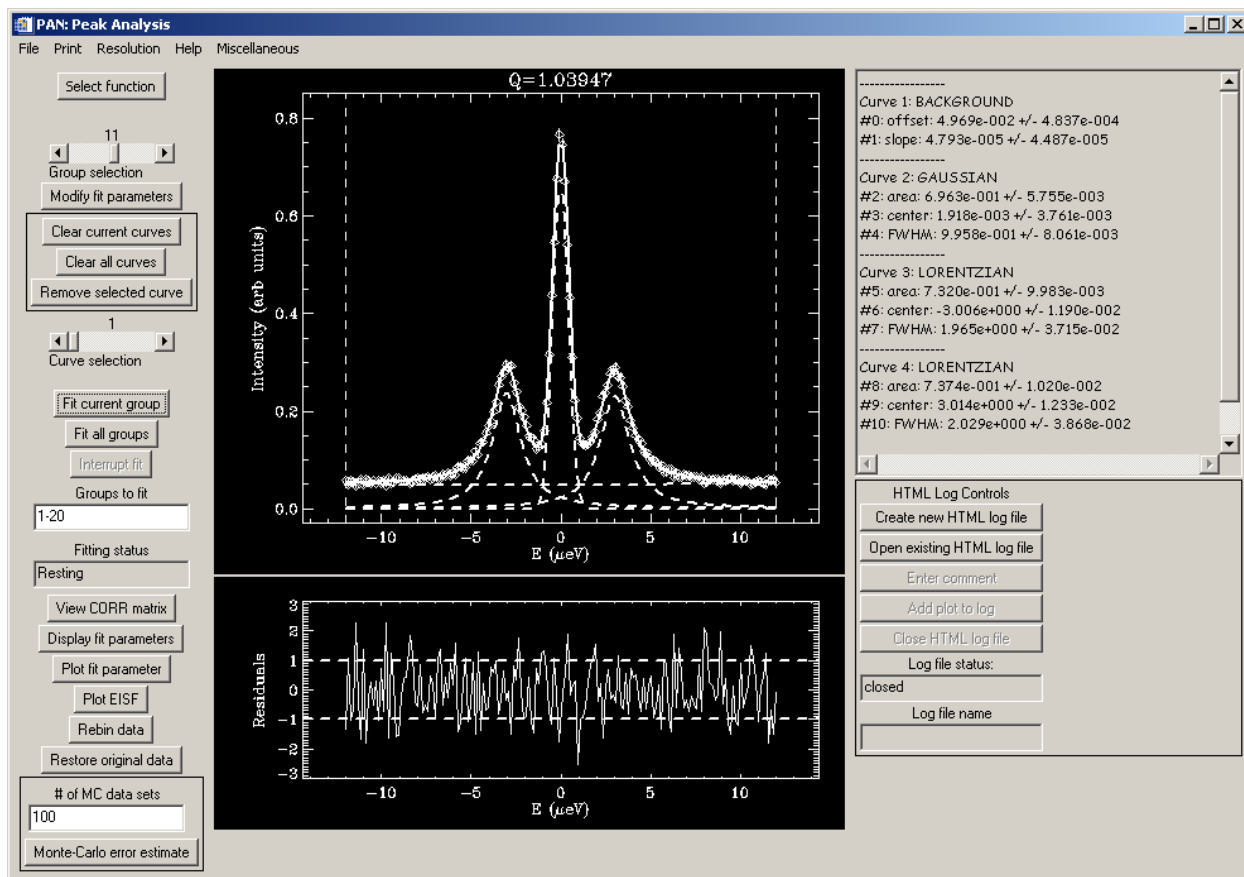# Peak Analysis: PAN
A general-purpose curve-fitting program

A user's guide

Rob Dimeo

October 4, 2002
(revised 07/14/05)

Contents

1.0   PAN Overview

1.1   What is PAN?

PAN is a general-purpose curve-fitting utility with a graphical user interface written in IDL. This means that you must have IDL running on your computer to run PAN. I wrote PAN using IDL v5.4 but it might work ok on earlier versions of IDL. You can use PAN to fit your ASCII data to any one of a library of fit functions and/or type in your own custom fit function. There is no limit to the number of functions to which you can fit your data but more functions result in more computation time.

PAN was written such that the amount of typing done by the user (in entering initial parameter guesses, for instance) is minimized. This was accomplished largely through the use of IDL's sophisticated API toolkit. One of the features of this program is that the user can select a function from the built-in library and use the mouse to specify amplitude, position and width, all with a few mouse clicks.

1.2   How do I get PAN?

You can download PAN from
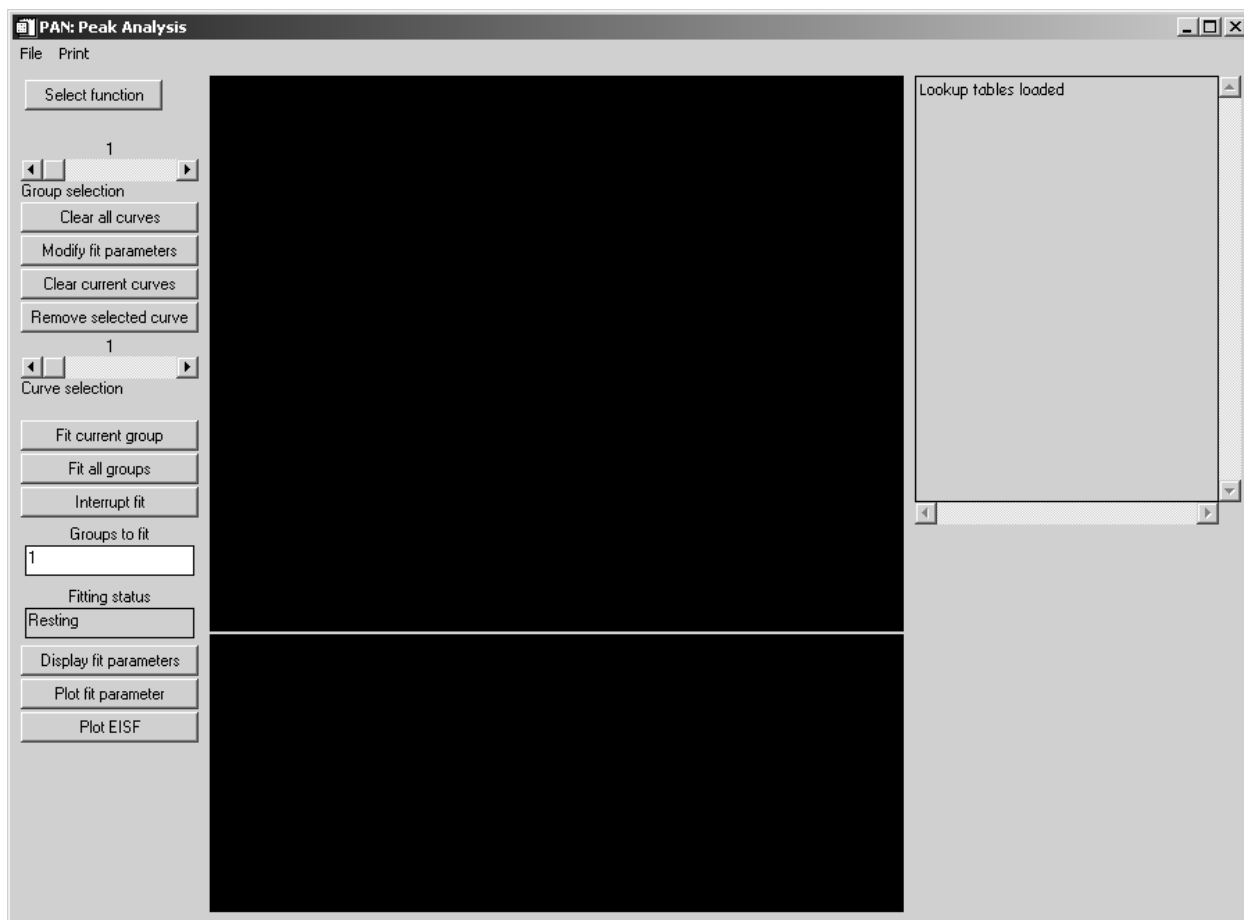ftp://ftp.ncnr.nist.gov/pub/staff/dimeo/pan.zip.

1.3   How do I launch PAN?

Simply unzip the file onto your computer and make sure that the new folders created on your computer are your local IDL path. Type PAN at the IDL command prompt to begin:

```
IDL> pan
```

After a few moments, you should see a screen similar to that shown in figure 1.

Figure 1 Main interface for PAN.

1.4   How do I get PAN to read in my data?

Your data can come in two varieties in which PAN will read it: (a) single group or (b) multiple groups

      a.   3-column tab- or space- delimited: x,z,zerr where x is the independent variable, z is the dependent variable, and zerr is the measurement error in the dependent variable.

            Example: Simple data set consisting of x, z and zerr

```
0.000            1.000            1.000
2.000            4.000            2.000
6.000            9.000            3.000
```

      b.   Grouped ASCII data: x,y,z,zerr where x and y are the independent variables, z is the dependent variable, and zerr is the measurement error in the dependent variable.

All lines that begin with a hash # are comment lines and are ignored by PAN. PAN expects the first number to appear without a hash mark to be the number of x values and the second without a hash to be the number of y values. The y values follow next and then two columns of z and zerr grouped by their y value.

Example: In this example we have 5 xvalues and 3 yvalues resulting in 15 total data (z) points. The value of z at x=-2.0 and y=1.0 is 200. The corresponding error, zerr, is 14.1.

```
# Number of x-values
5
# Number of y-values
3
# xvalues:
-2.0000000
-1.0000000
0.00000000
1.0000000
2.0000000
# yvalues:
0.250000
1.00000
1.75000
# Group: 0
229.00000 15.132746
451.00000 21.236761
4380.0000 66.181568
455.00000 21.330729
209.00000 14.456832
# Group: 1
200.00000 14.142136
426.00000 20.639767
3260.0000 57.096410
393.00000 19.824228
219.00000 14.798649
# Group: 2
203.00000 14.247807
320.00000 17.888544
2103.0000 45.858478
333.00000 18.248288
232.00000 15.231546
```

If you were to read this data file into PAN, you would see z plotted as a function of x and use the GROUP SELECTION slider bar to change the value of y.
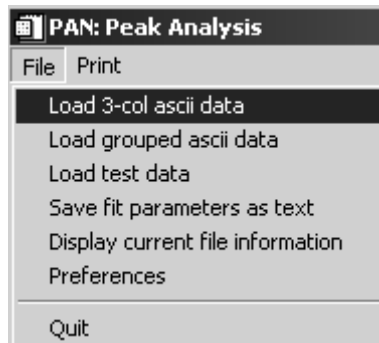
2.0   PAN Controls

Figure 1 shows the main interface for PAN. In this section we describe the controls and various menu options.

There are two menus, FILE and PRINT, located at the top of the main interface.

2.1   FILE menu

The file menu is shown in figure 2 below and the options are described below the figure.

Figure 2 File menu options.

Load 3-col ascii data
format described in section 1.4

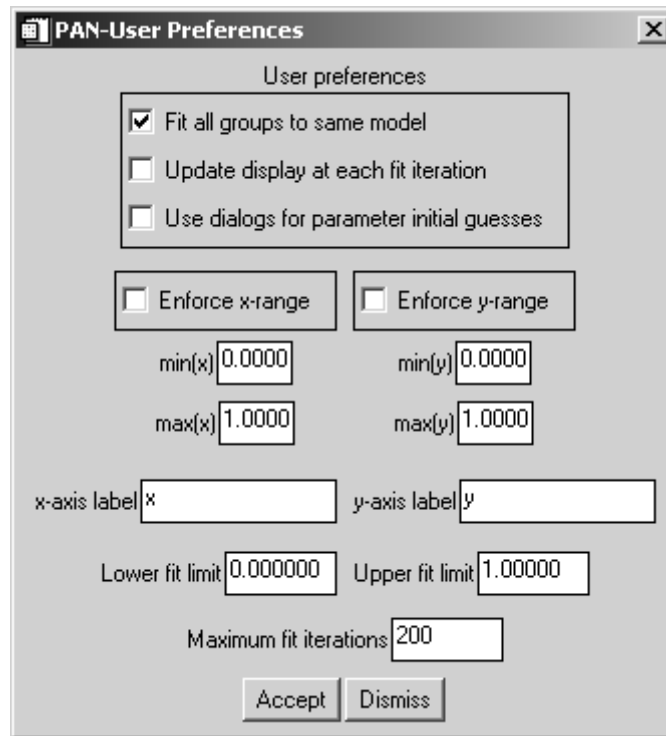Load grouped ascii data
format described in section 1.4

Load test data
used in the tutorial in section 4

Save fit parameters as text
all fit parameters, uncertainties and χ^2 values are saved into an ascii file

Display current file information
displays filename of current data file

Preferences
Launches dialog shown in figure 3. Details described below. Note that this is a persistent dialog and will close only after the user presses the Dismiss button. If you change one of the fields and press Accept, the dialog will remain but PAN's preferences will be updated.

Figure 3 Preferences dialog.

`Fit all groups to same model`
After you specify the curve components for a single data group, if you select `FIT ALL GROUPS`, each group will be fit with a model identical to the first one. Otherwise each group can have a different model.

`Update display at each fit iteration`
When this box is checked the display will be updated with each intermediate step as the fit algorithm converges to a solution. This slows down the fitting process so the default is that the box is not checked.

`Use dialogs for parameter initial guesses`
When this box is checked the user enters the initial guesses in a dialog box rather than using the mouse to specify location, widths, etc.

`Enforce x-range/Enforce y-range/min(x)/max(x)/min(y)/max(y)`
When checked, `Enforce x-range` and `Enforce y-range` cause the PAN display to show only the limits listed in `min(x), max(x), min(y), max(y)`. These text fields are editable and allow the user to change the viewing range. Note also that, when the `Enforce` boxes are unchecked, the user can change the viewing range by pressing the left mouse button and dragging a rubberband box around the desired view region. A right mouse button press results in autoscaling the range.

`x-axis label, y-axis label`
These text fields allow the user to specify the plot axes in the main PAN display.
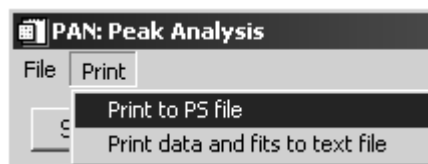
`Lower fit limit, Upper fit limit`
These text fields allow the user to change the range over which to fit the data. The fit region is delineated by vertical dashed lines.

`Maximum fit iterations`
From time to time the fit algorithm may get stuck and not converge to a solution. In this case, PAN will stop fitting when it has reached the maximum number of fit iterations. The default is 200 but the user can change this value.

2.2   PRINT menu

The `Print` menu is shown in figure 4 below and the options are described below the figure.

Figure 4 Print menu options.
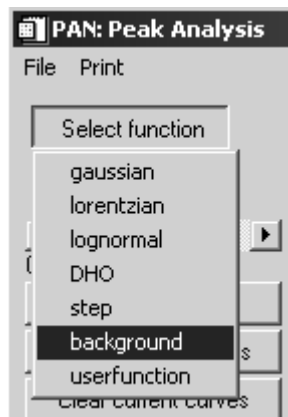
`Print to PS file`
This selection will print the contents of the main display window, the residuals window if a fit is present, the fit parameters with uncertainties and χ^2 (if fit is present).

`Print data and fits to text file`
This selection prints data from all groups into three columns (x,z,zerr) for each group if no fits are present. If fits are present the number of columns increases to $3+1+n_{curves}$. The first three columns are x,z,zerr. The fourth column is the overall fit. The remaining n_{curves} columns contain the fit components evaluated at x.
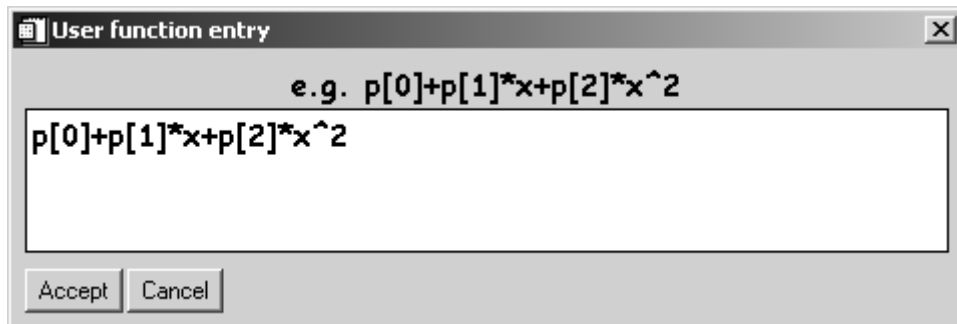
2.3   `Select function` dropdown menu

This menu, shown in figure 5, allows the user to select which curve to add to the overall fit function. After selecting one of these functions, the user is given instructions in the text display (to the right of the main plotting window) on how to specify the initial parameter values for the function using the mouse (if the preference has been set to always use a dialog then a dialog will appear).

Figure 5 Select function dropdown menu.

If the user selects `userfunction` then curve entry dialog appears (figure 6) allowing the user to enter his/her custom fitting function. Note that this custom fit function must be written in IDL syntax with the independent variable specified as x in the expression, and the parameters specified by elements of an array, p. The example shown in figure 6 is that of a quadratic equation with fit parameters p[0], p[1], and p[2].

Figure 6 User function curve entry dialog.

After pressing `Accept` the dialog shown in figure 7 appears allowing the user to enter the initial parameter values.

Figure 7 Parameter entry for user function dialog.

The remaining controls in the left hand control panel of the main PAN interface are listed below.

Group selection
This slider control allows user to view the different groups, i.e. change the y-value to y_o, say, to look at $z=z(x,y=y_o)$.

Clear all curves
If curves (or a fit) are present then this will remove all of the fits from all of the groups.

Modify fit parameters
Pressing this button launches a parameter modification dialog as shown in figure 8.

Figure 8 Parameter modification dialog.
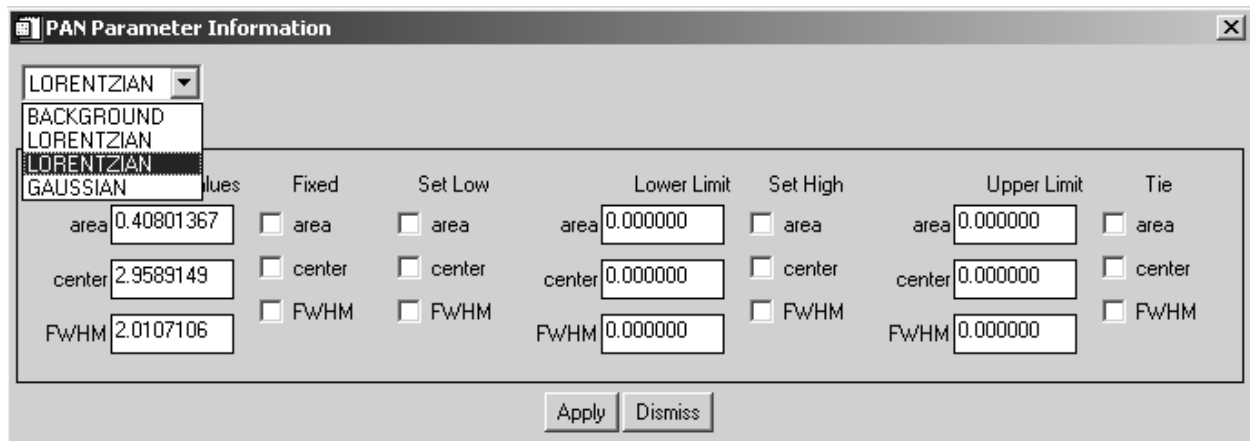
This dialog lists all of the curves in the currently selected group (as specified by the Group selection slider control. The example shown in figure 8 shows that one of the curves in the current fit is the sloping background. The current values are shown as well as a number of check boxes (all unchecked). Any of the these values can be fixed at some value by checking the box marked Fixed. Furthermore you can impose upper and lower limits on any of the parameters by checking the appropriate Set Low or Set High boxes and typing in the appropriate limits in the text field labeled Lower Limit or Upper Limit. In addition, any of the parameters can be tied together by checking the Tie box on the appropriate parameters.

When there are more than one curve in the fit you can access the other curves via the dropdown menu in the upper left corner of this dialog as shown in figure 9.

Figure 9 Parameter modification dialog showing multiple curve entries.

Clear current curves
If curves (or a fit) are present then this will remove the fit from the currently selected group.

Remove selected curve
Set the slider labeled Curve selection to the curve number you wish to remove from the currently selected group and press this button to perform this action.

Curve selection
Set the slider to the desired curve as listed in the right hand text panel of the main display (see figure 10).
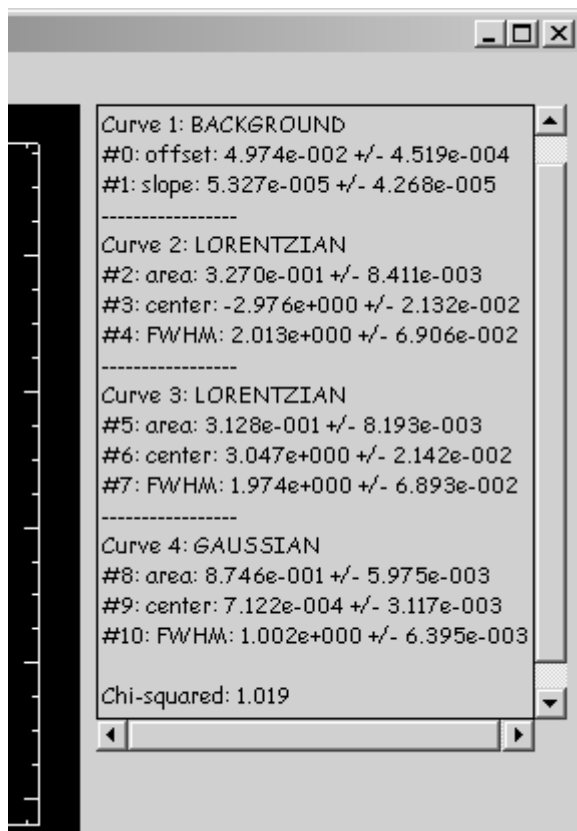
Figure 10 Right hand text panel displaying curve information.

Fit current group
Pressing this button results in the current group displayed to be fit to the model selected by the user.

Fit all groups
Pressing this button results in all of the groups listed in the text box labeled Groups to fit to be fit to the same model if this is the preference set in the Preferences dialog by the user.

Interrupt fit
If the fit "gets stuck" in a local minimum then you can press this button to stop the fit. Pressing it again resumes the fit. Note however that if it gets stuck and you don't press this then the fit routine will continue trying to fit it until the maximum number of fit iterations are reached (as specified in the Preferences dialog).

Groups to fit

This text box specifies which groups will be fit if the user presses the `Fit all groups` button.
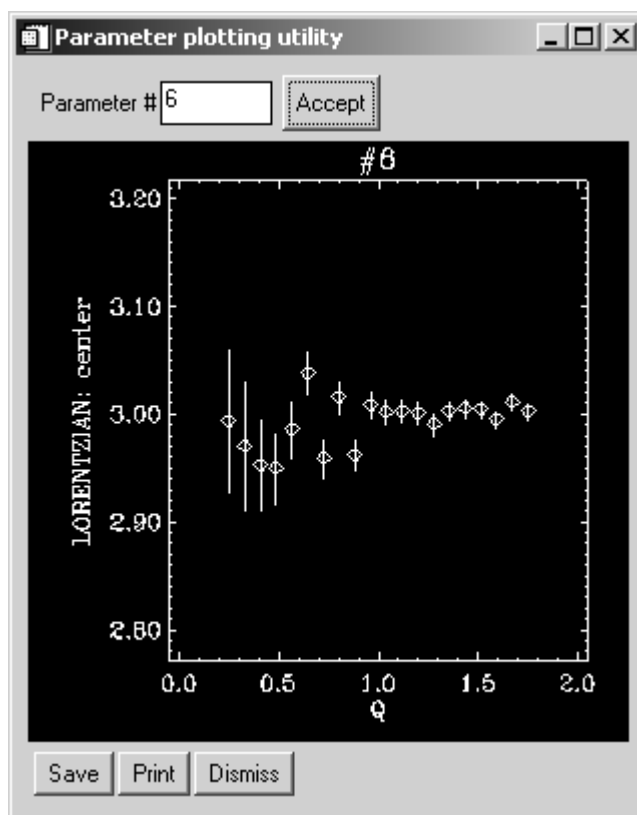
`Fitting status`
If fitting this reads "fitting" else it reads "resting". This is useful when deciding if the fit routine is "stuck".

`Display fit parameters`
Displays the current fit parameters in the text box on the right hand side of the main interface (see figure 10).

`Plot fit parameter`
If you have fit multiple groups with the same model then you can plot the group dependence of one of the parameters (see figure 11).

Figure 11 Example of plotting the group dependence of parameter #6 of the fit.

`Plot EISF`
This is a function that neutron scatterers use called the Elastic Incoherent Structure Factor. It has a similar look as the `Plot fit parameter` interface

does and the user specified those parameters that correspond to the elastic integrated intensity and the quasielastic integrated intensity.

3.0   Fit functions available in PAN

The functions available in PAN are:

Gaussian
Lorentzian
Lognormal
Damped harmonic oscillator
Background
Smooth step function
User function

Description of PAN Fit Functions

Gaussian

$$f(x; A, x_0, W) = \frac{A}{\sqrt{2\pi(W/2.354)^2}} e^{-\frac{1}{2}\left(\frac{x-x_0}{(W/2.354)}\right)^2}$$

A: integrated area
x_0:   mean of distribution
W: full-width at half-maximum

Lorentzian

$$f(x; A, x_0, W) = \frac{W}{2\pi} \frac{A}{(x - x_0)^2 + (W/2)^2}$$

A: integrated area
x_0:   mean of distribution
W: full-width at half-maximum

Lognormal

Distribution function for a random variable whose natural log is *normally* distributed.   Note that this distribution uses absolute values so that it is symmetric about E = 0.

$$f(x; A, x_0, \sigma) = \frac{A}{2|x|\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\ln(|x|-|x_0|)}{\sigma}\right)^2}$$

A: integrated area

x_0:   mean of underlying gaussian distribution

σ: standard deviation of the underlying gaussian distribution

You can use these parameters to get the mode, x_{mode}, and standard deviation, SD, of the lognormal distribution via:

$$x_{mode} = \exp\left(x_0 + \frac{\sigma^2}{2}\right)$$

$$SD = \sqrt{\exp(2x_0 + \sigma^2)(\exp(\sigma^2) - 1)}$$

The parameters x_0 and σ are actually displayed whenever you choose to display the fit parameters.

Damped Harmonic Oscillator

$$f(x; A, x_0, W) = N(x)\frac{AW}{2\pi}\left[\frac{1}{(x - x_0)^2 + (W/2)^2} - \frac{1}{(x + x_0)^2 + (W/2)^2}\right]$$

Note here we are explicitly assuming that the units of the independent variable, x, are meV so that the temperature units are Kelvin.

A: amplitude of function (NOT AREA!)

x_0:   center of each Lorentzian component

W: full-width at half maximum of the lorentzian component

N(x): Bose occupation factor defined as

$$N(E; T) = \left(1 - \exp\left(-\frac{x}{kT}\right)\right)^{-1}$$

T: Temperature in Kelvin

Smooth Step Function

This function is sometimes more useful than a sloping background if the measured background has a smooth transition from one flat level to another.

$$f(x; A, x_0, W) = \frac{A}{2}\left(1 + \text{erf}\left(\frac{x - x_0}{\sqrt{2(W/2.354)^2}}\right)\right)$$

A: Amplitude of transition
x_0:   location of midpoint of transition
W: full-width at half maximum of the transition

Sloping Background

$$f(x; m, b) = mx + b$$

m: slope
b: y-intercept

Decay

$f(x; A, \lambda) = (A/\lambda)\exp(-x/\lambda)\,\Theta(x)$
where $\Theta(x)$ is the Heaviside step function so that this function is one-sided.

A: area of function
λ: decay constant

KWW

$f(x; A, \lambda, \beta) = A\,\exp(-(x/\lambda)^\beta)\,\Theta(x)$
where $\Theta(x)$ is the Heaviside step function so that this function is one-sided.

A: amplitude of function (not area)
λ: decay constant
β: stretching exponent

4.0   Analyzing data with PAN: An example

We can analyze some simulated neutron scattering data to show how to use PAN in some detail. If you are a non-neutron scatterer, don't worry…this is just noisy, peaky data used to illustrate how to use PAN. In particular, we will load in the test data set, fit an appropriate model to the data, and extract the group-dependence of some important parameters.

First launch PAN if it is not already running. Next, select `File→Load test data`. In a few moments a data set that contains a large central peak and two small satellite peaks corrupted by noise should appear. Also, the `GROUPS TO FIT` text field should have been updated to read "1-20" indicating that there are now 20 groups (or Q-values for you neutron scatterers). You can move the slider control labeled `GROUP SELECTION` and the data displayed in the window will change. As you move to larger detector numbers, the central peak will be accompanied by two growing satellite peaks (this data simulates a rotational tunneling spectrum taken on a neutron backscattering instrument). Move the slider control so that you are viewing group (detector) 4. Click the right mouse button so that you autoscale. If you wish, you can hold the left mouse button down and drag a rubberband-type zoom box across the plot to enlarge a particular portion of the data window. A right mouse click will always result in zooming out all the way.

We will fit this group first and then fit the remaining groups automatically. The data looks like it has 4 components.   The central peak looks approximately Gaussian, the two satellite peaks are approximately Lorentzian, and there is an underlying offset (a background level). Choose `background` from the drop-menu called `Select function`. Move the cursor into the plot window and hold down the left mouse button. A dashed line should appear across the extents of the data plot. As you move the cursor up and down in the data window while holding down the left mouse button, the dashed line should follow the cursor. This action is used to specify the offset of the overall model function. Once you are satisfied with the level, release the left mouse button. Now hold down the left mouse button again and move the cursor around the data window. You should see the slope of the line change and pivot about the midpoint of the x-axis. Once you are satisfied with the slope, release the left mouse button. For this function, the slope should be quite small.
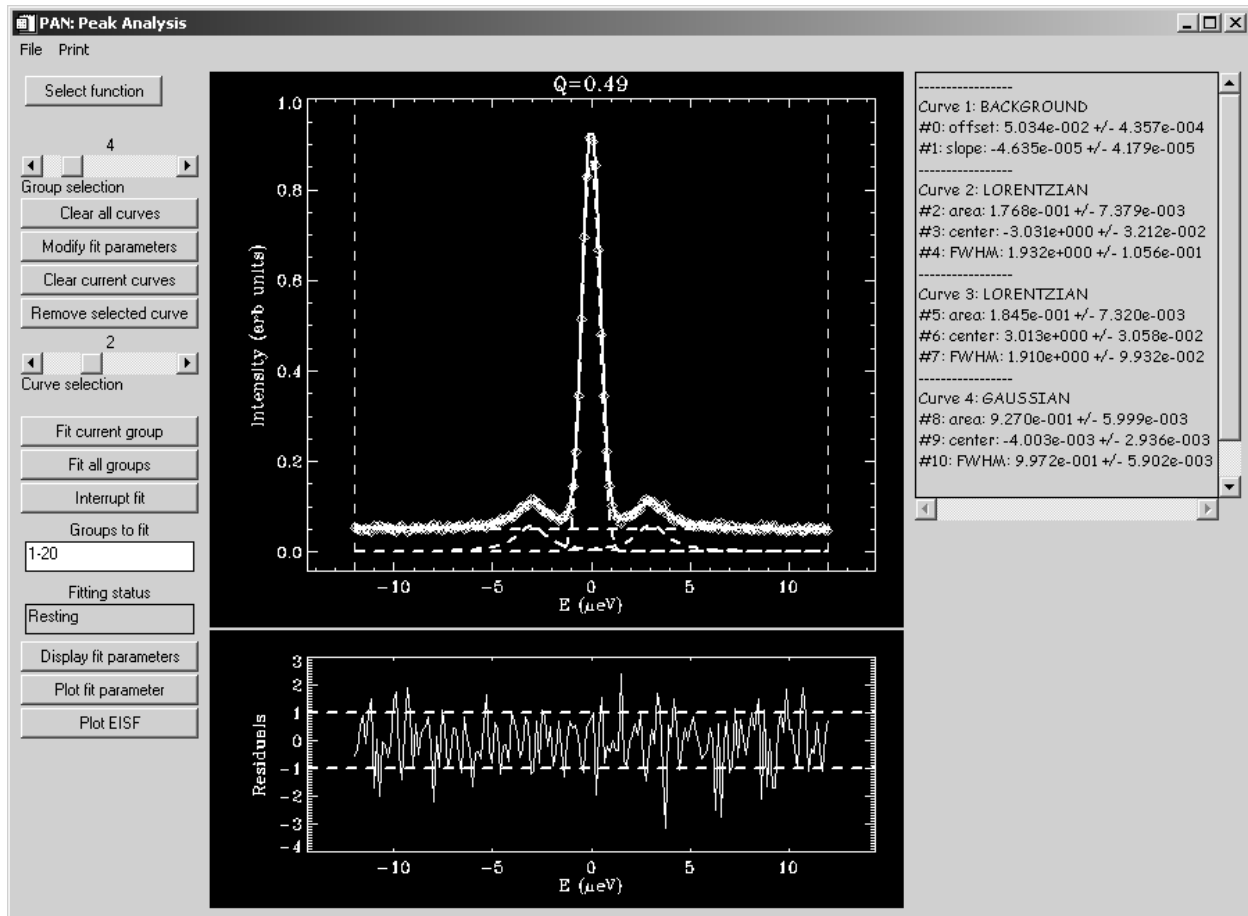
Next we will add a Gaussian curve to model the central peak. Choose `Gaussian` from the `Select function` menu. Move the cursor over the plot window and hold down the left mouse button. A Gaussian should follow your cursor movements both in location and in height. Change the vertical position to change the Gaussian peak height and change the horizontal position to change the Gaussian peak center. Once you are satisfied that the peak has the correct amplitude as that of the central data peak and is located at the correct position, release the left mouse button. Next hold down the left mouse button and move the cursor horizontally, thus changing the width of the curve. Once you are satisfied with the width release the left mouse button.

Next add a Lorentzian to fit the satellite peak to the left of the central peak. The method to change the amplitude, center, and width of the Lorentzian is exactly the same as for the Gaussian. Finally, add another Lorentzian to model the satellite peak to the right of the central peak.

You have just specified the initial guesses for the fit using the mouse and now it is time to fit the model to the data. You can look at the actual guess values by pressing the `Display fit parameters` button and the values should appear in the text box on the right hand side of the main interface. Next press the button named `Fit current group`. In a moment you should see the result of a least-squares fit of the model to the data. Your final answer may be different but mine is shown below in figure 12. You can see the residuals plotted in the bottom window which, in addition to the χ^2 value, provide a visual goodness-of-fit. The residuals are defined as

$$r_i = \frac{z_i^{\text{fit}} - z_i^{\text{data}}}{\sigma_i} ,$$

where σ_i are the uncertainties on the data points (zerr). If you have modeled the data well then the residuals should vary between ±1, as shown by the horizontal dashed lines. It looks like we have a good fit in figure 12.

Figure 12 Fit of model to data for current group.

The model parameters with uncertainties are found in the text panel displayed to the right of the data window.

The next part of the analysis we want to do is to fit all of the groups. However we need to be careful at this stage. Scroll the `Group selection` slider control down to group 1 and notice how small the satellite peaks have become. There is a possibility that the least-squares routine will have trouble fitting this same model to this data. Therefore we would like to impose a constraint on the area and width of the satellite peaks. Move the `Group selection` slider control back to group 4 and press the `Modify fit parameters` button. Set the lower limit for the areas and FWHM (full-width at half maximum) for the Lorentzians (select them using the dropdown menu as shown in figure 9). This means that you need to check the `Set low` boxes for the Lorentzians and make sure that the `Lower limit` boxes equal 0.0.

When you have finished this, press the button labeled `Fit all groups`. You should see fits being performed for each group starting with group 1 and finishing with group 20. Once it is finished you can change the `Group selection` slider to inspect each fit. Whenever a new group is displayed with its fit, its fit parameters, uncertainties, and χ^2 value are also displayed in the text panel to the right of the data window for you to examine.

Now we would like to look at the dependence of the peak position of the Lorentzian to the right of the central peak on the group (i.e. the Q-dependence of the peak position). First we should note which parameter and curve this corresponds to in terms of our notation discussed above. Inspection of the parameters (see figure 12) indicates that this parameter is #6. Press the `Plot fit parameter` button and type in 6 in the field labeled `parameter #`. Press Accept and you should see data that looks similar to that shown in figure 11. You can save these parameters and errors in a three-column formatted ASCII file, print to a postscript file, zoom into the window using the same mouse operations as in the main PAN application, and/or exit the plot utility. You may choose to examine the group-dependence of any of the fit parameters by simply specifying the curve and parameter using the technique discussed above.