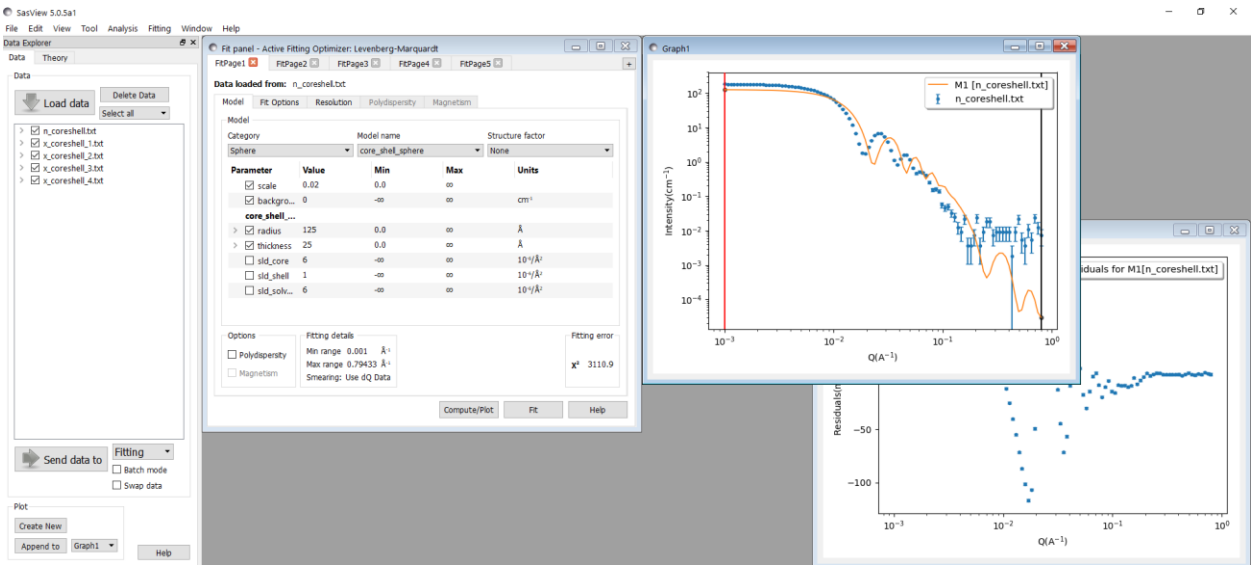
In order to demonstrate how to use this option, we will use five sets of synthetic data, corresponding to a core-shell model and generated under well-controlled conditions. The parameters employed to generate the 5 data sets are given in the following table:

	M1	M2	M3	M4	M5
Parameter	**n_coreshell.txt**	**x_coreshell_1.txt**	**x_coreshell_2.txt**	**x_coreshell_3.txt**	**x_coreshell_4.txt**
scale	0.010	0.010	0.012	0.010	0.012
background	0.01	0.001	0.001	0.001	0.001
radius	150.0	150.0	150.0	165.0	165.0
thickness	40.0	40.0	40.0	40.0	40.0
sld_core	6.0	4.0	4.0	4.0	4.0
sld_shell	1.0	4.0	4.0	4.0	4.0
sld_solvent	6.0	6.0	6.0	6.0	6.0

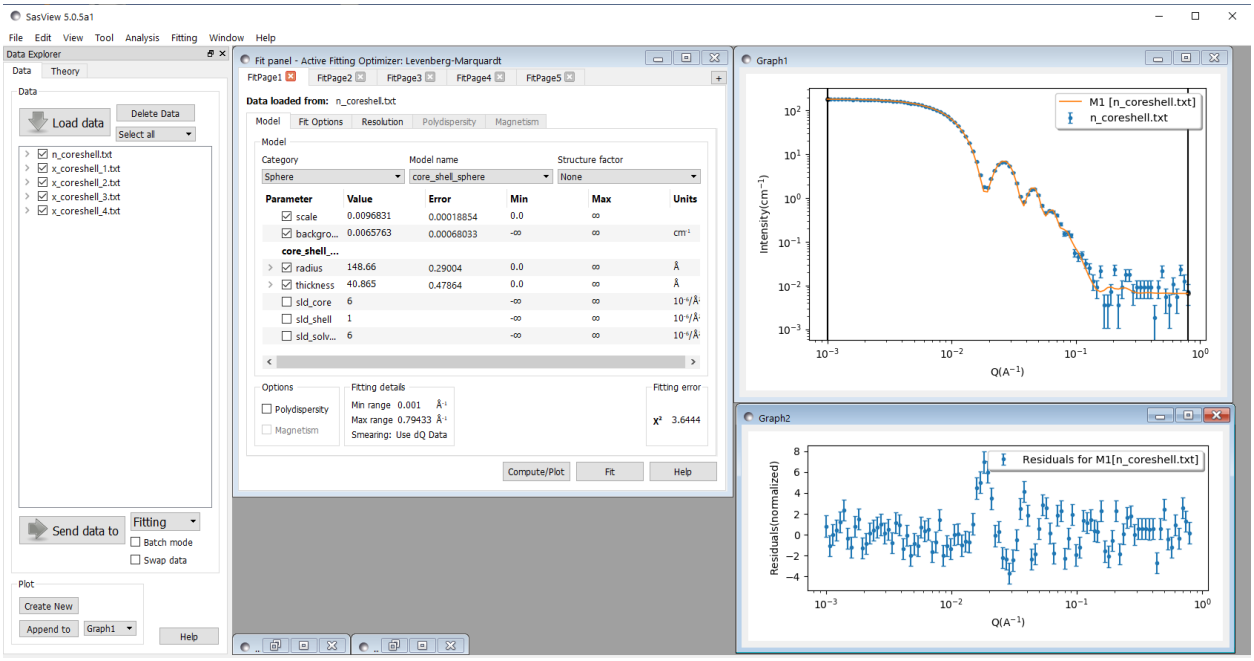
The first of those sets (M1) is a "neutron-like" data set, while M2-M5 are considered "x-ray" models. M1 differs from M2-M5 in the number of points (100 *vs* 1000), the Q-resolution ($\Delta Q/Q \approx 7\%$ *vs* 0.7% for $Q > 0.1$ Å^{-1}), and the error bars (mean relative error of 15 % *vs* 1.2 %). However, the "x-ray" models are completely insensible to the thickness of the core-shell particles, as the scattering length densities of the core and the shell are equal. We see that M1+M2 corresponds to the ideal case, where both sets of data have been generated with exactly the same sample parameters (volume fraction (scale), core radius, and shell thickness). But M3, M4 and M5 have been created using a different set of parameters (either for the scale or the radius or both), implying that no model will work well when trying to fit simultaneously M1 (or M2) together with any of them.

Individual fits

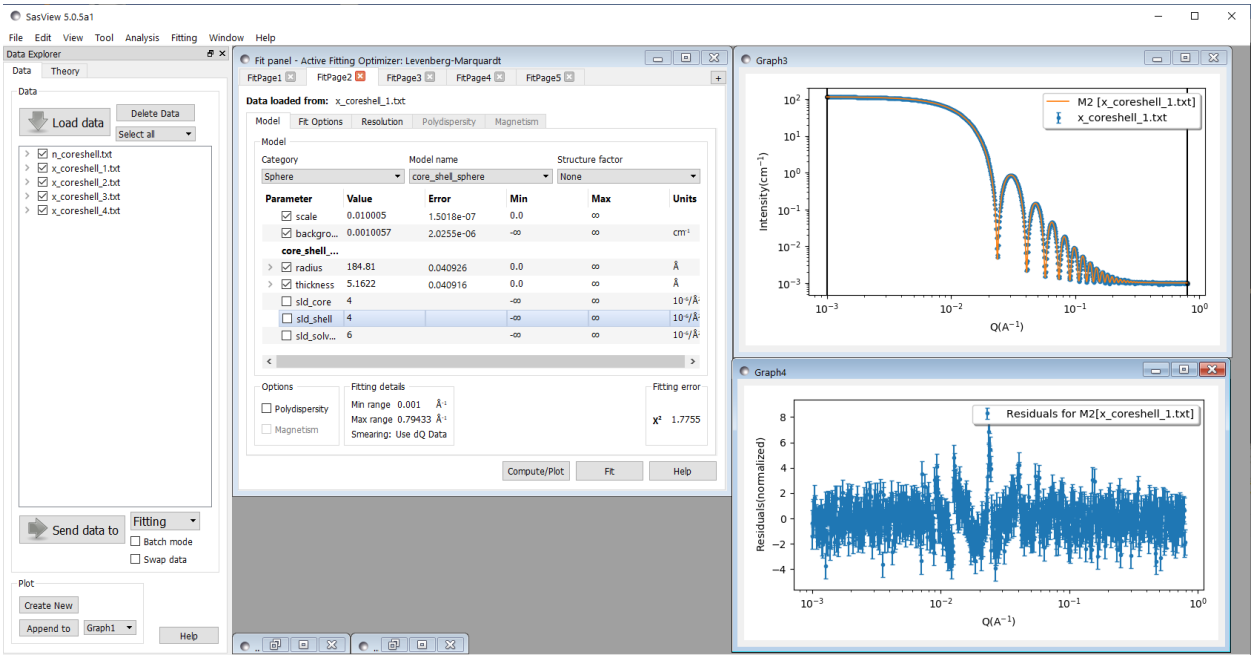
We will start by fitting each data set individually. You can consult the tutorial "Basic 1D Fitting in SasView" if you need additional instructions. We will assume that the scattering lengths of the particles (core and shell) and the solvent are well known, and that we are interesting in determining the volume fraction (scale parameter) and the dimension of the particles (radius and thickness). The initial FitPage for the neutron data set (M1) would look like something like this:



And after fitting the data (using the default options, i.e. Levenberg-Marquardt algorithm with 200 steps and a tolerance of 1.5×10^{-8}), we get a reasonably good fit, with a reduced χ^2 of 3.6 and final parameters close to the expected results: ϕ = 0.0097 ± 0.0002, radius = 148.7 ± 0.3 Å and thickness = 40.9 ± 0.5 Å:



The fits obtained for the "x-ray" data sets (M2-M5) are even better, with reduced $\chi^2 \approx$ 1.8-1.9. As expected, we are not able to get the real radius and thickness, but we obtain the correct value for the sum of the radius and thickness.



The set of parameters obtained when fitting each of the 5 data sets individually are summarized in the following table[1]:

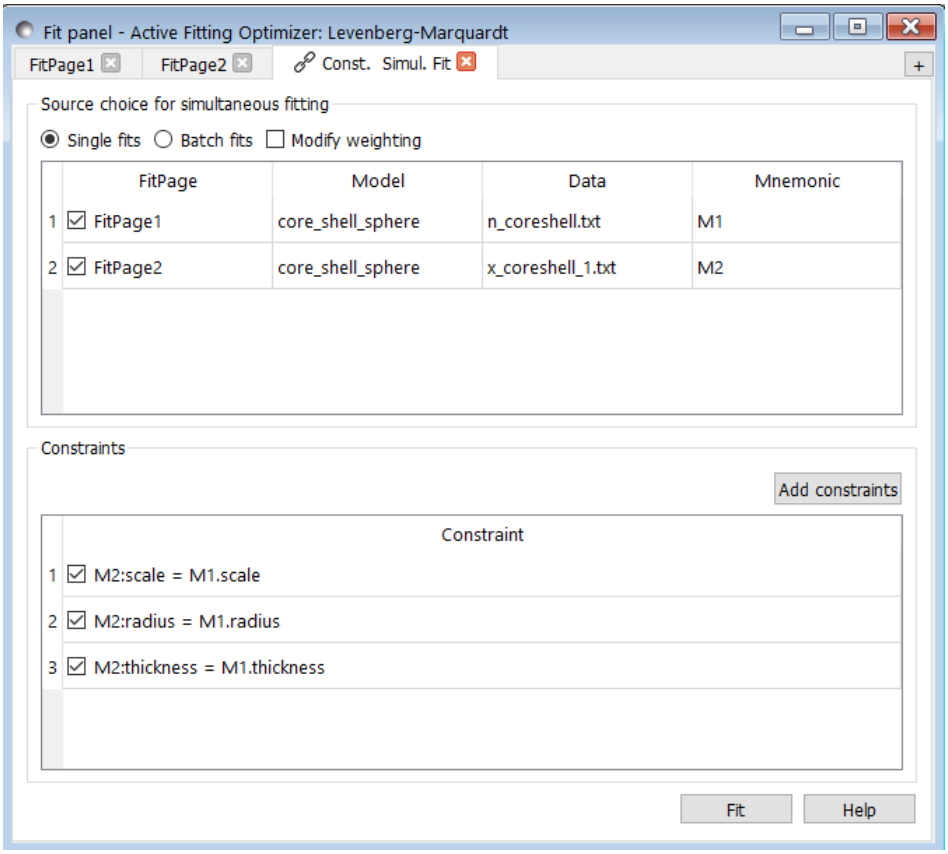
Parameter	M1	M2	M3	M4	M5
scale	0.0097 ± 0.0002	0.0100 ± 2×10^{-7}	0.0120 ± 2×10^{-7}	0.0100 ± 2×10^{-7}	0.0120 ± 2×10^{-7}
background	0.0066 ± 0.0007	0.0010 ± 2×10^{-6}	0.0010 ± 2×10^{-6}	0.0010 ± 2×10^{-6}	0.0010 ± 3×10^{-6}
radius	148.7 ± 0.3	184.81 ± 0.04	176.90 ± 0.09	154.4 ± 0.6	200.78 ± 0.03
thickness	40.9 ± 0.5	5.16 ± 0.04	13.07 ± 0.9	50.6 ± 0.6	4.19 ± 0.03
R + thickness	189.5 ± 0.6	189.97 ± 0.06	189.97 ± 0.12	205.0 ± 0.9	204.97 ± 0.05
χ^2	3.6	1.8	1.8	1.9	1.9

Simultaneous fit

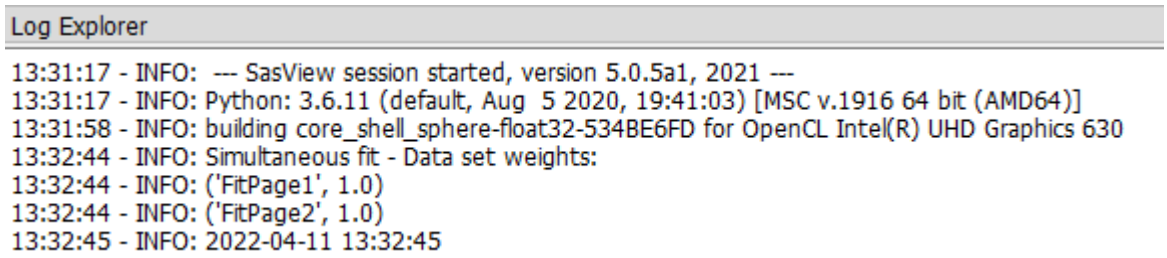
Now let's see what happens when we fit two data sets simultaneously, using the Constrained or Simultaneous Fit option in the Fitting menu. Again, if needed detailed instructions are given in the tutorial "Simultaneous 1D Fitting in SasView".

[1] We used the default Levenberg-Marquardt algorithm, which implies that the algorithm will find the local minimum closest to the initial guess. Therefore, you could get different results depending on the initial value of the parameters. However, in most cases you should get a set of results very close to those given in the table.

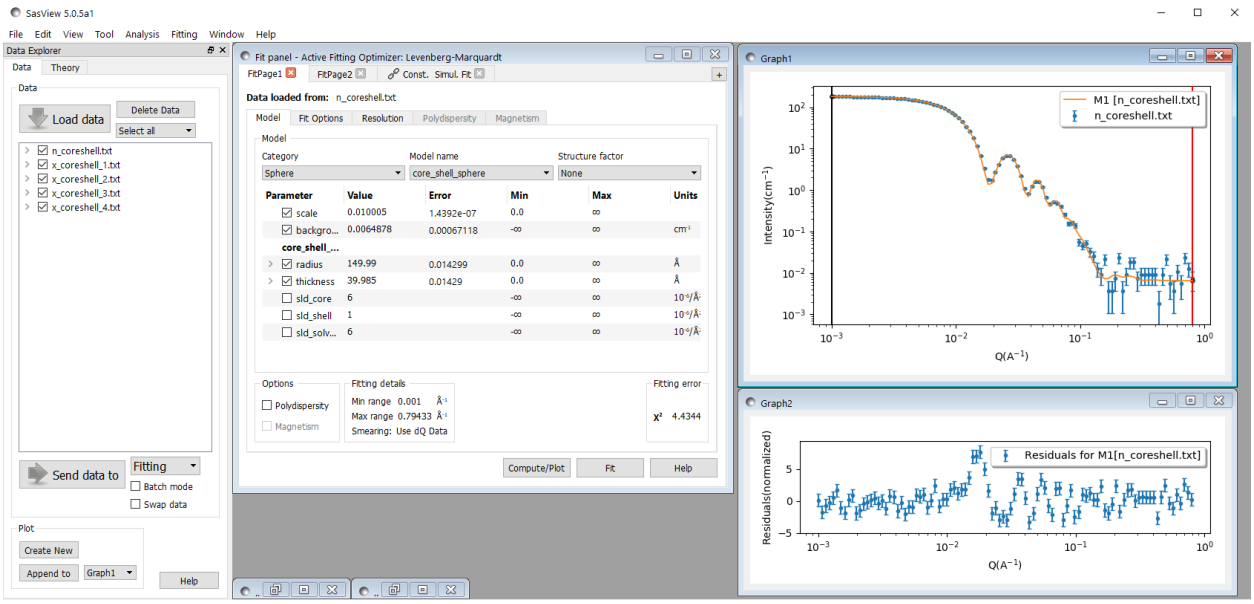
We will start fitting simultaneously the sets M1 and M2, assuming that the sample is the same and therefore constraining the volume fraction, the radius and the thickness to be the same for both data sets. We define the three needed constraints in the corresponding tab:

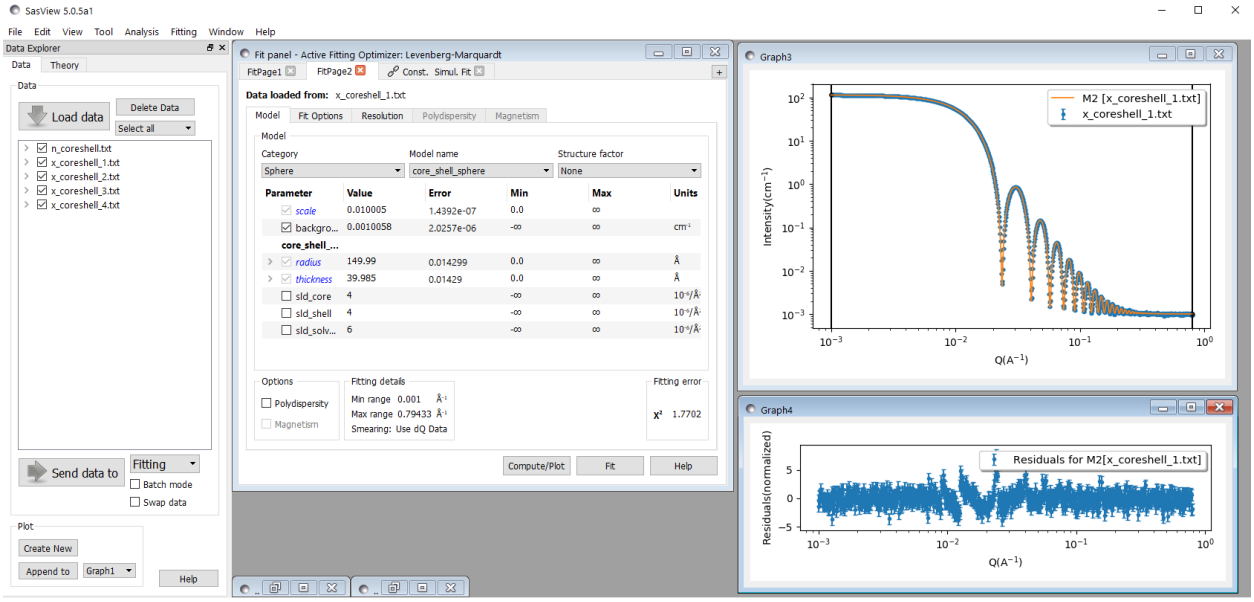


And we run the simultaneous fit by pressing the Fit button in this tab. Information about the specific weight attributed to each data set (by default all the sets have a weight of one) is given in the log explorer window:



We see that in this ideal case we obtain good fits for both data sets:





For M2, the reduced χ^2 is the same as the one obtained when fitting the set individually, while for M1 is also comparable to that reference value (= 4.4 *vs* 3.6 before). And more importantly, even if M1 has a much smaller influence in the global fit because the lesser number of point and larger error bars, the additional information provided by this data set is enough to drive the simultaneous fit to the correct answer for the radius and the thickness of our core-shell particles.

Simultaneous fit of non-compatible data sets

A different situation arises when the data sets that we fit simultaneously are not "compatible", meaning that the optimal solutions for each of them are different. This is the case, for example, if we try to fit simultaneously M1 + M2 or M2 + M3, as M3 has been generated using a different volume fraction. In this case it is obvious that we cannot get a unique set of parameters that is adequate for both sets. Such situation could arrive if we have used two different samples and we have not been careful enough to ensure that they were exactly the same. Unfortunately, other systematic errors will have similar consequences and they are much harder to correct for or even to determine. What solution will we obtain in these cases?

The following table shows the results obtained by fitting simultaneously M1 and each of the "x-ray" sets (M2-M5):

Parameter	M1+M2	M1 + M3	M1 + M4	M1 + M5
scale	$0.0100 \pm 1\times10^{-7}$	$0.0120 \pm 2\times10^{-7}$	$0.0100 \pm 1\times10^{-7}$	$0.0120 \pm 2\times10^{-7}$
background	0.0065 / 0.0010	0.0055 / 0.0010	0.0068 / 0.0010	0.0062 / 0.0010
radius	149.99 ± 0.01	154.24 ± 0.01	166.67 ± 0.01	170.67 ± 0.01
thickness	39.98 ± 0.01	35.74 ± 0.01	38.30 ± 0.01	34.29 ± 0.01
R + thickness	189.97 ± 0.02	189.98 ± 0.02	204.97 ± 0.02	204.96 ± 0.02
χ^2	4.4 / 1.8	12.3 / 1.8	244 / 1.9	294 / 1.9

As expected, as soon as it is not possible to get an unique optimal solution for both sets, the "heavier" sets dominates the fit and the information contained in M1 is almost fully neglected. We can see how the reduced χ^2 increases considerably for M1, while for M3-M5 it remains close to the value obtained in the individual fits, and how the scale and/or radius+thickness values obtained are the optimal ones for the "x-ray" set and deviate considerably from M1. However, it is still interesting to notice that the information contained in M1 serves at least to drive the solution to a reasonable value for the shell thickness.

As a reference, it is also interesting to see what happens when we fit two incompatible data sets that have a similar statistical weight:

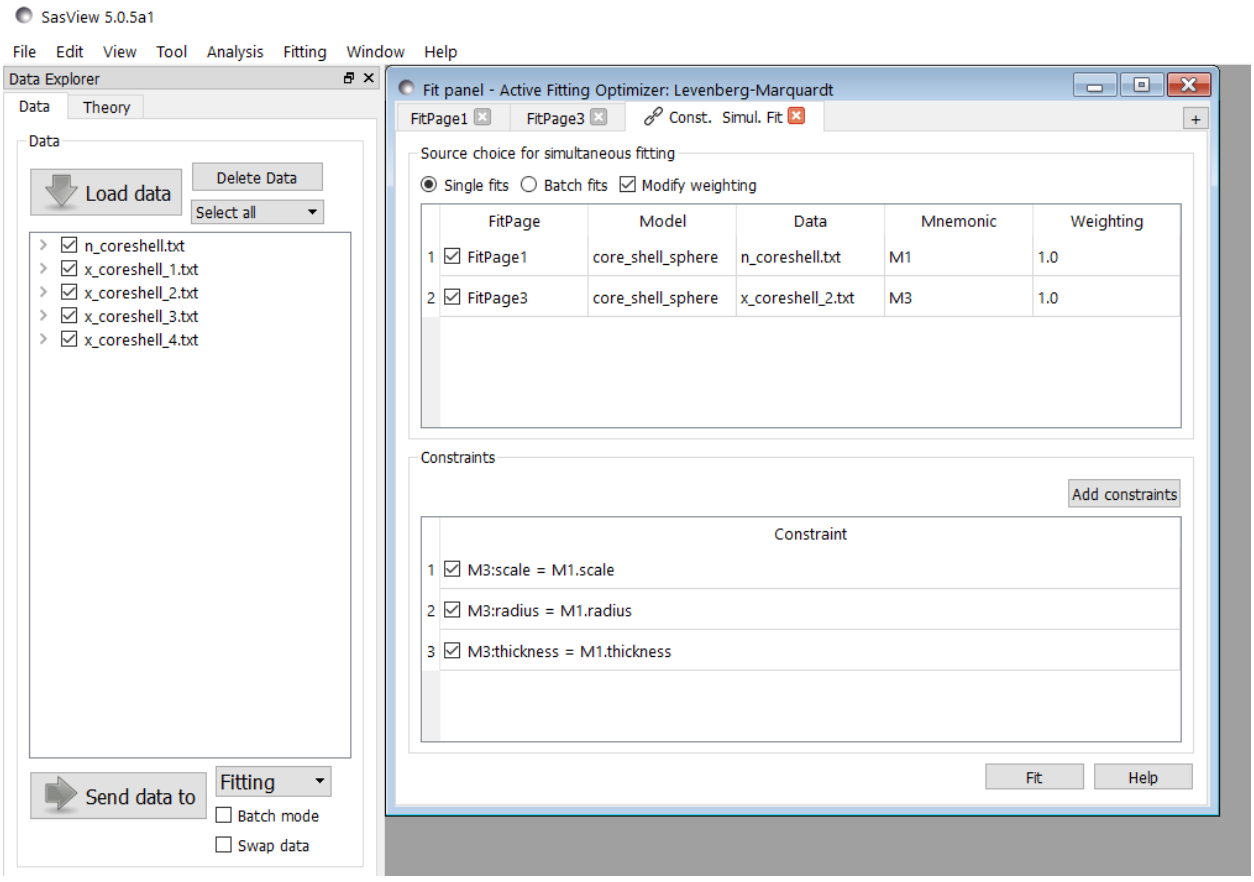
Parameter	M2 + M3	M2 + M4	M2 + M5
scale	$0.0108 \pm 1\times10^{-7}$	$0.0098 \pm 9\times10^{-8}$	$0.0105 \pm 1\times10^{-7}$
background	0.0008 / 0.0011	0.0004 / 0.0001	0.0004 / 0.00005
radius	152.2 ± 0.3	138.7 ± 0.4	158.6 ± 0.2
thickness	37.8 ± 0.3	58.7 ± 0.4	37.6 ± 0.2
R + thickness	190.0 ± 0.4	197.3 ± 0.5	196.3 ± 0.3
χ^2	2×10^5 / 3×10^5	3×10^5 / 4×10^5	7×10^5 / 2×10^6

We see that in this case none of the fits is good, as we obtain parameter values in between the two possible answers, so none of the two data sets is correctly reproduced.

As said at the beginning, the first thing to do in these cases is to try to understand the origin of such discrepancies and, if possible, correct for them. Another possibility consists in fitting individually each data set, either i) iterating over each data set and using the results for one set as the initial guess for the following one, and/or ii) introducing soft-constraints or reasonable parameter limits between the models. For example, in our M1+M3 example we could impose that the thickness parameter should be 40 ± 5 Å, based on our result for M1, and then refine the scale factor using M3 or M1, depending on our confidence in each data set.

Simultaneous fit using modified weights

However, the procedure above can be quite time consuming and the "Modify weighting" option is a quick way to explore what happens if the weight of the dominant sets is artificially decreased. After checking the corresponding button, the tab interface will change to show an additional column named Weighting:



The default value for all the data sets is 1, meaning that the weight of all data sets will be scaled down using the "lightest" set as a reference, in order to try to get a good balance where each data set has a similar importance in determining the final solution.

How is this done? Note that there is not a clear way of determining exactly the role of each data set in the total fit, so this is done in an empirical way and probably you will need to adjust the weights manually to get the desired balance.

First, a single weight per data set is computed as:

$$W_D = \sum_{i=1}^{N_p} \frac{1}{RE_i^2},$$

where N_p is the number of points in dataset D and RE_i is the relative error of point i, defined as $RE_i = \frac{\sigma_i}{I_i}$. We see that data sets with more points or smaller relative errors will have larger weights, so the appropriate weight for each data set is calculated using the minimum of those weights, i.e.:

$$W'_D = U_D \sqrt{\frac{\min(W_D)}{W_D}}$$

where U_D is the weighting given by the user in the interface.

The weights W'_D are then used by Bumps to compute the residuals as:

$$R = \sum_D W'_D \left[\sum_i^D \left(\frac{I_i - M_i}{\sigma_i} \right) \right]$$

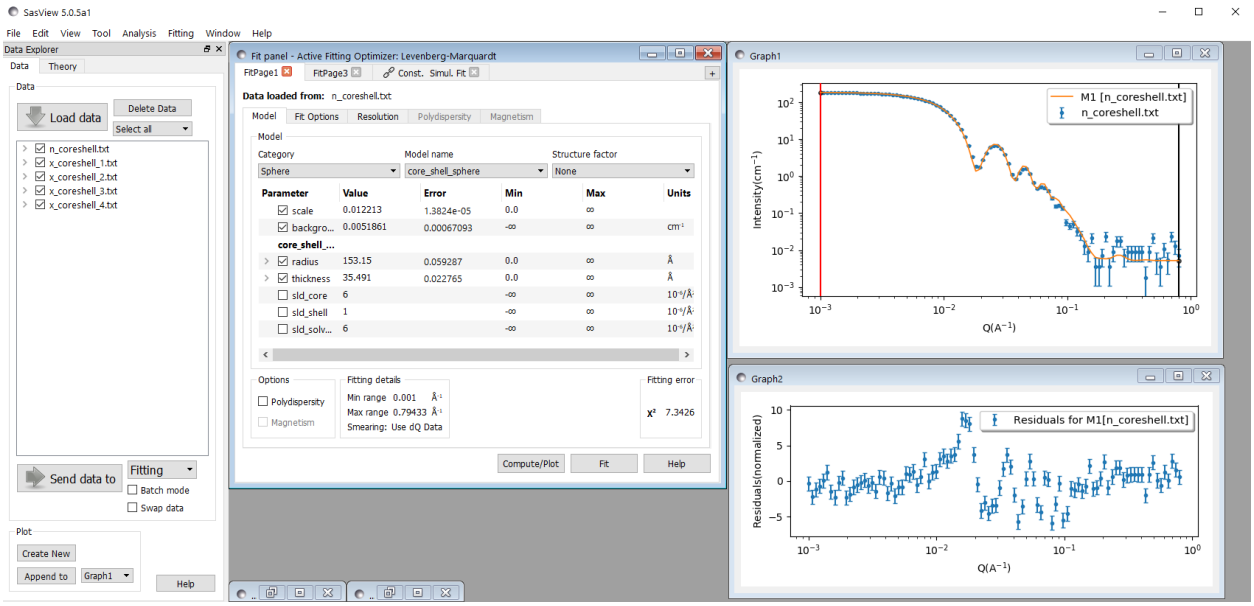
where M_i is the modelled intensity for each data point and the minimized quantity is R^2.

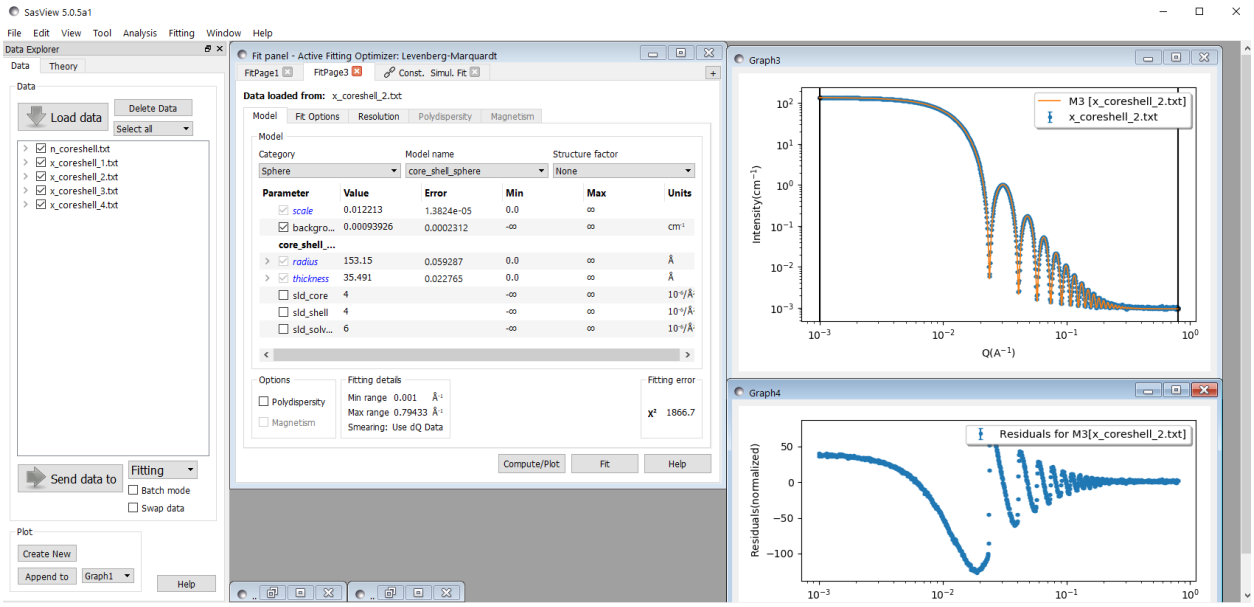
So if we repeat now the fit for the sets M1 + M3, we see that the weight of the M3 set is reduced by a factor of about 100, i.e. roughly a factor of 10 due to the difference in the number of points in M1 and M3, and another factor of 10 due to the difference in their relative errors.

Log Explorer

```
17:20:26 - INFO:  --- SasView session started, version 5.0.5a1, 2021 ---
17:20:26 - INFO: Python: 3.6.11 (default, Aug  5 2020, 19:41:03) [MSC v.1916 64 bit (AMD64)]
17:20:48 - INFO: building core_shell_sphere-float32-534BE6FD for OpenCL Intel(R) UHD Graphics 630
17:21:04 - INFO: Simultaneous fit - Data set weights:
17:21:04 - INFO: ('FitPage1', 1.0)
17:21:04 - INFO: ('FitPage3', 0.009713959263473316)
17:21:05 - INFO: 2022-04-11 17:21:05
```

And if we compare the fits and reduced χ^2 with those obtained before, we observe that M1 has gained some influence in the fit (χ^2 = 12.3 →7.3) at the expense of M3 (χ^2 = 1.8 →1867):





If needed, it is then possible to adjust the user weights to reduce or increase further the influence of a given set. For example, reducing further the weight of M3 we get results that approach more and more to those of M1 fitted alone:

Parameter	M1+M3	M1+M3	M1+M3	M1+M3	M1+M3
User weights	None	1 / 1	1 / 0.1	1/0.03	1 / 0.01
Applied weights	1 / 1	1 / 0.0097	1 / 0.00097	1 / 0.00029	$1 / 9.7\times10^{-5}$
scale	$0.0120 \pm 2\times10^{-7}$	$0.0122 \pm 1\times10^{-5}$	$0.0122 \pm 8\times10^{-5}$	0.0107 ± 0.0002	0.0098 ± 0.0002
background	0.0055 / 0.0010	0.0052 / 0.0012	0.0050 / 0.0008	0.0059 / 0.0011	0.0065 / 0.0013
radius	154.24 ± 0.01	153.15 ± 0.06	151.7 ± 0.1	150.1 ± 0.2	148.9 ± 0.3
thickness	35.74 ± 0.01	35.49 ± 0.02	35.7 ± 0.1	38.4 ± 0.4	40.5 ± 0.5
R + thickness	189.98 ± 0.02	188.64 ± 0.06	187.4 ± 0.2	188.5 ± 0.4	189.4 ± 0.5
χ^2	12.3 / 1.8	7.3 / 1867	5.0 / 21994	$3.9 / 5\times10^{5}$	$3.7 / 1\times10^{6}$

Using the sets M1 and M4 with different user weights, we get:

Parameter	M1+M4	M1+M4	M1+M4	M1+M4	M1 + M4
User weights	None	1 / 1	1 / 3	1 / 5	1 / 0.1
Applied weights	1 / 1	1 / 0.0099	1 / 0.03	1 / 0.049	1 / 0.00099
scale	$0.0100 \pm 1\times10^{-7}$	$0.0109 \pm 1\times10^{-5}$	$0.0102 \pm 5\times10^{-6}$	$0.0101 \pm 3\times10^{-6}$	$0.0118 \pm 8\times10^{-5}$
background	0.0068 / 0.0010	0.0060 / 0.0002	0.0067 / 0.0009	0.0069 / 0.0010	0.0053 / −0.003
radius	166.67 ± 0.01	160.92 ± 0.06	165.51 ± 0.03	166.22 ± 0.02	151.8 ± 0.1
thickness	38.30 ± 0.01	37.16 ± 0.02	38.06 ± 0.02	38.21 ± 0.01	36.3 ± 0.1
R + thickness	204.97 ± 0.02	198.08 ± 0.06	203.57 ± 0.03	204.43 ± 0.03	188.0 ± 0.2
χ^2	244 / 1.9	106 / 51107	218 / 1754	238 / 255	$4.9 / 6\times10^{5}$

And with M1 and M5:

Parameter	M1+M5	M1+M5	M1+M5	M1+M5
User weights	None	1 / 1	1 / 0.5	1 / 2
Applied weights	1 / 1	1 / 0.0099	1 / 0.0049	1 / 0.020
scale	$0.0120 \pm 2\times10^{-7}$	$0.0131 \pm 1\times10^{-5}$	$0.0139 \pm 2\times10^{-5}$	$0.0124 \pm 8\times10^{-6}$
background	0.0062 / 0.0010	$0.0044 / -4\times10^{-5}$	0.0040 / −0.0019	0.0055 / 0.0008
radius	170.67 ± 0.01	164.39 ± 0.06	159.91 ± 0.07	168.12 ± 0.04
thickness	34.29 ± 0.01	33.21 ± 0.02	32.56 ± 0.02	33.84 ± 0.02
R + thickness	204.96 ± 0.02	197.60 ± 0.06	192.47 ± 0.08	201.96 ± 0.04
χ^2	304 / 1.9	133 / 59774	$52 / 2.2\times10^{5}$	228 / 8440

This concludes our example showing how the user weights can be employed to give more or less relative weight to one set over another, and therefore direct the fit in a desired direction. The variability of the results obtained shown in the tables above tells us that one should be very careful when modifying the weights, so our two final advices are:

1. If possible, avoid using this option.
2. If needed, use it to try to get a set of parameters that are reasonably consistent with all the available information (all the data sets) and then try to understand why the weight of some of the sets needs to be modified. Are the error uncertainties well estimated? Is the instrumental resolution well known? Are the samples on all data sets really equivalent?

Discarding the error bars

It is possible to modify also the way the uncertainties are employed using the Weighting box in the Fit Options tab for each page. For example, setting Weighting to None makes that statistical uncertainties are completely discarded and each data point has the same weight in the fitting, independently of its error bar. This weighting scheme can be combined with the Modify weighting one available in the Constrained or Simultaneous Fit tab. For example, in this particular example (Weighting = None), checking also the Modify weighting option will have the effect of "correcting" for the different number of data points in each data set.

Fitting again M1 + M3, but setting Weighting to None in both FitPages we obtain the following results when fitting both sets individually:

Parameter	M1	M3
scale	0.02 ± 0.02	$0.0120 \pm 3\times10^{-5}$
background	−0.04 ± 0.17	0.0008 ± 0.04
radius	156 ± 10	171 ± 12
thickness	28 ± 18	19 ± 12
R + thickness	184 ± 21	190 ± 16
χ^2	0.13	8×10^{-5}

We see that the individual fits are much less reliable if the information about the data uncertainties is neglected, resulting in very large uncertainties for the parameters obtained. In this case, we cannot expect a value of the reduced χ^2 close to 1 for a good fit, so the numerical values returned by the fit are not particularly helpful to tell us if the fit was successful or not.

And fitting them simultaneously:

Parameter	M1+M3	M1+M3
User weights	None	1 / 1
Applied weights	1 / 1	1 / 0.316
scale	$0.0121 \pm 3\times10^{-5}$	$0.0123 \pm 7\times10^{-5}$
background	0.192 / −0.03	0.055 / −0.102
radius	154.0 ± 0.1	153.0 ± 0.3
thickness	35.57 ± 0.04	35.35 ± 0.09
R + thickness	189.6 ± 0.2	188.4 ± 0.3
χ^2	0.38 / 0.004	0.17 / 0.08

In this case, the results obtained with and without the Modify weighting option are very similar and we also see that the weight applied to the set M3 is $\sqrt{N_{M1}/N_{M3}}$, as expected if only the number of points is used to calculate the data set weight.

An example with real data

Finally we can test this option on real data using the three sets already employed in the tutorial "Simultaneous 1D Fitting in SasView"[2]:

M1 = core contrast AOT me_SANS

M2 = drop contrast AOT me_SANS

M3 = shell contrast AOT me_SANS

As before, we start by fitting each data set individually to have a reference of the optimal ()set of parameters for each of them:

Parameter	M1	M2	M3
scale	0.029 ± 0.003	0.0506 ± 0.0002	0.0164 ± 0.0003
background	0.03 ± 0.01	0.114 ± 0.003	0.177 ± 0.004
radius	26.7 ± 0.2	29.8 ± 0.2	15.9 ± 3
thickness	0.5 ± 1.1	1.5 ± 0.2	21.4 ± 0.4
χ^2	4.3	35.6	10.5

It should be noted that those are the parameters that minimize χ^2 for each data set, but not necessarily the true parameters, as different contrasts will be or less sensitive to the radius and shell thickness of the particles. Fitting the 3 sets simultaneously we get:

Parameter	M1+M2+M3	M1+M2+M3	M1+M2+M3
User weights	None	1/1/1	0.1/1/0.3
App weights	1/1/1	1/0.31/0.61	0.1/0.31/0.18
scale	0.05/0.05/0.08	0.06/0.06/0.06	0.04/0.05/0.10
background	0.10/0.14/0.08	0.06/0.16/0.08	0.16/0.13/0.08
radius	25.28 ± 0.08	24.3 ± 0.1	26.3 ± 0.6
thickness	6.71 ± 0.10	8.0 ± 0.2	5.6 ± 0.6
χ^2	6.4/35.7/16.3	5.4/37.0/17.3	8.5/35.6/15.3

[2] Consult the tutorial for a full description of the scientific problem and details on how to set and perform the simultaneous fit. Note also that the results presented here have been obtained using a fixed Gaussian polydispersity of 0.1 for both the radius and shell thickness, and the default Levenberg-Marquard algorithm, while previously the polydispersity was also adjusted and the more robust DREAM algorithm was used to fit the data, which explains why the results are different.