2021 ISIS Virtual Reflectivity Training Course

18th November 2021

An Introduction to NR Data Analysis Using RasCAL in IDAaaS

Introduction

During the following demonstration, we will introduce the RasCAL package for the analysis of reflectometry data and perform some fits to some simple data. As outlined in previous lectures, RasCAL uses 'slab' models to describe the interfacial structure under investigation, and calculates the theoretical reflectometry profile using Abeles matrix formalism (introduced in Lecture 1). This demonstration will be performed live, and all information given here will be covered during the session. This document provides a guide for you to reproduce the analysis yourself.

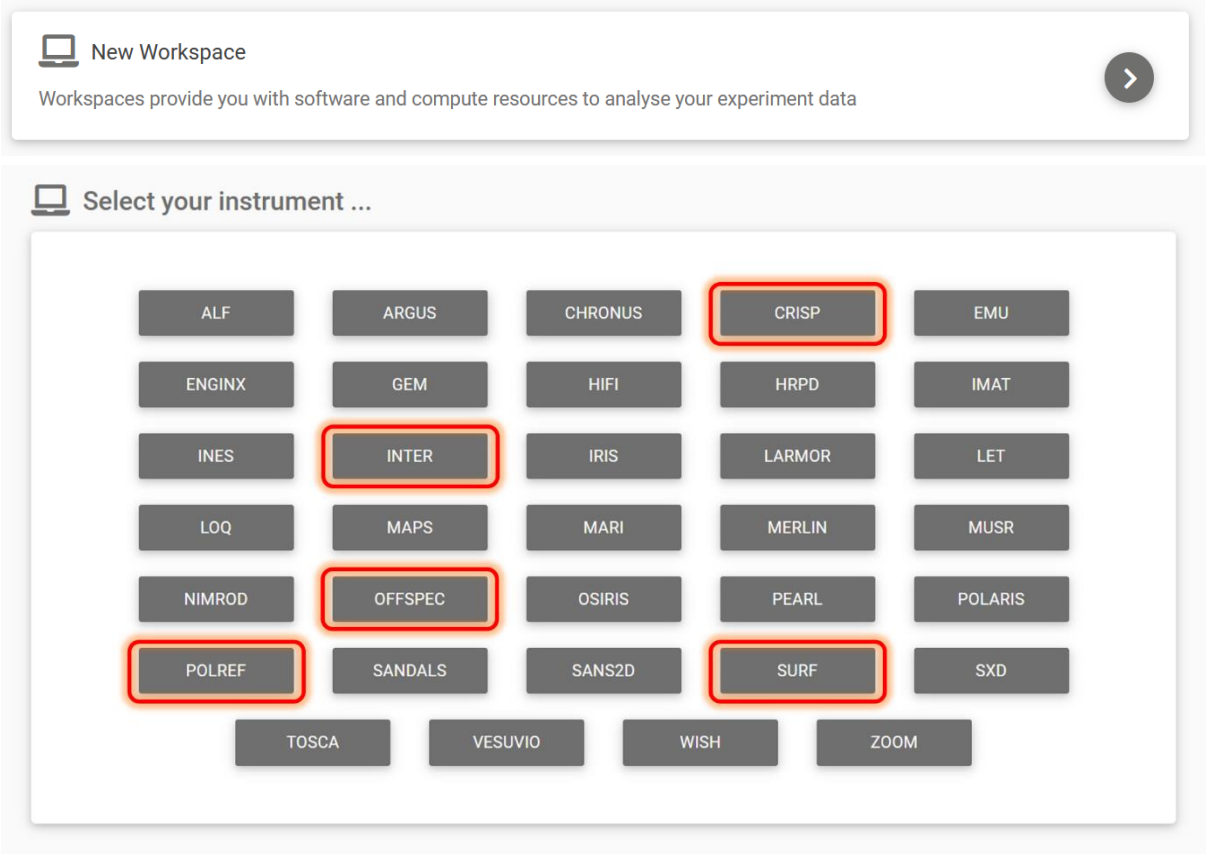
We advise you go through this guide in advance of the Lipid Bilayer Case Study to familiarize yourself with some of the core functionality of RasCAL.

Throughout this session you will:

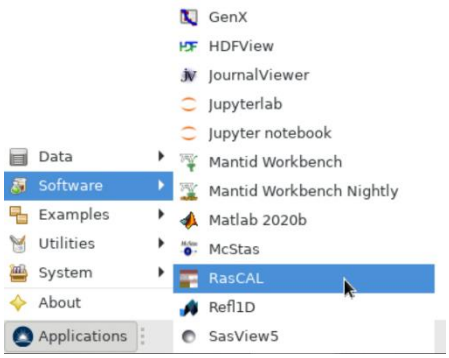
- Be introduced to RasCAL software, including navigation in the GUI, loading experimental data and building a simple model.
- Become familiar with the influence of each layer parameter on the observed reflectivity data.
- Fit NR data collected from a clean silicon substrate in D_2O.
- Perform Markov-Chain Monte-Carlo (MCMC) error analysis and interpret error analysis.
- Co-refine a model using two isotopic contrasts describing the same interfacial structure to improve confidence in fitted parameters.

Part 1: Introduction to RasCAL, loading data and building a model.

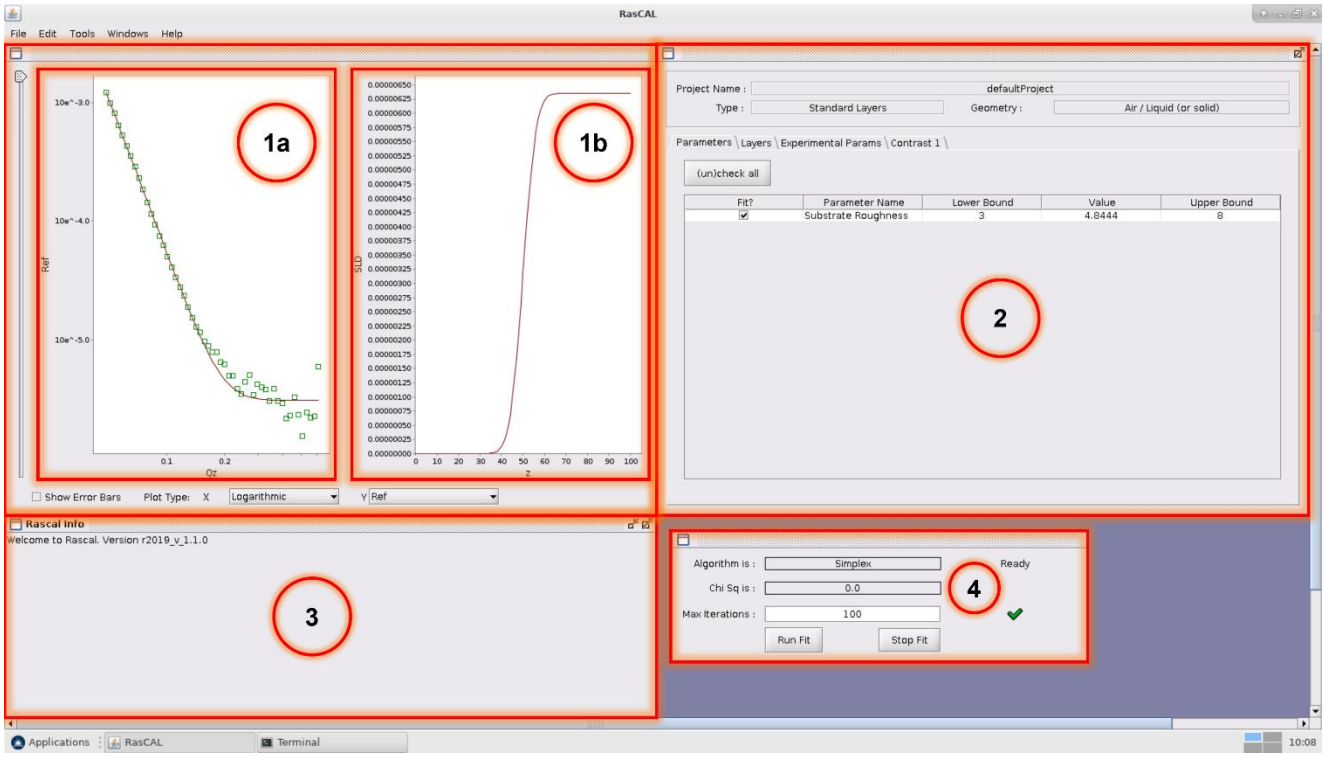
To launch RasCAL, navigate to IDAaaS within a web browser, log in and launch reflectometry workspace using any of the ISIS reflectometry instrument suite (INTER, POLREF, OFFSPEC, SURF, CRISP). If you don't already have an existing environment, create a New Workspace as shown below:



Once you have launched an IDAaaS session, open RasCAL, located within Applications > Software > RasCAL



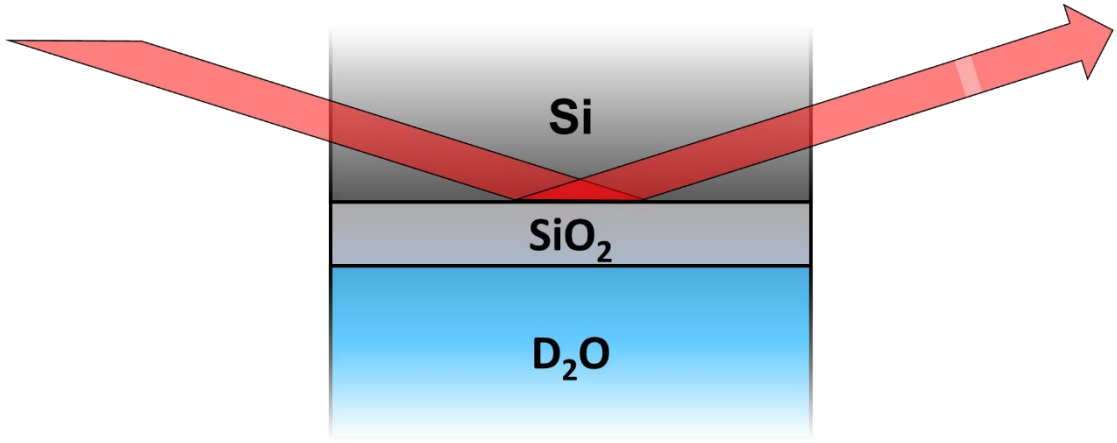
You will now be met with the main RasCAL main interface. Windows may be overlapping, so first select Windows > Tile to arrange windows sensibly for your display resolution. Here are the key components of the user interface:



1. Model inspection panel. This allows you to view:
    a. The reflectivity data and theoretical reflectivity profile described by your model.
    b. The SLD profile corresponding to your model, from which the theoretical reflectivity profile is calculated.
2. Model building panel. This allows you to change parameters associated with your model, as well as experimental parameters such as the Qz resolution and scale factors.
3. Information Panel. This will display information relating to the operation of RasCAL, such as fitting progression information and file operations (eg. Saving, loading files)
4. Fitting panel. Here, the fitting algorithm currently being used is displayed, along with the normalized Chi Squared parameter (ie, goodness of fit), the number of iterations to be performed, and buttons to start and stop fitting the data.

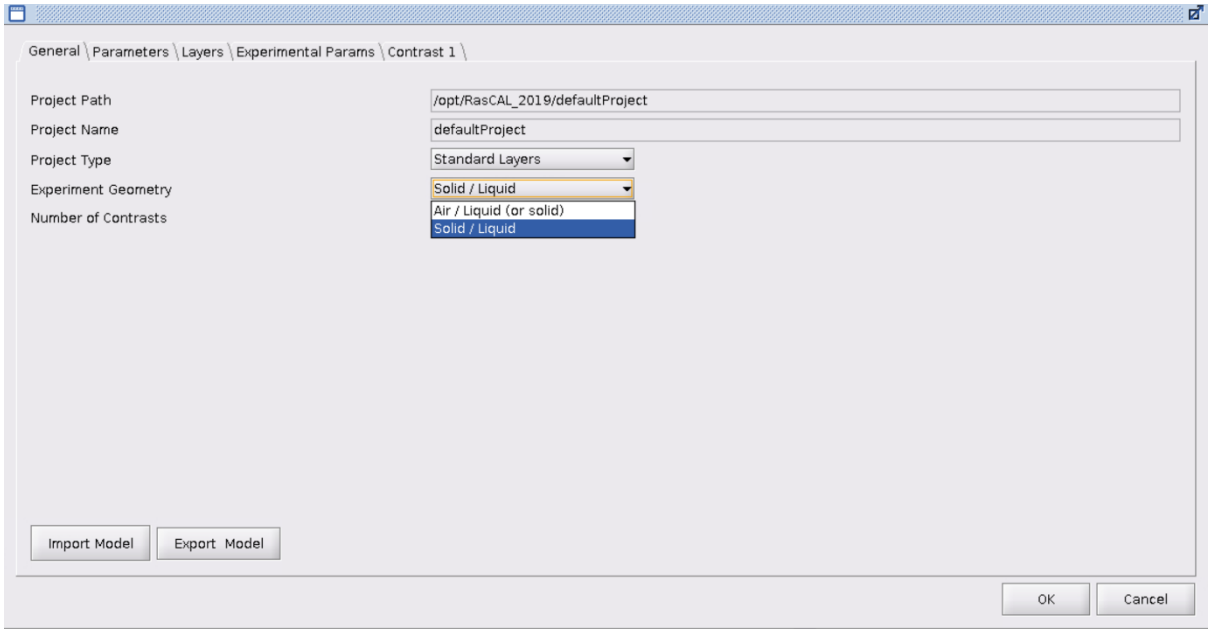
Here, we will build a model to describe a simple Silicon-water interface. Silicon is a common substrate for reflectometry experiments, and measurements of the clean substrate are frequently taken at the start of an experiment to understand the initial substrate structure and to improve the ambiguity in parameters required for downstream structural analysis.

Silicon will spontaneously form a thin amorphous oxide layer on the surface, which is permeable to bulk solvent. Our model will therefore require a single interfacial layer, corresponding to hydrated SiO_2, between two bulk phases, Si and water. In order to satisfy the conditions for total external reflection, the neutron beam must travel through the material with the lower refractive index (and therefore lower SLD) and reflect from the interface with the higher refractive index. Therefore, our model will have a 'bulk-in' (or superphase) of Si, and a 'bulk-out' (or subphase) of D_2O:

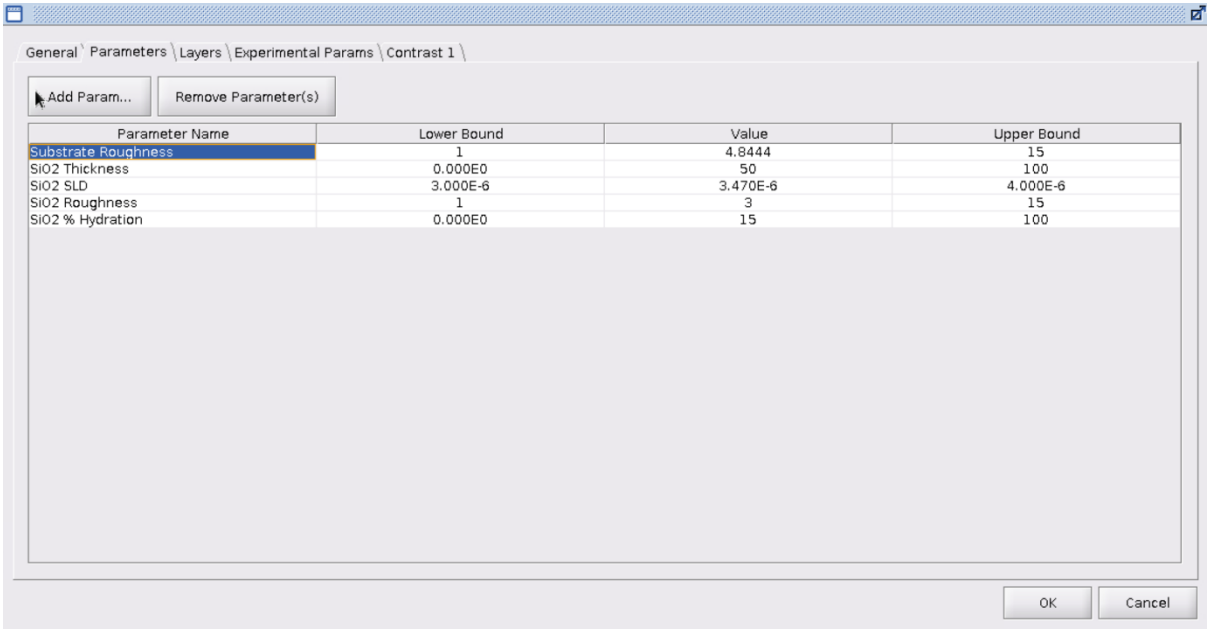


Building a model.

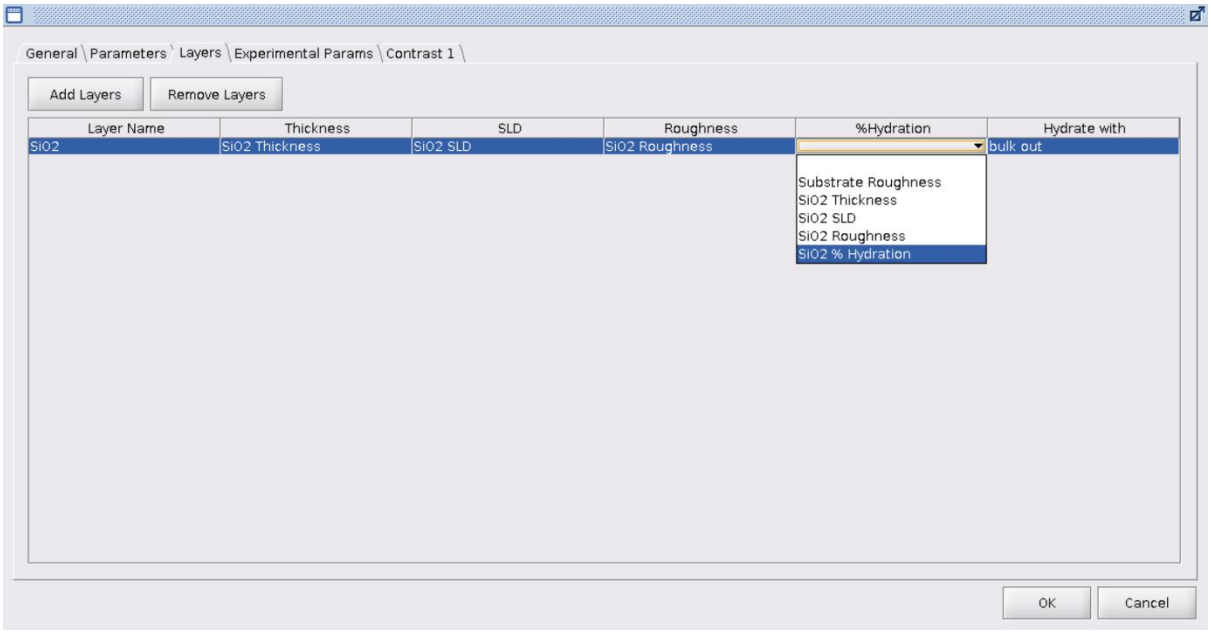
In the RasCAL window, select ‘Edit > Edit Project’. In the ‘General’ tab of the resultant window, change the Experiment Geometry to Solid / Liquid:



Next, we want to add the parameters required for your model. In RasCAL, a sample is defined as a series of slabs, flanked by bulk phases. Each slab layer is described by four parameters: Thickness, SLD, Roughness and Hydration. Here, we have a single interfacial layer corresponding to SiO_2, between bulk Si and solvent. ‘Substrate Roughness’ describes the roughness between bulk Si and the first interfacial layer, so should be retained as a parameter. Four additional parameters should be added, and given initial values, upper and lower values as shown below.

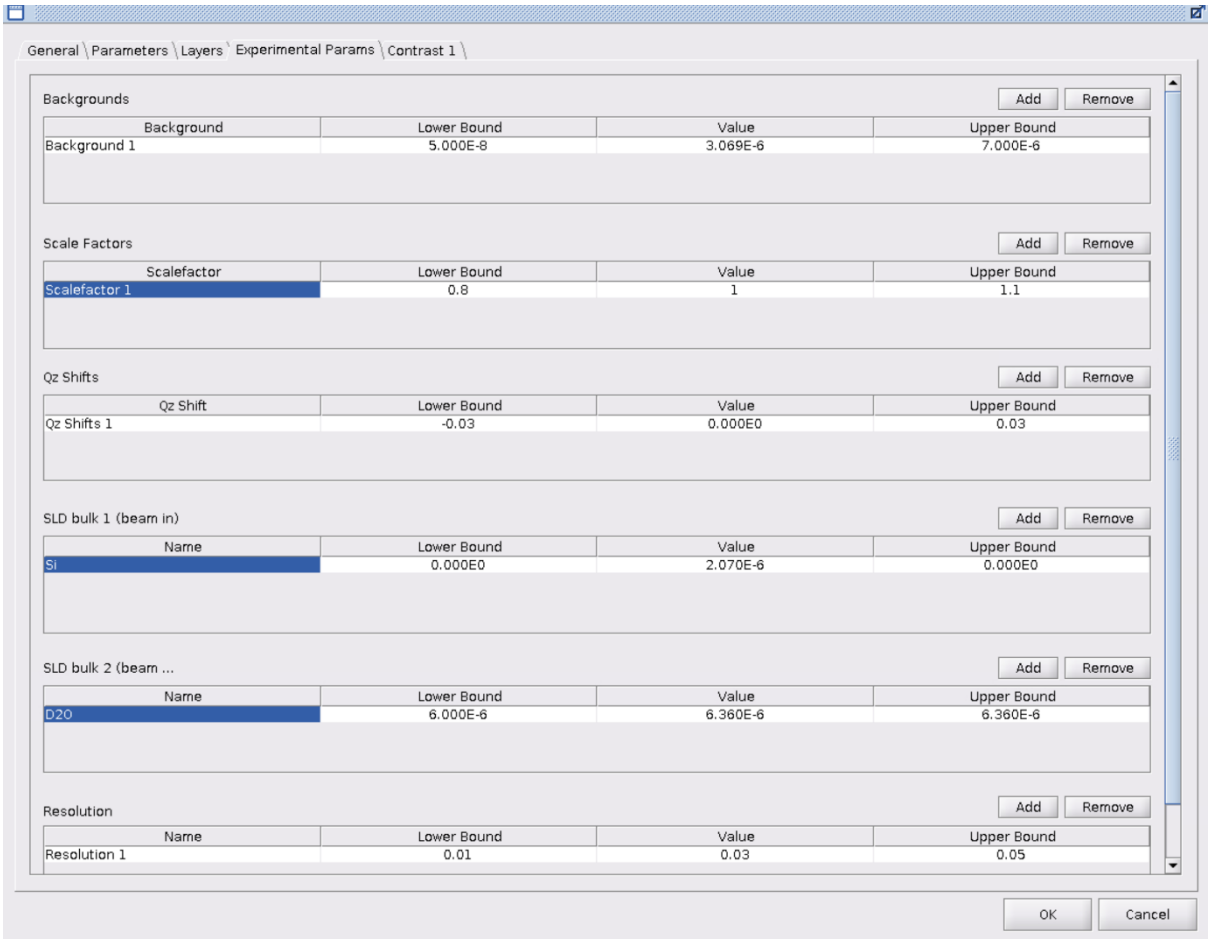


The ‘Layers’ tab is used to assign parameters to each interfacial layer in your model. For this sample, add a single layer corresponding to SiO_2, name the layer (**N.B**. Spaces **are** permitted in both parameter and layer names within the GUI), and select the parameters describing thickness, SLD, Roughness and Hydration from the drop-down menu.

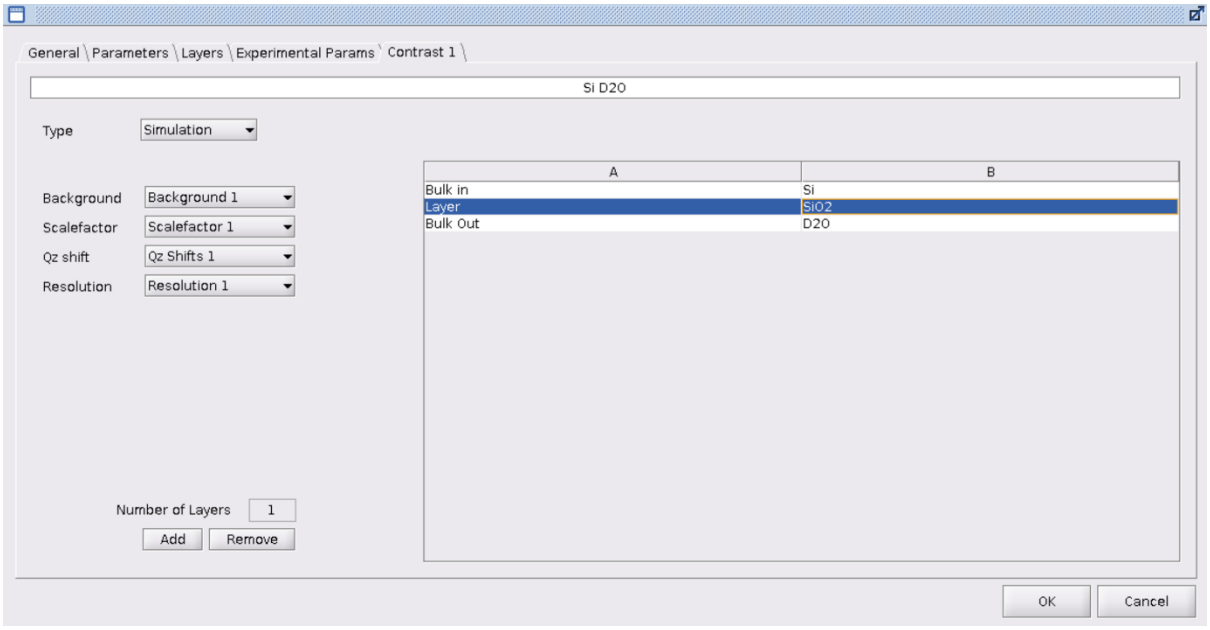


The ‘Experimental Parameters’ tab contains instrumental parameters such as backgrounds, scale factors, resolution as well as the SLDs of bulk phases.

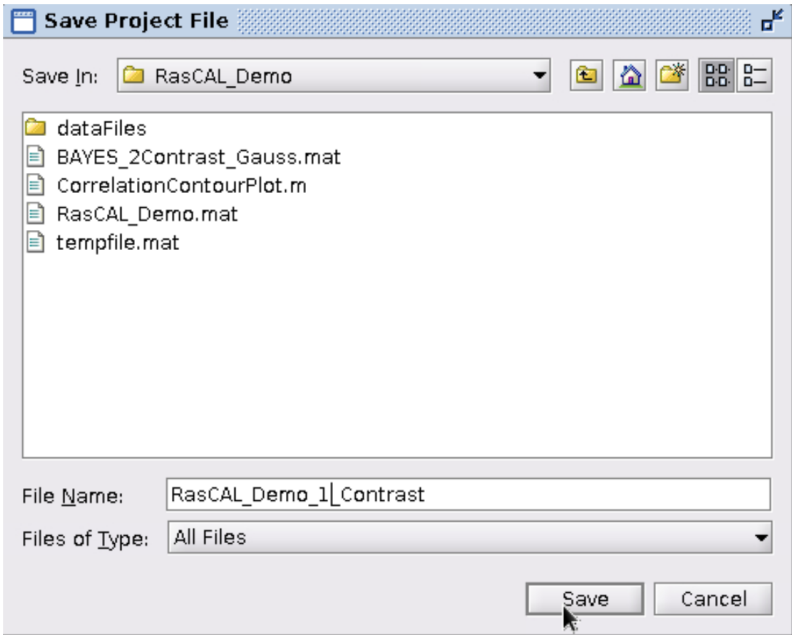
Reflectivity is defined as the ratio of reflected neutrons to incident neutrons (I/I_0), so should have a maximum intensity of 1 below the critical edge in the region of total external reflection. The scale factor parameter should therefore be around 1, though slight misalignment (amongst other issues) could lead to a slightly lower scale factor. The resolution is defined by the slit settings, and should be fixed based on the instrument configuration. In this case, the Q_z resolution was 3%, so the resolution should be fixed at 0.03. We can calculate the SLD of Si and D2O as 2.07×10^{-6} Å^{-2} and 6.36×10^{-6} Å^{-2}, so these values should also be entered here. Set your parameter values and bounds as shown below:



In the ‘Contrast’ tab, you assign instrumental parameters and layers to each contrast that you have measured. You can also choose to select the data to fit to here, or simply simulate data. We will start with simulated data of a single contrast. Select the Bulk In phase as Si, and the Bulk Out phase as D2O. To add an additional layer, select the ‘Bulk Out’ phase and then click the ‘Add’ Button (bottom left) to add a layer above the subphase. Choose SiO2 from the drop-down menu to assign the layer defined in the Layer tab to this sample.



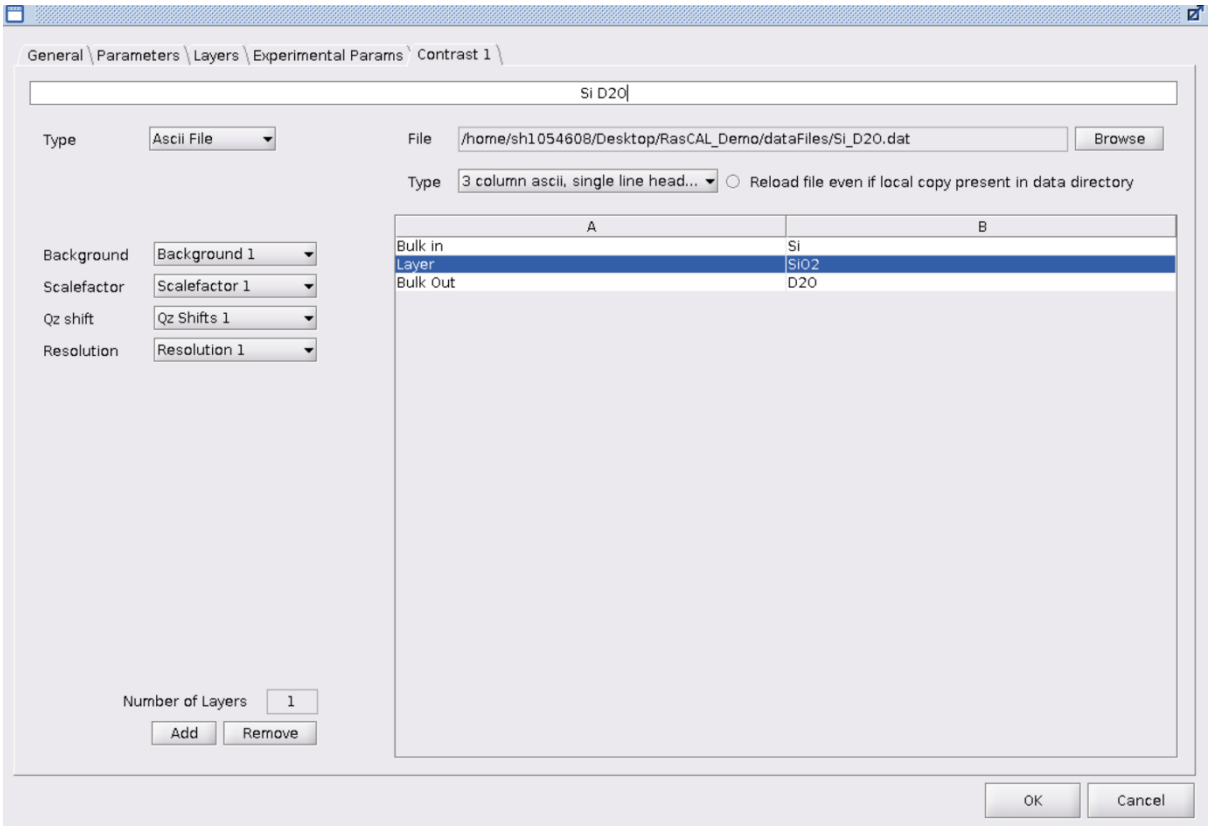
Press OK to build the model and return to the RasCAL main interface. Before we do anything further, save the project as a new file (‘File’ > ‘Save as new file’). Navigate to your IDAaaS home desktop and save this file within the ‘RasCAL_Demo’ directory:



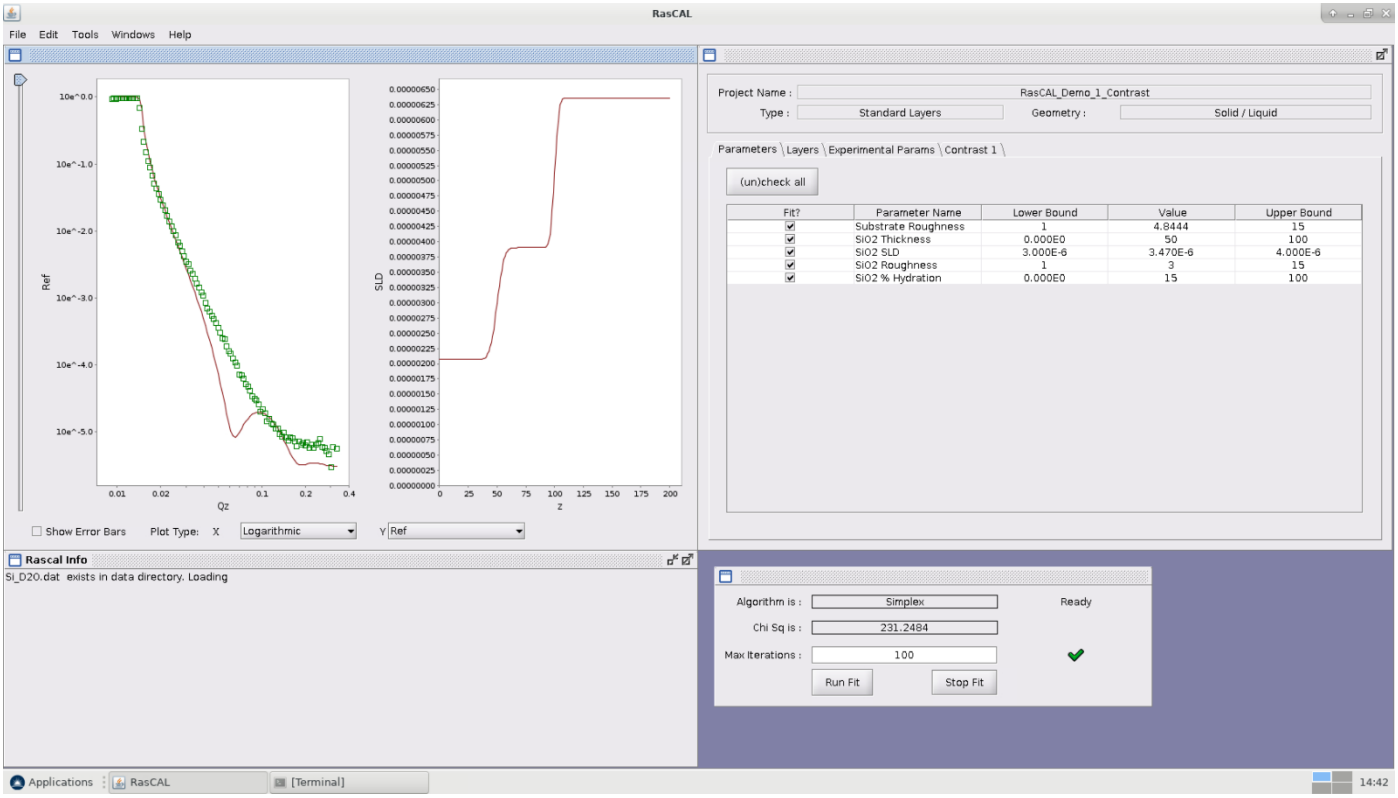
Loading Data

Before loading data, we need to make sure that RasCAL is looking in the correct directory. Therefore, load the file you have just saved in the RasCAL_Demo directory on your desktop (‘File’ > ‘Load’ and select ‘RasCAL_Demo_1_Contrast.mat’).

To add the data, go back to the Experimental Setup window (‘Edit’ > ‘Edit Project’) and on the Contrast 1 tab, change ‘Type’ from Simulation to ‘Ascii File’. Click the ‘Browse’ button and select ‘Si_D2O.dat’ from the ‘dataFiles’ directory within the ‘RasCAL_Demo’ directory.



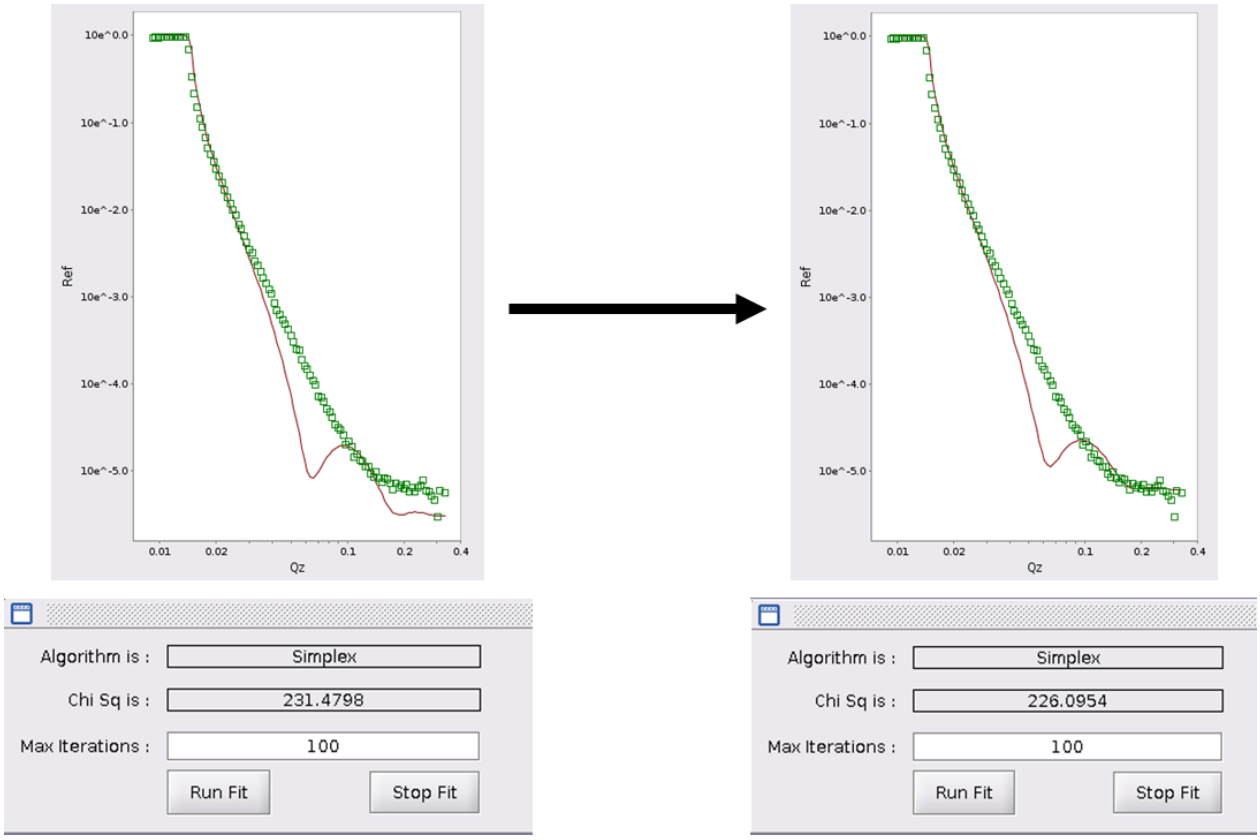
Press OK, and you should now see your data and the reflectivity profile calculated from your model parameter values:



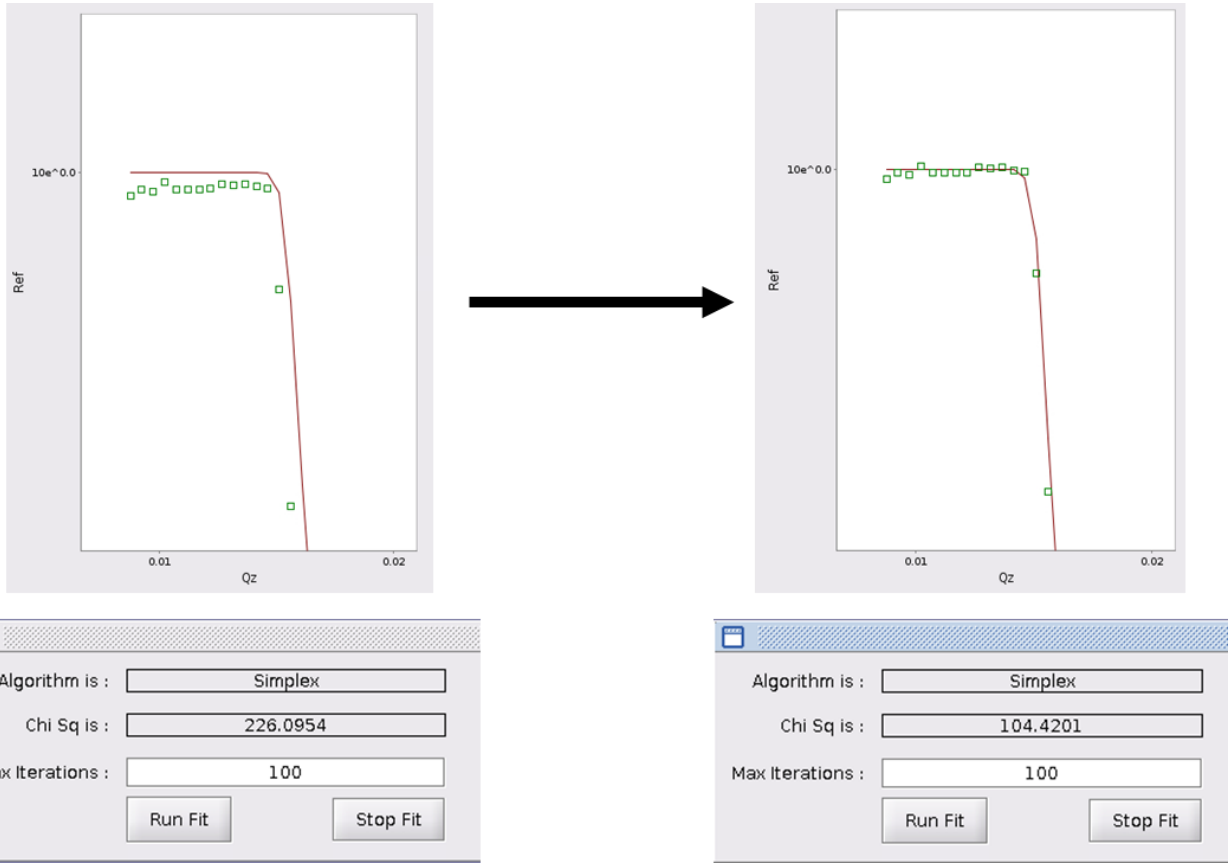
Part 2: Fitting a single contrast dataset.

Before we algorithmically fit the data, we should manually adjust some parameters to be closer to the correct values. This can help you to find the global minimum and not get stuck in local minima with nonsensical parameter values. The first thing to note, is that the incoherent background is far too low. We should change this parameter to be closer to the correct value. And parameter can be manually changed in two ways, either by double clicking on the value and entering a different value, or by checking the tick box by the parameter, then selecting ‘Tools’ > ‘Show Sliders’.

Use either of these methods to change the background to 6.1E-6. You should notice two things. First, the background now looks visibly at a correct level, and secondly, the Chi2 value has decreased, indicating the model now better describes the data.

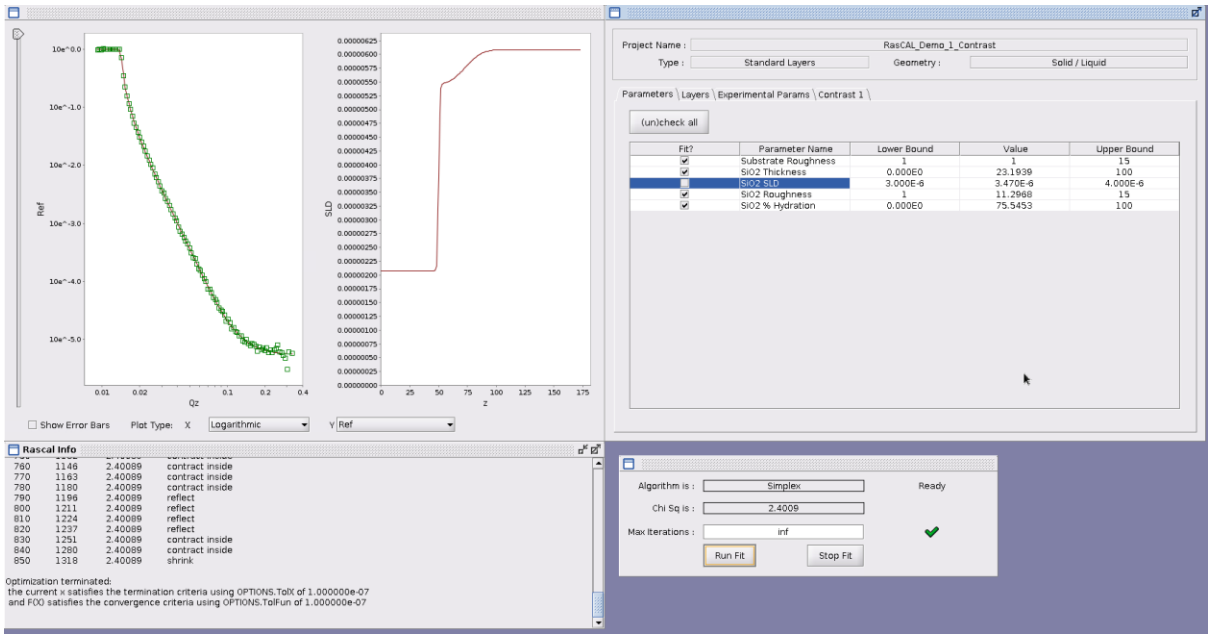


Next, we want to focus on the critical edge. Here, we are sensitive to the scalefactor (as below the critical edge, the reflectivity should, by definition, be equal to 1), and the SLD of D_2O. While we can calculate the SLD of D2O as 6.36×10^{-6} $Å^{-2}$, D_2O can exchange with atmospheric moisture. This means that there will be a small fraction of H2O or HOD present, so you will frequently observe a slightly lower SLD than the theoretical maximum.



Now we can setup a fit. We need to select parameters which we want to fit (check the ‘Fit?’ box for each parameter to fit, leave unchecked to fix the parameter). We know the SLDs of SiO_2 and Si, so these should be fixed. Likewise, we know the resolution, and we do not need any Qz shifts, so these should also be fixed. Other parameters (Substrate Roughness, SiO_2 Thickness, SiO_2 Roughness, SiO_2 % Hydration, Background 1, Scale, D_2O SLD) should be checked for fitting. In the Fitting panel, change ‘Max iterations’ to ‘inf’ and click ‘Run Fit’.

You should see the fit and SLD profile change as RasCAL performs the Simplex fit, as the Chi2 drops. You should end up with a good fit to the data, as shown below.



Once you are happy with the fit, save the file.

We will now perform Bayesian error estimation. First, save your fit so far, by selecting ‘File’ > ‘Save’. Launch the Bayesian analysis dialogue by selecting ‘Tools’ > ‘Bayesian Analysis’. All of the fitting parameters, best fit values and bounds will be automatically carried over from the previous best fit. There are a few options here you can change:

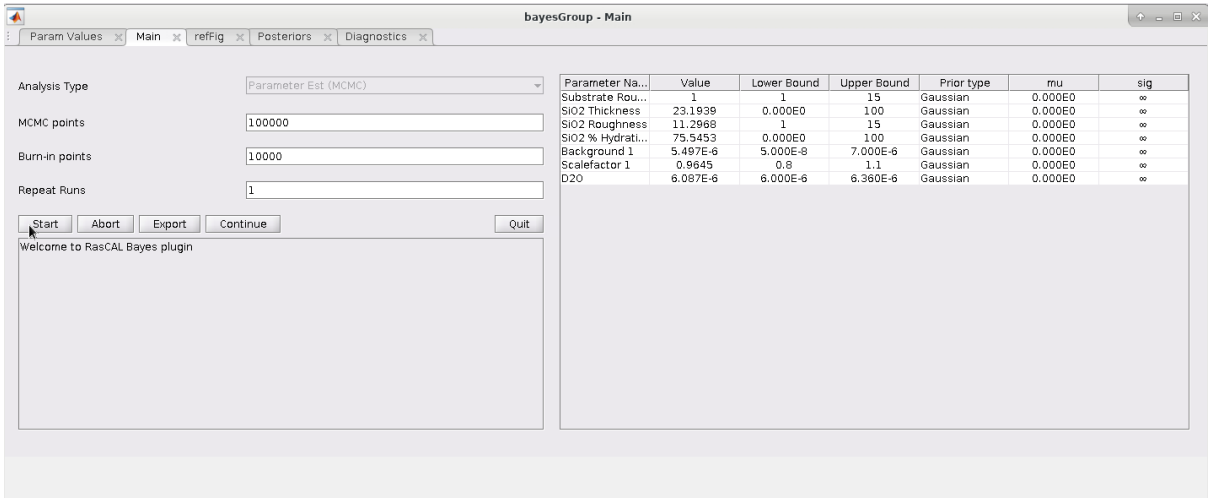
MCMC points: Number of Markov chain Monte Carlo iterations to perform

Burn-in points: Number of iterations to perform BEFORE iterations are taken for analysis. This should be high enough to find the global minima so values around the minima can be collected.

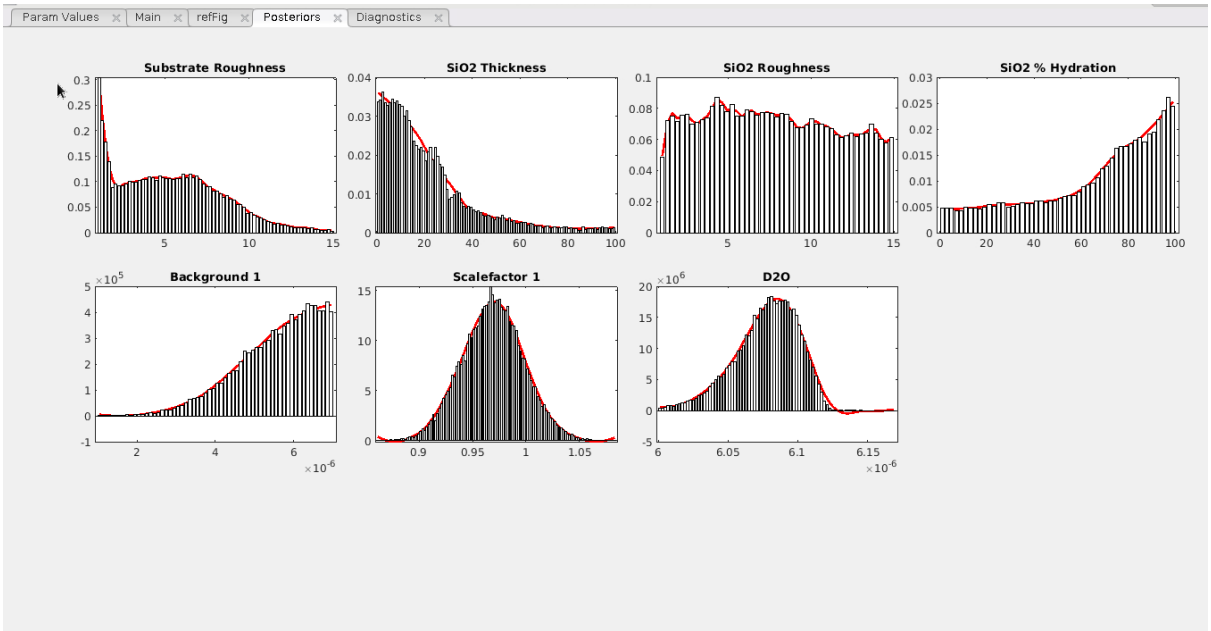
Repeat Runs: Number of times to repeat the analysis, which are then averaged.

Prior distribution: These options allow you to modify your prior distribution. If left as default, RasCAL will assume a uniform distribution between the lower and upper bounds for each parameter. If you have prior information on the sample, you can change to a Gaussian prior, with a mean value of ‘mu’, and a standard deviation of ‘sig’. For our purposes, we have no prior knowledge of the sample, so will proceed using uniform priors.

For this example, 100000 MCMC points, 10000 Burn-in points and 1 repeat run will be sufficient. Enter these options as shown below and click ‘Start’.



This analysis should take around 1-2 minutes. After which you should see the posterior distributions for each parameter:



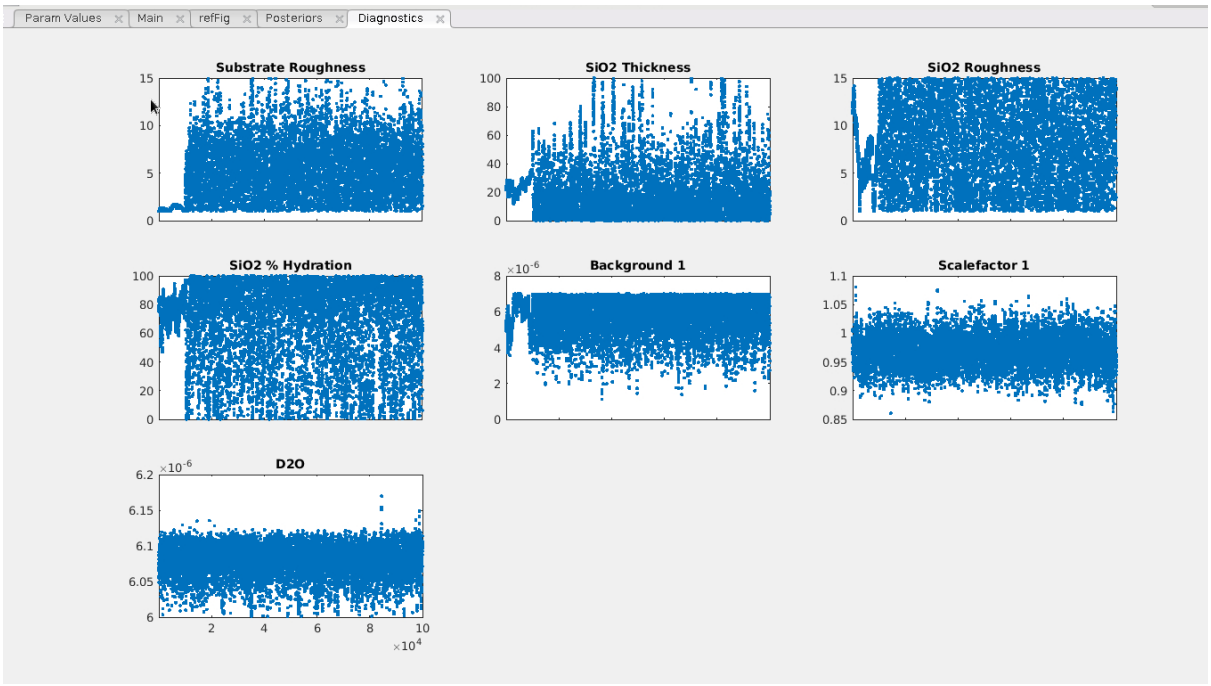
There are a few noteworthy points here. The posterior distributions are generally quite broad, indicating ambiguity in the analysis. Particularly, the SiO_2 Roughness is effectively constant over the fit range, showing the data is not able to unambiguously describe this parameter. Similarly, there are broad distributions for the substrate roughness, SiO_2 Thickness and SiO_2 Hydration. In contrast, we have good confidence in the scalefactor and D2O SLD. This is due to the critical edge in the data explicitly describing these parameters.

Clicking on the 'Param Values' tab will show you the best values and confidence intervals calculated from these posterior distributions:

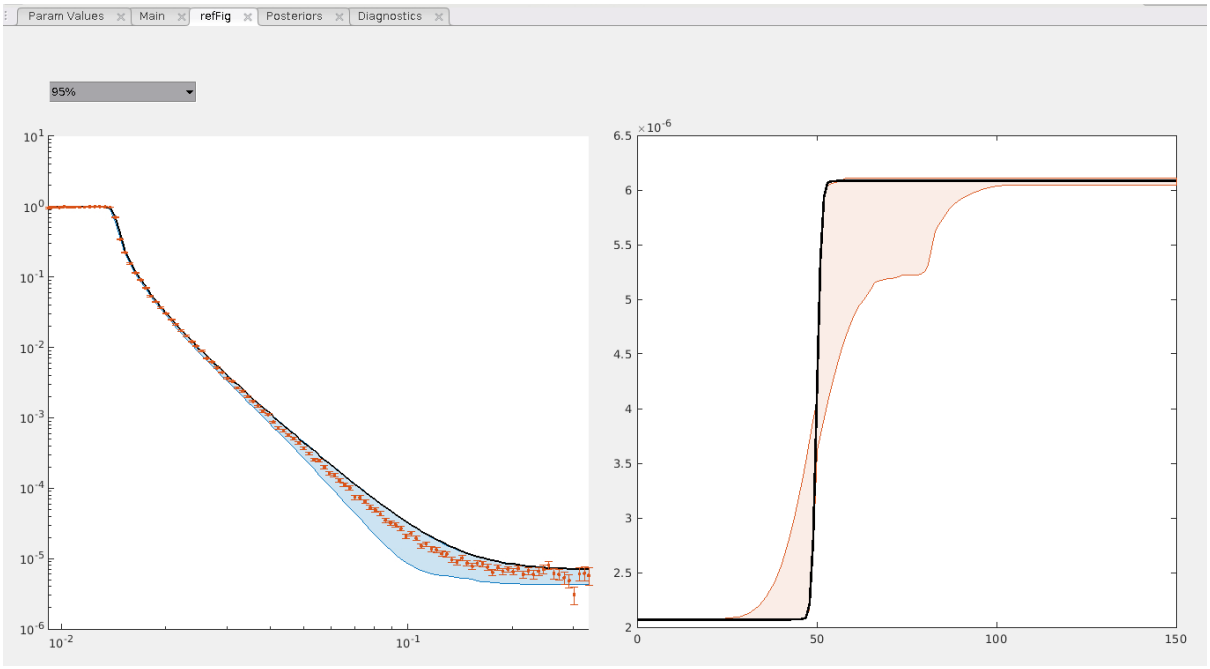
Param Values | Main | refFig | Posteriors | Diagnostics

Parameter Name	Best Value	95% CI	65% CI	25% CI
Substrate Roughness	1.1	[1.000e+00, 8.728e+00]	[1.012e+00, 6.534e+00]	[4.851e+00, 7.162e+00]
SiO2 Thickness	0.5	[1.084e-03, 3.459e+01]	[2.392e-01, 2.286e+01]	[1.560e+01, 2.770e+01]
SiO2 Roughness	4.35	[1.114e+00, 1.099e+01]	[4.884e+00, 1.404e+01]	[7.525e+00, 1.100e+01]
SiO2 % Hydration	99	[6.499e+01, 9.999e+01]	[6.191e+01, 9.999e+01]	[2.949e+01, 6.645e+01]
Background 1	6.950E-6	[4.292e-06, 7.000e-06]	[5.223e-06, 6.952e-06]	[5.544e-06, 6.234e-06]
Scalefactor 1	0.971	[9.243e-01, 1.011e+00]	[9.433e-01, 9.966e-01]	[9.500e-01, 9.693e-01]
D2O	6.085E-6	[6.047e-06, 6.113e-06]	[6.063e-06, 6.103e-06]	[6.061e-06, 6.078e-06]

Clicking on the ‘Diagnostics’ tab will show you the values obtained for each parameter for each MCMC iteration. Ideally, you should see well mixed parameters that appear to randomly vary throughout the analysis. Here, you can see that for a few parameters, there are some trends in a few parameters over the first few runs. This indicates the fit had not fully equilibrated prior to the MCMC analysis, so should be repeated with a larger number of burn-in points to accurately estimate the uncertainty.



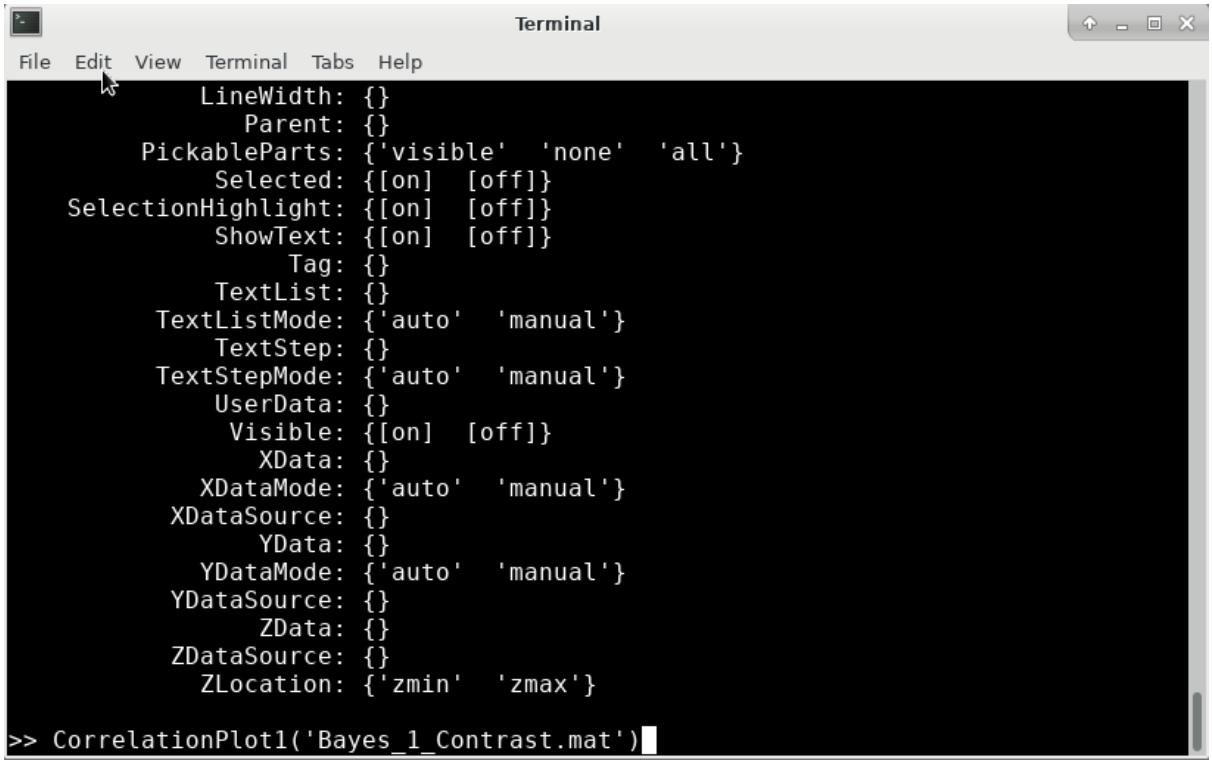
Clicking on the ‘refFig’ tab will allow you to see the best fit and corresponding SLD profiles as solid lines, with the confidence intervals displayed as shaded regions.



As you can see, there is currently a high level of uncertainty associated with the best fit parameters of the model. In addition to posterior distributions, it is also important to consider correlations between parameters. Thickness and SLD/hydration are frequently correlated, so can further increase the uncertainty in the analysis.

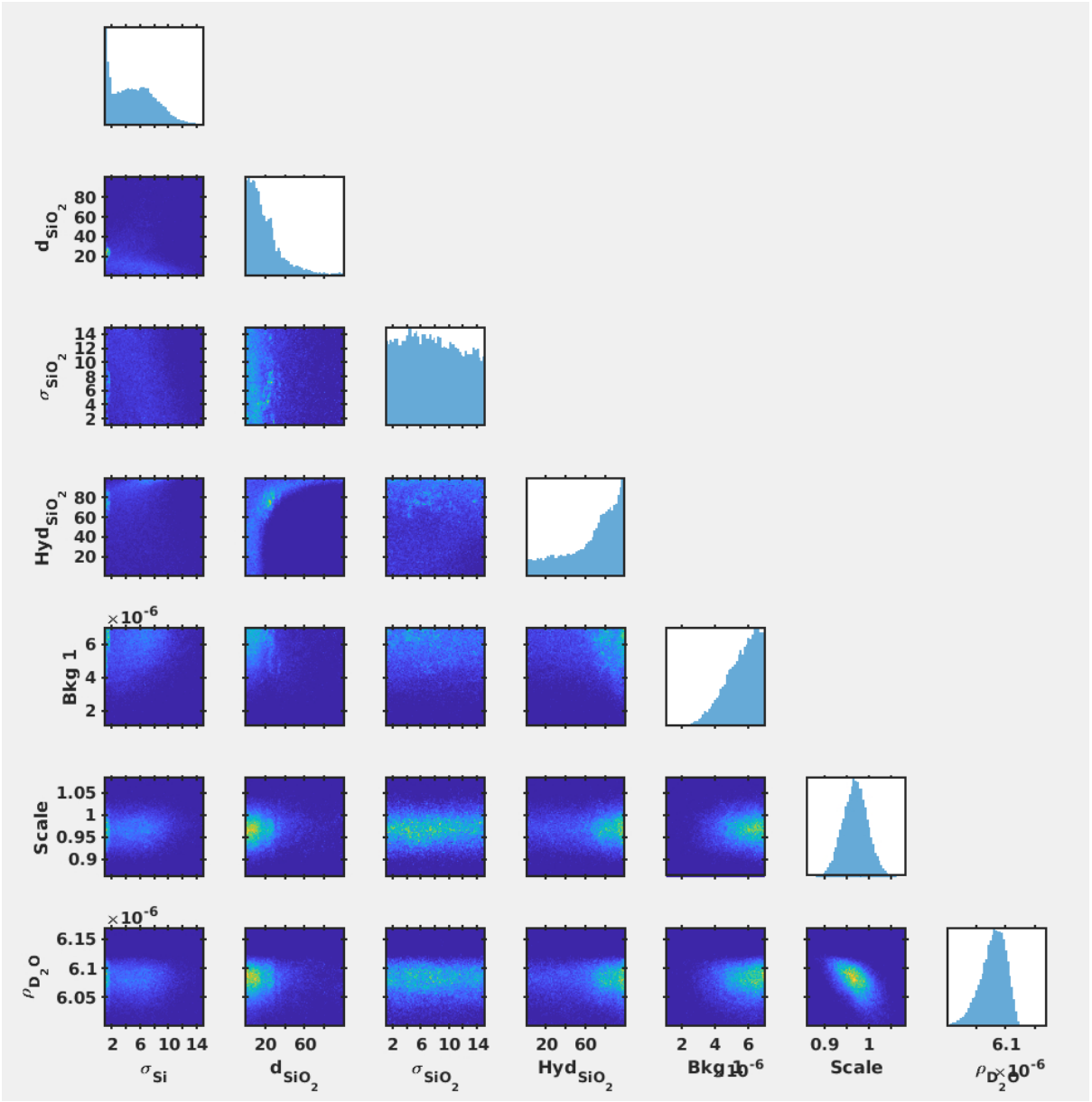
Here, you can inspect the correlations between parameters by using the 'CorrelationPlot' function included in the 'RasCAL_Demo' directory. To use this function, on the 'Main' Tab of the Bayesian analysis window, select 'Export' and save the analysis as open the Matlab terminal open on the taskbar, and enter 'Bayes_1_Contrast' in the Save Name box, and press 'OK'. Close the Bayesian analysis interface by pressing 'Quit'. Select the Matlab terminal which is open on the task bar, and enter:

```
CorrelationPlot1('Bayes_1_Contrast.mat')
```



Note: CorrelationPlot is included here for demonstration purposes, but is not distributed with RasCAL. With the exception of axes labels, this is a general function that should work with any Bayesian analysis exported from RasCAL. Feel free to use this function (with some simple modification) for your own analysis if you wish.

After a few seconds the following figure should be generated (you may need to expand the window to see the individual plots):



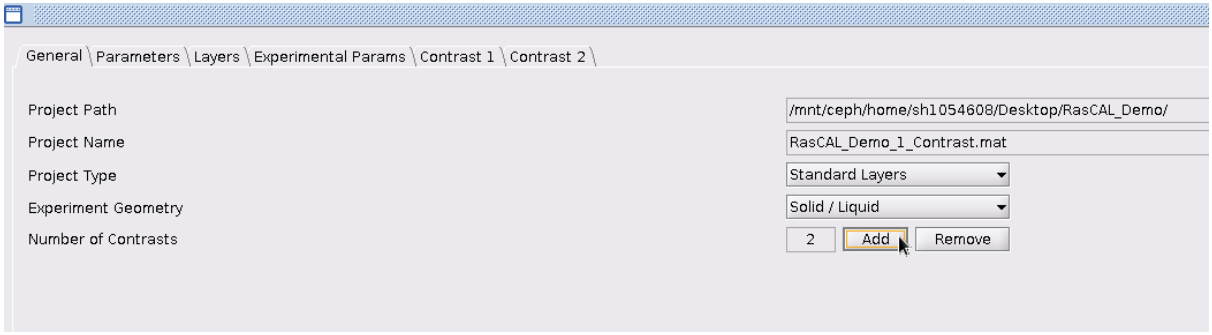
Here, the posterior distributions for each parameter are shown as histograms, and the correlations are shown as contour plots for the parameter labelled on the X axis, against the parameter shown on the Y axis. If a parameter pair is non correlated, it will appear as a circular contour. If the parameter pair is correlated a diagonal contour will be shown. Here, you can see that there is are correlations between several parameter pairs.

Do not close the terminal after this, but feel free to minimize!

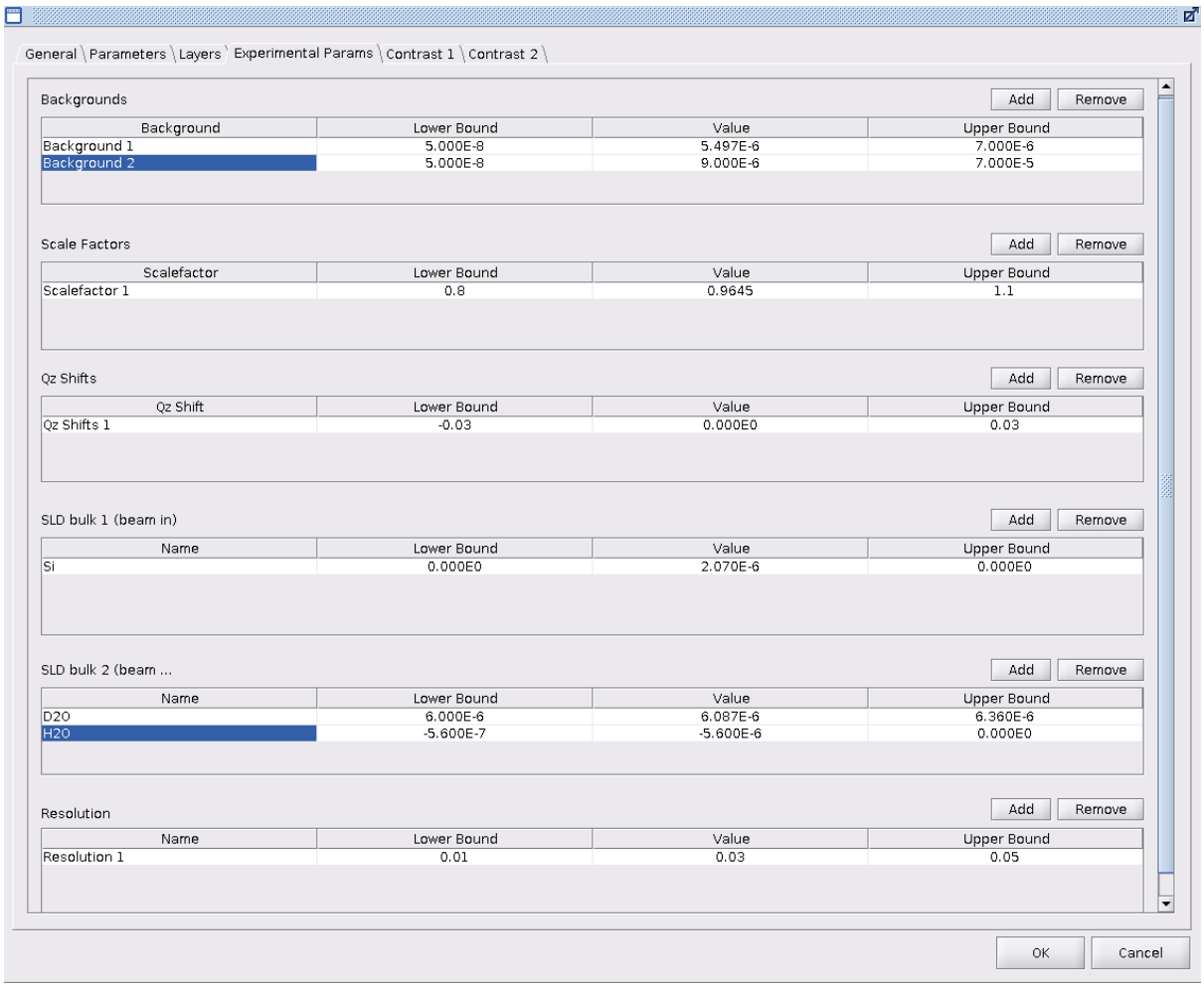
Part 3: Co-refinement of multiple contrasts.

As we showed in Part 2, we can obtain a good fit to the NR data collected from a clean Si-D_2O interface, but Bayesian error estimation shows that we have very little confidence in the model. Particularly, the interfacial roughness and hydration is poorly resolved. This can be improved by collecting an additional isotopic contrast. As you will have heard in previous lectures, there is a large difference in the scattering lengths of deuterium and protium. We can exploit this difference by collecting NR data of the same sample in H_2O (SLD = -0.56×10^{-6} Å^{-2}) in addition to D_2O (SLD = 6.36×10^{-6} Å^{-2}). This will allow us to improve our confidence in the model.

First, we need to setup the model for two contrasts. In the main RasCAL window, enter the project editing window ('Edit' > 'Edit Project') and add an additional contrast by clicking 'Add' on 'Number of Contrasts'.

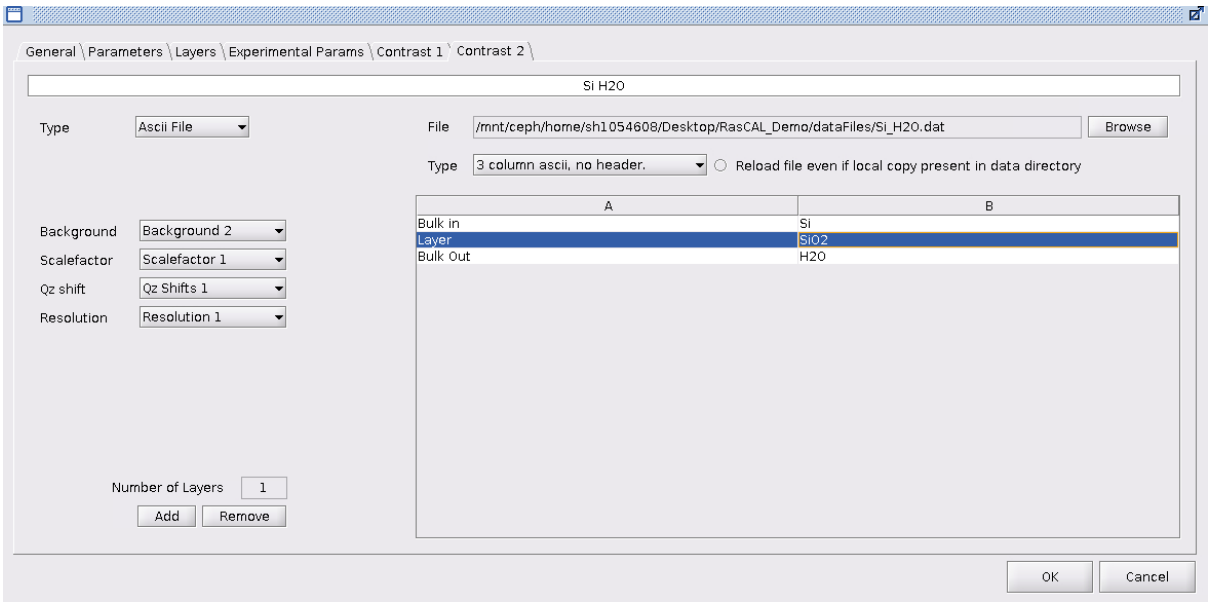


As the interfacial structure is the same between both datasets, we only need to include some additional experimental parameters. The incoherent background is higher for samples with a greater proportion of H_2O, so add an additional background by selecting 'Add' under the 'Backgrounds' section. As mentioned above, the SLD of H_2O is substantially different to that of D_2O, so an additional 'bulk out' (i.e. subphase) SLD is required. Enter values and bounds as shown below:



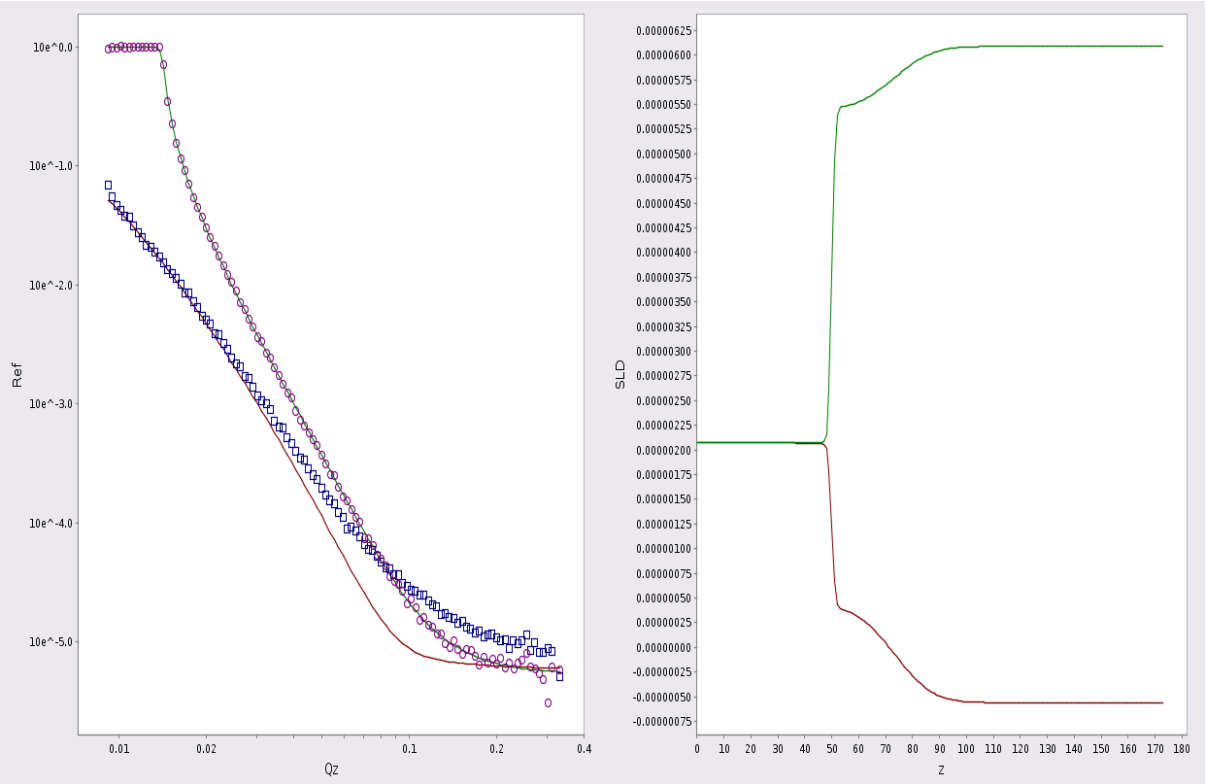
Finally, on the ‘Contrast 2’ tab, we need to assign parameters to the additional dataset. Under background, select ‘Background 2’. Change ‘Bulk Out’ to H2O, and add a layer between H2O and Si, selecting ‘SiO2’.

Change ‘Type’ from ‘Simultation’ to ‘Ascii File’. Click the ‘Browse’ button and select ‘Si_H2O.dat’ from the ‘dataFiles’ directory within the ‘RasCAL_Demo’ directory.



Click ‘OK’ and then save as a new file (‘File’ > ‘Save As New File’), as ‘RasCAL_Demo_2_Contrast.

You will now see both datasets loaded, with the previous model applied to both contrasts. Note that while the model produces a good fit to the D_2O contrast, the fit to the H_2O contrast is poor, indicating that the previous model does not accurately describe the interfacial structure.

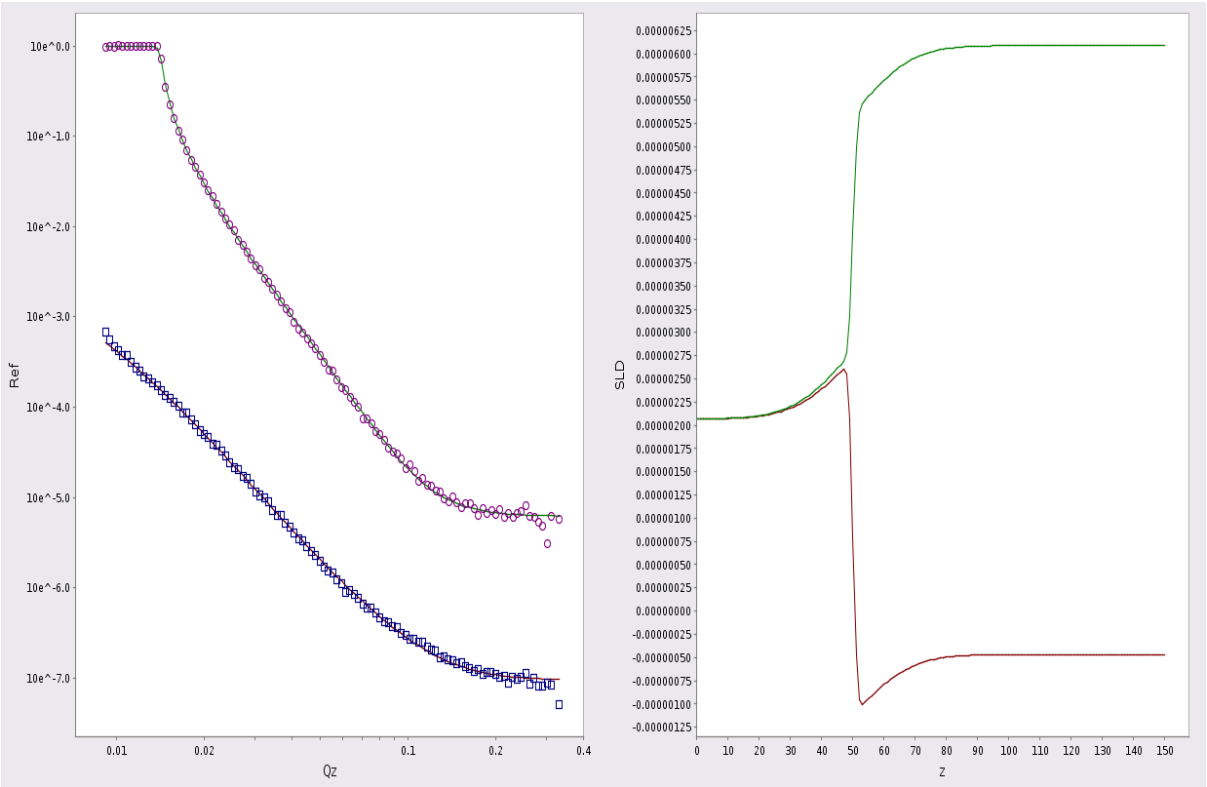


We will fit these data in a similar way to the previous example. As before, adjust 'Background 2' (describing the background for the H_2O contrast) to a sensible value. Fit the data with the same parameters checked as previously, with the addition of 'Background 2' and 'H2O SLD'.

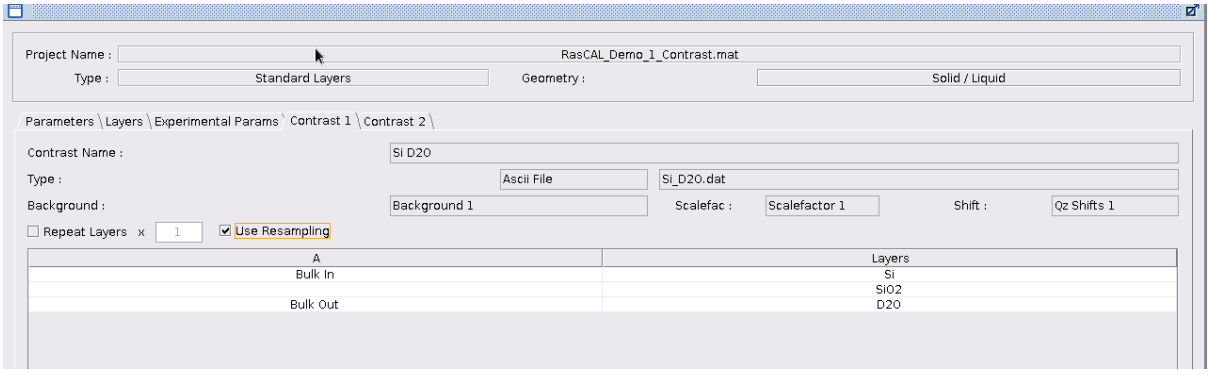
Fitted parameters should be:

- **Substrate Roughness**
- **SiO2 Thickness**
- **SiO2 Roughness**
- **SiO2 Hydration**
- **Background 1**
- **Background 2**
- **Scalefactor1**
- **D2O SLD**
- **H2O SLD**

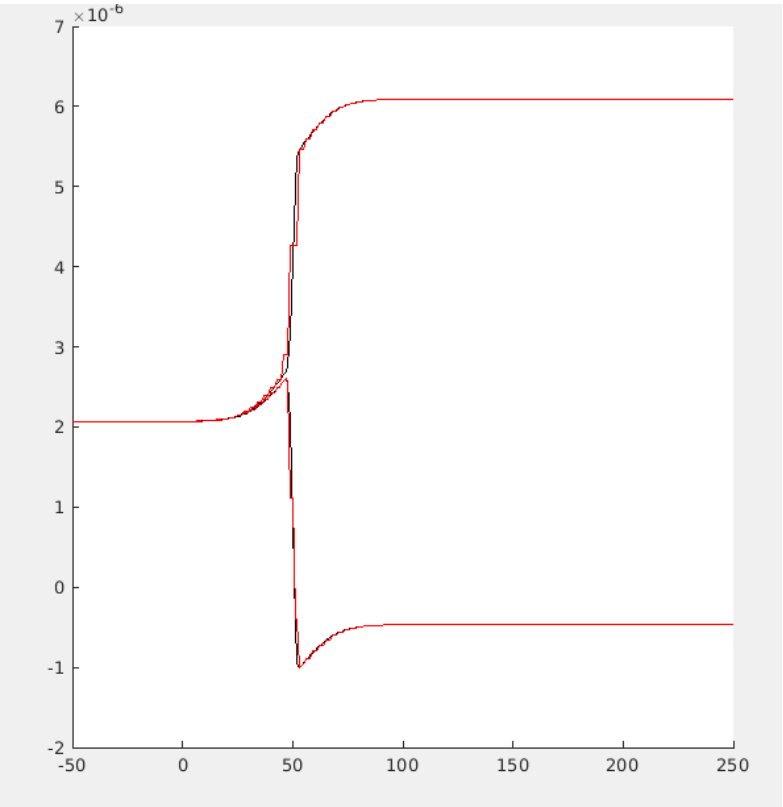
When you fit these data, you may end up with a result similar to below.



While this appears to produce a good fit to the experimental data, it is a unrealistic SLD profile. You will note that the roughness of the Si-SiO_2 interface is very high, compared to a very low roughness associated with the SiO_2-bulk solvent interface. This roughness of the Si-SiO_2 interface appears to 'bleed' through to the subsequent interface. By default, roughnesses are calculated using the Nevot-Croce error functions. This apparent 'bleeding' of the roughness into the subsequent layers is not accounted for in the calculated reflectivity profiles. We can correct this using 'microslicing', where the roughness between interfaces is resampled as multiple thin layers of zero roughness, allowing the calculation of an accurate reflectivity profile for the displayed SLD profile. In RasCAL, this is done by checking the 'Use Resampling' option under the 'Contrast 1' and 'Contrast 2' tabs:



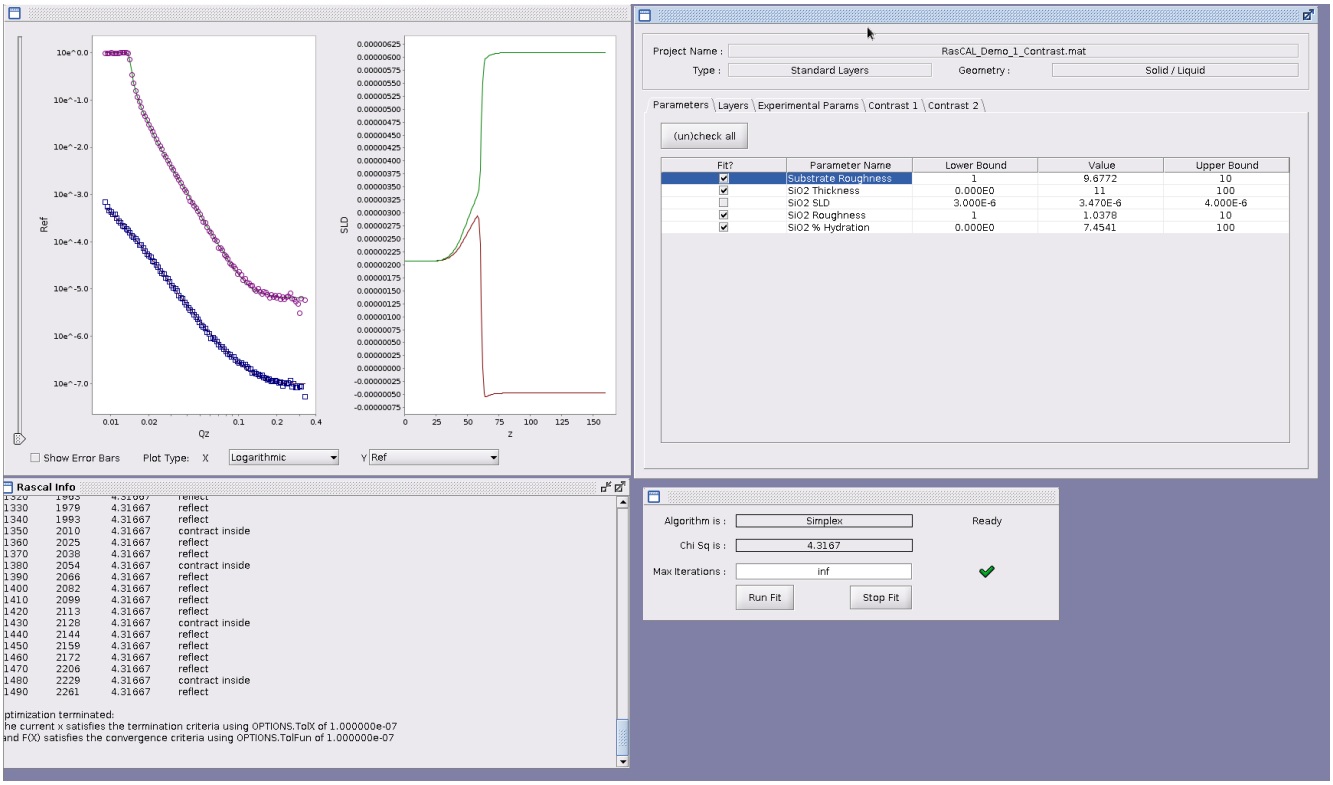
You can see how the roughnesses are microsliced by selecting ‘Tools’ > ‘Preferences’ and clicking on the Check button on the ‘Resample’ tab. Here the black curve uses Nevot-Croce roughness, and the red curve uses the resampled microsliced roughness.



Despite this correction, the resultant fit is not a realistic representation of the interface, suggesting the fit has got ‘stuck’ in a local minimum. We therefore need to manually adjust the fit parameters to sensible values, and change the parameter bounds to stop the fit getting stuck in this unrealistic local minimum. Initially change the following parameters and bounds and then rerun the fit:

Parameter	**Low Bound**	**Value**	**High Bound**
Substrate Roughness	1	5	10
SiO2 Thickness	0	15	50
SiO2 Roughness	1	5	10
SiO2 % Hydration	0	25	100

You should now have a fit with similar parameters as shown below:

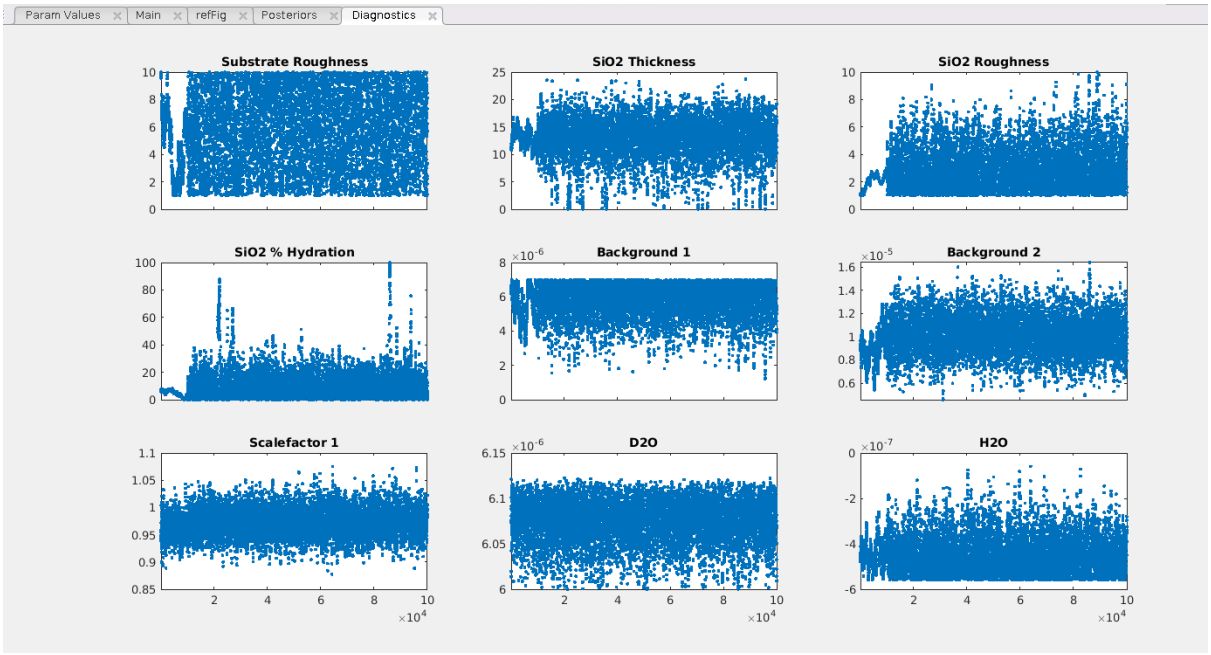


Save your analysis by selecting ‘File’ > ‘Save’.

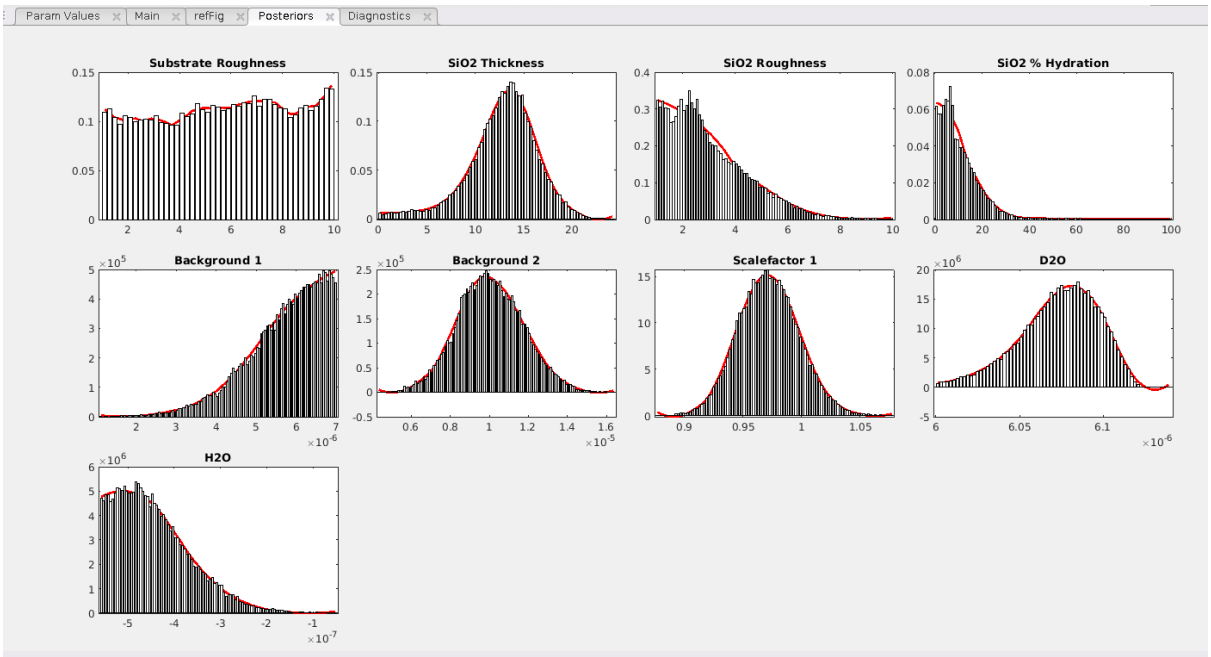
Now, we want to run the Bayesian analysis, with the same settings as we had previously. This will allow us to see the influence of including an additional contrast on the posterior distributions of each of the fitted parameters.

Note: Now we have an additional dataset loaded and are using resampled roughnesses, this will take a little longer than the previous analysis.

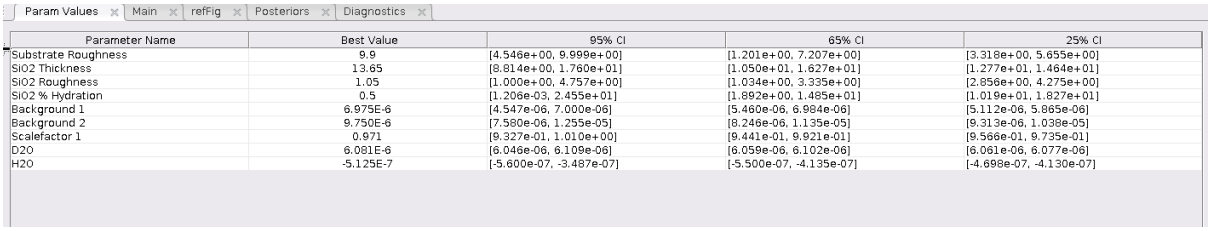
Once the analysis has finished, click onto the ‘Diagnostics’ tab to see the parameter values through each MCMC iteration. As with the previous analysis, good sampling can be observed for each parameter, but the number of burn-in points should be increased to ensure that the global minimum has been identified prior to the sampling.



As before, you will be able to see the posterior distributions of the fitted parameters. Note that we now have a lot more confidence in the thickness, roughness and hydration of the SiO_2 layer but the substrate roughness remains ambiguous:

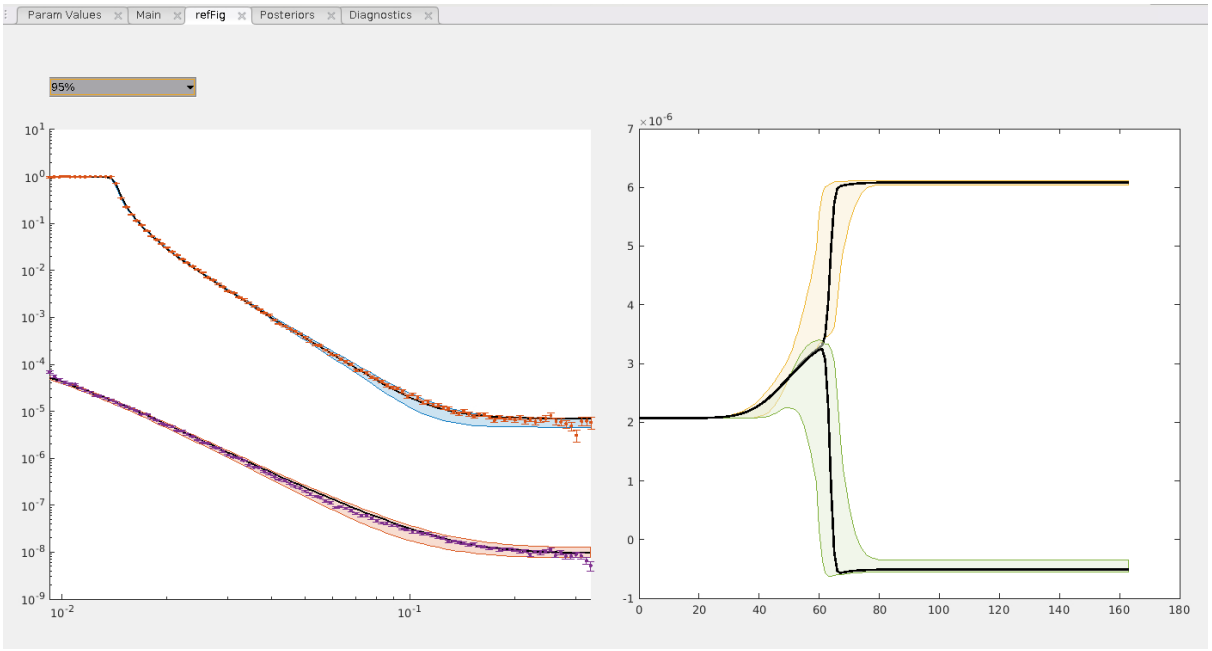


This can also be identified in the parameter best fit and confidence intervals, where the uncertainty is improved in all cases bar the Substrate roughness, which remains ambiguous.



Param Values | Main | refFig | Posteriors | Diagnostics

Parameter Name	Best Value	95% CI	65% CI	25% CI
Substrate Roughness	9.9	[4.546e+00, 9.999e+00]	[1.201e+00, 7.207e+00]	[3.318e+00, 5.655e+00]
SiO2 Thickness	13.65	[8.814e+00, 1.760e+01]	[1.050e+01, 1.627e+01]	[1.277e+01, 1.464e+01]
SiO2 Roughness	1.05	[1.000e+00, 4.757e+00]	[1.034e+00, 3.335e+00]	[2.856e+00, 4.275e+00]
SiO2 % Hydration	0.5	[1.206e-03, 2.455e+01]	[1.892e+00, 1.485e+01]	[1.019e+01, 1.827e+01]
Background 1	6.975E-6	[4.547e-06, 7.000e-06]	[5.460e-06, 6.984e-06]	[5.112e-06, 5.865e-06]
Background 2	9.750E-6	[7.580e-06, 1.255e-05]	[8.246e-06, 1.135e-05]	[9.313e-06, 1.038e-05]
Scalefactor 1	0.971	[9.327e-01, 1.010e+00]	[9.441e-01, 9.921e-01]	[9.566e-01, 9.735e-01]
D2O	6.081E-6	[6.046e-06, 6.109e-06]	[6.059e-06, 6.102e-06]	[6.061e-06, 6.077e-06]
H2O	-5.125E-7	[-5.600e-07, -3.487e-07]	[-5.500e-07, -4.135e-07]	[-4.698e-07, -4.130e-07]

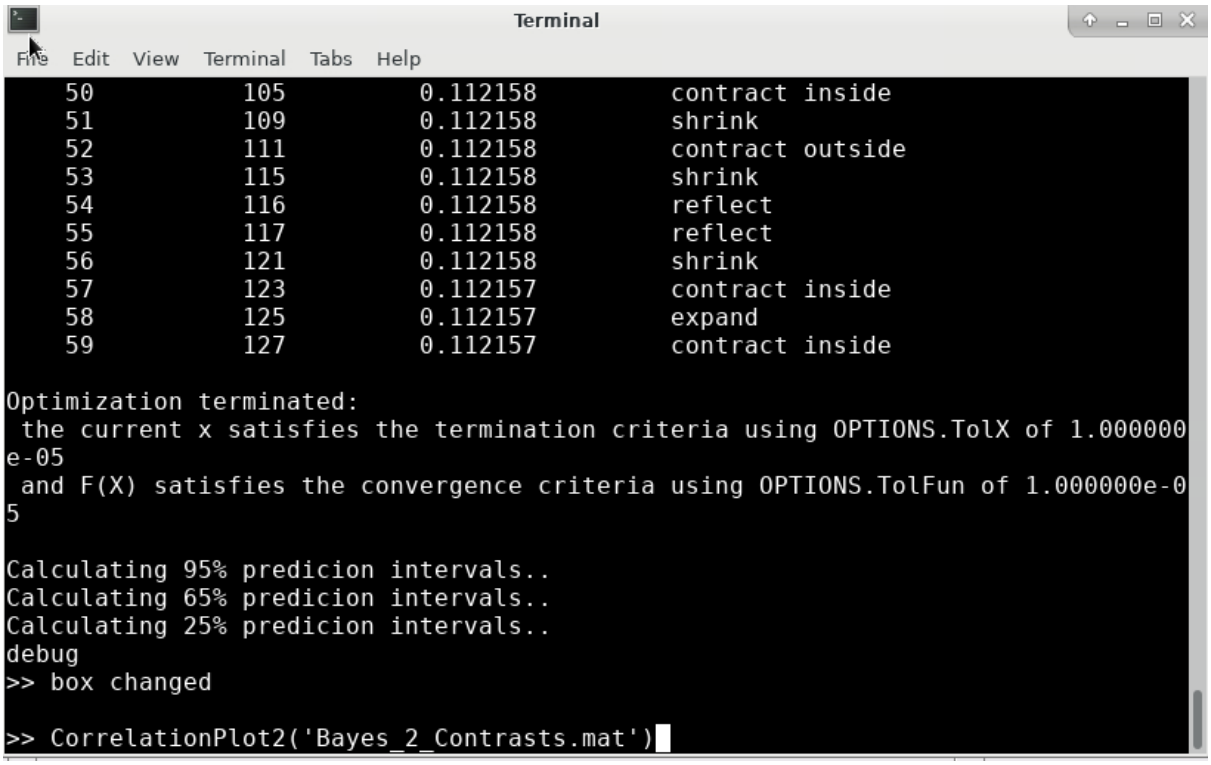


Similarly, inspecting the calculated reflectivity and SLD profiles, while ambiguity is still present, the 95% confidence intervals have shrunk considerably compared to the same model fitted to a single contrast dataset.

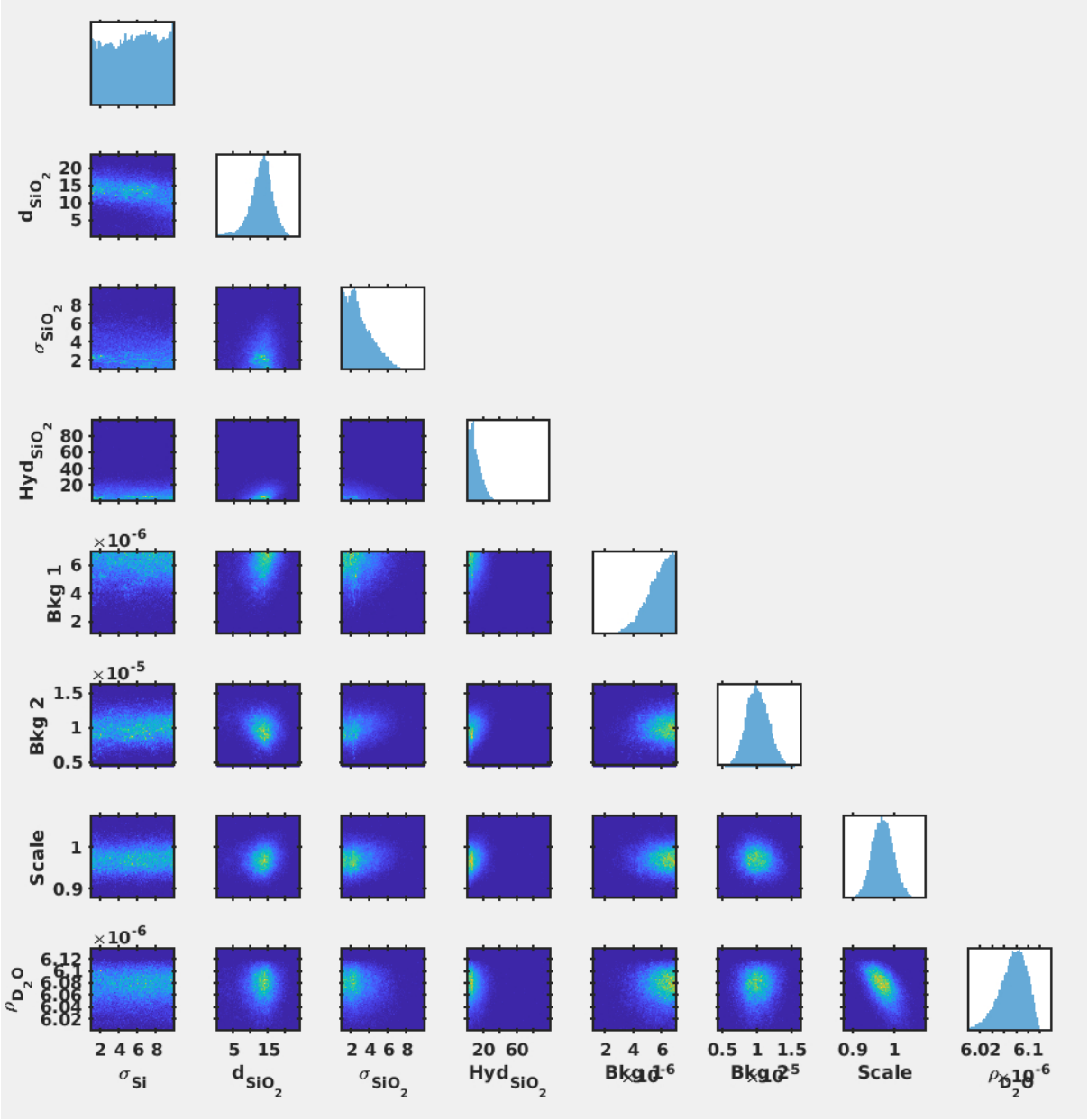
Finally, we can inspect how the inclusion of an additional isotopic contrast has influenced the correlations between fitted parameters.

As before, on the ‘Main’ tab of the Bayesian analysis dialogue, export the analysis as ‘Bayes_2_Contrasts’ and close the Bayesian analysis window by selecting ‘Quit’. Open the Matlab terminal from the task bar and type:

```
CorrelationPlot2(‘Bayes_2_Contrasts.mat’)
```



After a few seconds, the following figure will be generated (you may need to expand the figure window to see the individual plots):



Here we can see a substantial improvement in the correlations between parameters, particularly involving the thickness, roughness and hydration of SiO_2.

This demonstration has hopefully introduced you to analysis of simple neutron reflectometry data in RasCAL, and demonstrated how the collection of additional isotopic contrasts of the same interfacial structure can improve the ambiguity associated with the resulting model parameters.