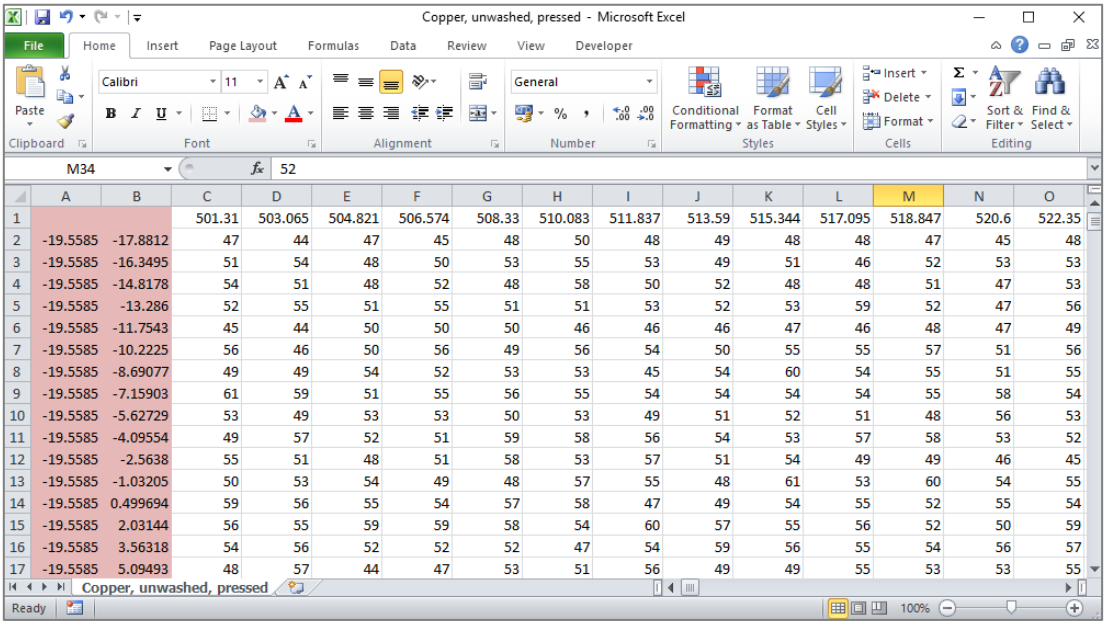
Instructions for Raman Map Fitting Program

The program is designed to open Raman Data as a .csv file, complete all peak fitting and then process the subsequent data before returning the fitted parameters and histogram as .csv files and a 2D histogram plot as a .png file. Whilst any Raman data set in the correct format can be used, it is recommended that as large an acquisition window as possible is used when collecting the map. This provides more background points, improving the quality of the fit. If using the 532nm green laser, two windows can be used that cover 500cm^{-1} – 3350cm^{-1}.

This guide will use the following convention; anything displayed by the program will be shown in **bold**, whilst any user inputs will be shown in *italics*. Any further notes that may help with errors but not necessary for operation will be shown in grey.

File Format

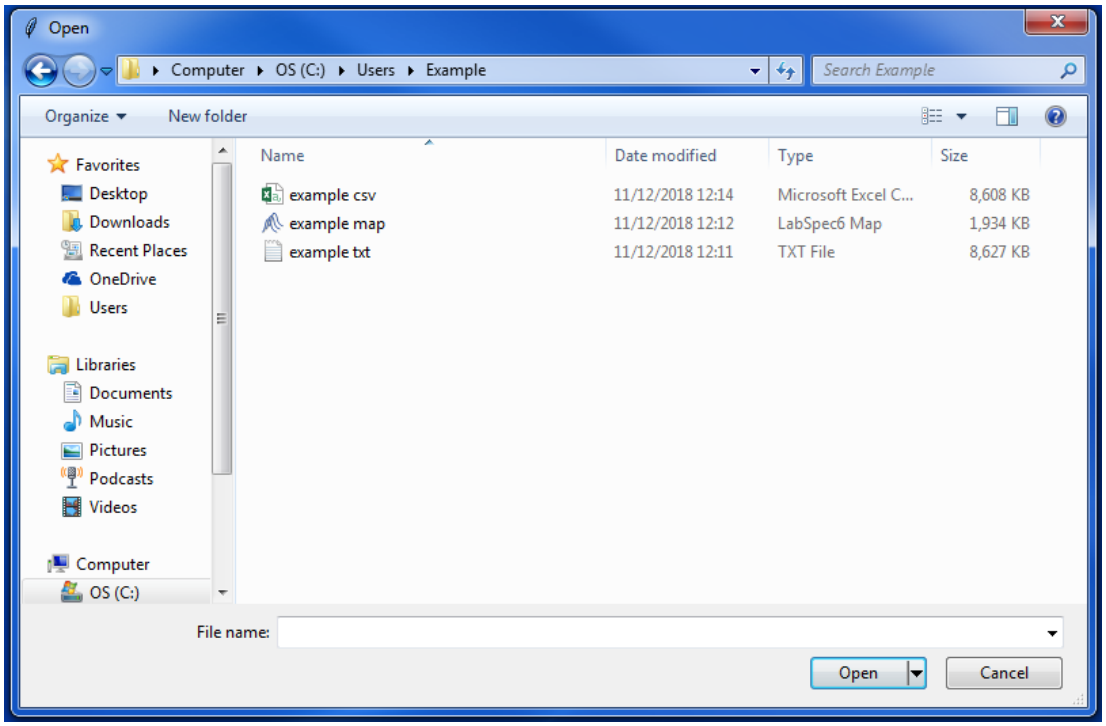
There is a small amount of reformatting required to allow the fitting program to read the data, specifically removing columns containing the spatial coordinates and saving the files in .csv format. It is easiest to do this by opening the .txt map data in excel, removing the two leftmost columns and saving the data as a Comma Delimited File.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1			501.31	503.065	504.821	506.574	508.33	510.083	511.837	513.59	515.344	517.095	518.847	520.6	522.35
2	-19.5585	-17.8812	47	44	47	45	48	50	48	49	48	48	47	45	48
3	-19.5585	-16.3495	51	54	48	50	53	55	53	49	51	46	52	53	53
4	-19.5585	-14.8178	54	51	48	52	48	58	50	52	48	48	51	47	53
5	-19.5585	-13.286	52	55	51	55	51	51	53	52	53	59	52	47	56
6	-19.5585	-11.7543	45	44	50	50	50	46	46	46	47	46	48	47	49
7	-19.5585	-10.2225	56	46	50	56	49	56	54	50	55	55	57	51	56
8	-19.5585	-8.69077	49	49	54	52	53	53	45	54	60	54	55	51	55
9	-19.5585	-7.15903	61	59	51	55	56	55	54	54	54	54	55	58	54
10	-19.5585	-5.62729	53	49	53	53	50	53	49	51	52	51	48	56	53
11	-19.5585	-4.09554	49	57	52	51	59	58	56	54	53	57	58	53	52
12	-19.5585	-2.5638	55	51	48	51	58	53	57	51	54	49	49	46	45
13	-19.5585	-1.03205	50	53	54	49	48	57	55	48	61	53	60	54	55
14	-19.5585	0.499694	59	56	55	54	57	58	47	49	54	55	52	55	54
15	-19.5585	2.03144	56	55	59	59	58	54	60	57	55	56	52	50	59
16	-19.5585	3.56318	54	56	52	52	52	47	54	59	56	55	54	56	57
17	-19.5585	5.09493	48	57	44	47	53	51	56	49	49	55	53	53	55

Running the Fitting Program

Double click to open the program as usual. The command window should open immediately but there may be a delay in initialising the open file window. Select the .csv file to be fitted; note the same file name will be used for the output**:**



Be aware that file locations saved on network drives e.g \\vfuser01.dur.ac.uk\Chemistry\, will result in an error when saving the file due to the double backslash.

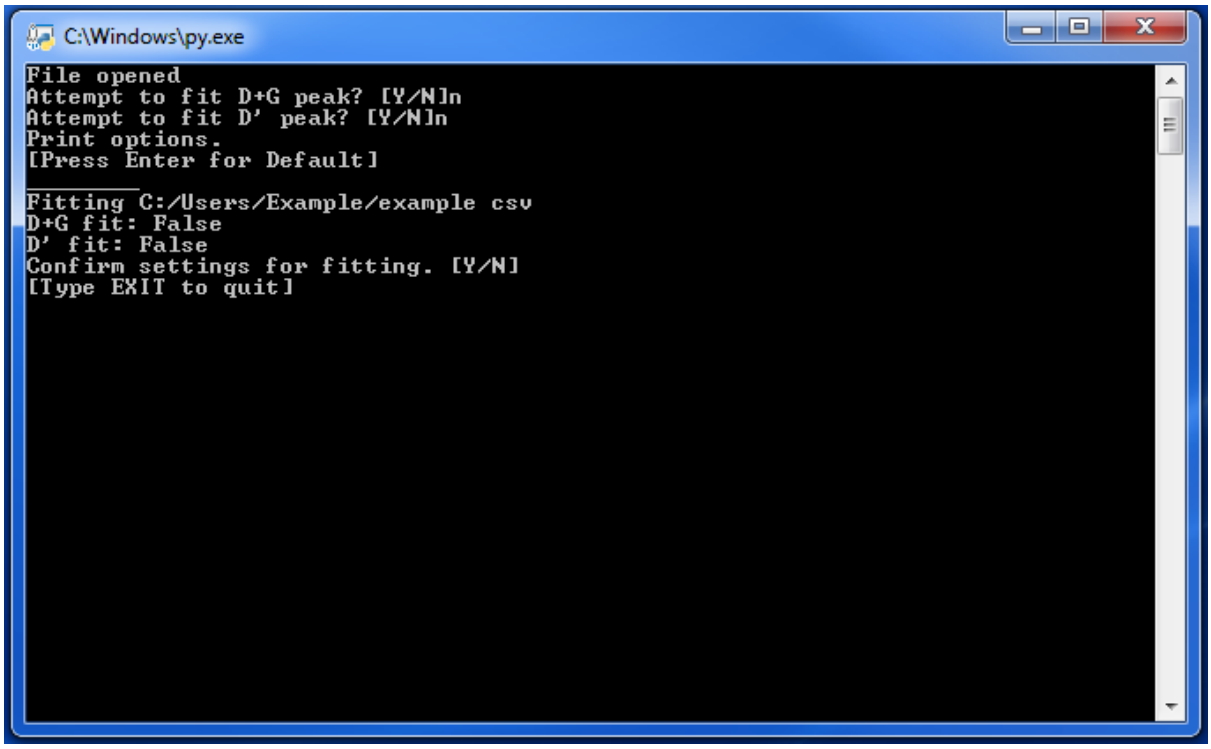
If there is an error with the graphical interface for opening files, a text-based system should appear. If the file is saved to the desktop, entering the name as displayed should be sufficient for the program to find the file. If this does not work it may be necessary to type a full file location in here. e.g *C:\Users\Stuart\Map Data\FileName*

The main control interface will confirm the file was opened and request inputs to control which peaks will be fitted. Remember that adding more peaks will increase the time taken; this effect is extreme when the additional points are of low intensity.

If fitting the D' peak, ensure a clear shoulder is present on the G peak. If no clear shoulder is present the mathematically optimum fit is often a very wide D' peak that artificially decreases the apparent height of the G peak.

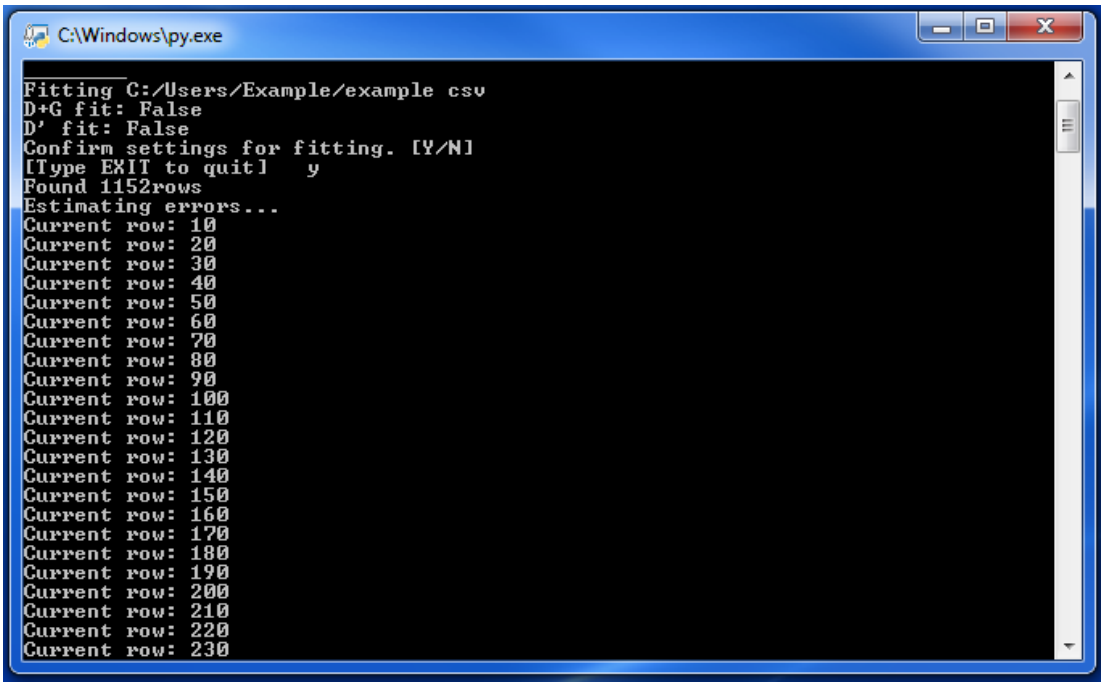
Attempt to fit D+G peak? [Y/N] *Y* (to fit the D+G peak), *N* (to not include the D+G)
Attempt to fit D' peak? [Y/N] *Y* (to fit the D' peak), *N* (to not include the D')

Print options. [Press Enter for Default]
Press Enter to continue fitting.
Typing *reports* will return a full fit report for every spectrum fitted
Typing *null_plots* will return a plot for every spectrum rejected by the program



Before commencing fitting, you must confirm the details of the fitting program.

File location
D+G fit: True or **False**
D' fit: True or **False**
Confirm settings for fitting. [Y/N]



The progress of the fitting will be shown throughout. Initially the errors of each line are estimated, shown as:
Current row: ## in multiples of 10

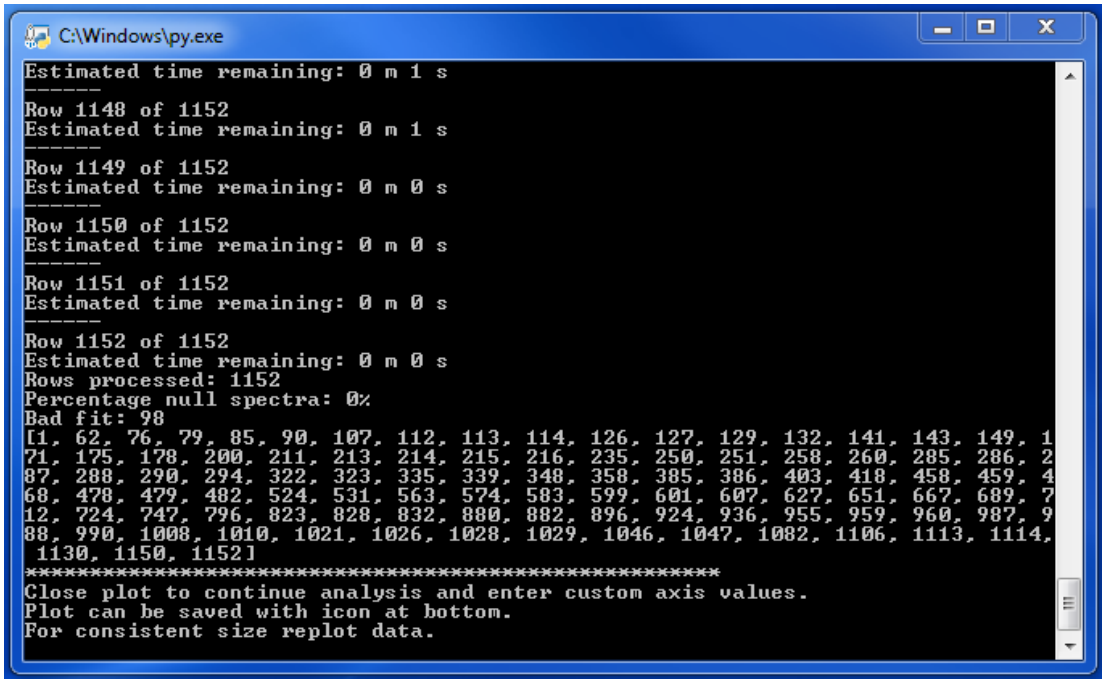
Followed by the progress of the spectrum line fitting:

Row ## of ####
Estimated time remaining: # m #s

The error estimation iterates over regions not expected to contain any graphitic peaks[1], calculating the standard deviation of 11 sequential data points. These are all be close enough together that the same value would be expected, σ is therefore an approximation of the uncertainty of each data point. The series of deviations is then averaged to produce a single value that is used as the uncertainty in every count (y-axis) data point.

The time estimation multiplies the number of rows left to fit with the average time taken to fit every previous row. As such the number will be less accurate at the start and very noisy spectra early in the fitting may artificially increase the expected time.

Once the fitting is complete the command prompt will display key outputs from the fitting and an interactive window will display the 2D histogram.

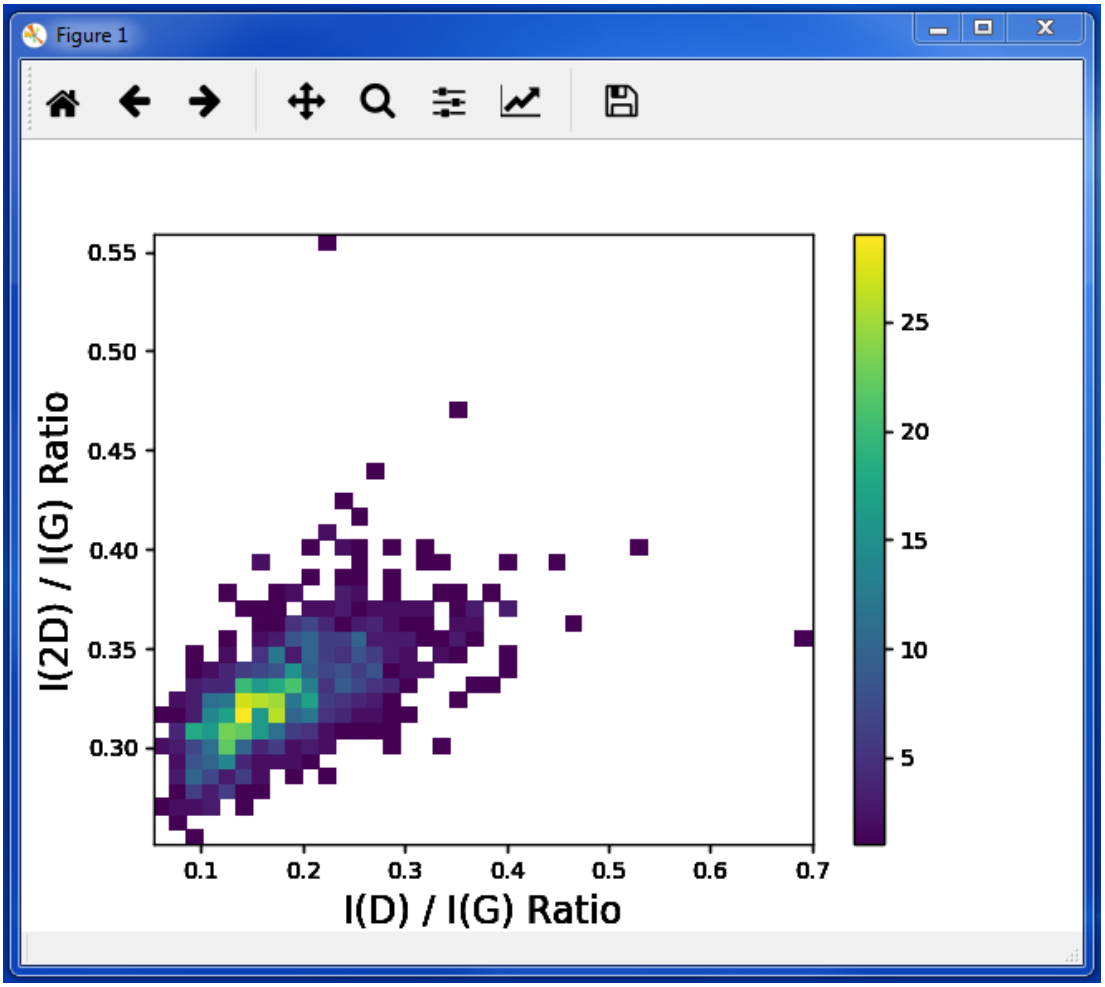


Null spectra = no peaks were clearly identifiable above background noise.
Bad fit = the signal to noise was not sufficient to extract the error data; all parameters rejected.

[1] See section on Goodness of Fit, pg.7 for exact range used.

Modifying the Histogram

The interactive window allows the plot area to be moved or zoomed and the plot saved. If the window is closed a new plot with a new axis range can be produced and saved with a consistent size to the same folder as the original file.

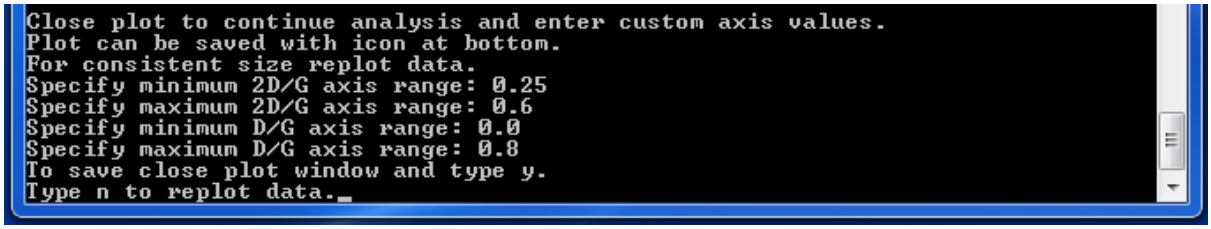


Unfortunately, you can't control the axis limits with the plot still open, so either remember or note down the axis limits required, close the plot window and enter the desired limits in the command window. Example:

Specify minimum 2D/G axis range: *0.25*

These will set the axis limits for a new 2D histogram plot that will appear once all axis limits are entered. A warning will be displayed if data lies outside the range specified.

Changing the axis in this manner does alter the size of the histogram bins. The number is constant at 40 bins, so 0.0 – 1.0 will make each bin 0.025 in size.



Upon closing the second plot window, the option will appear to re-plot the 2D histogram (for example if the figure has not been saved) or complete the fitting and export the histogram data.

To save close plot window and type y.
Type n to replot data.

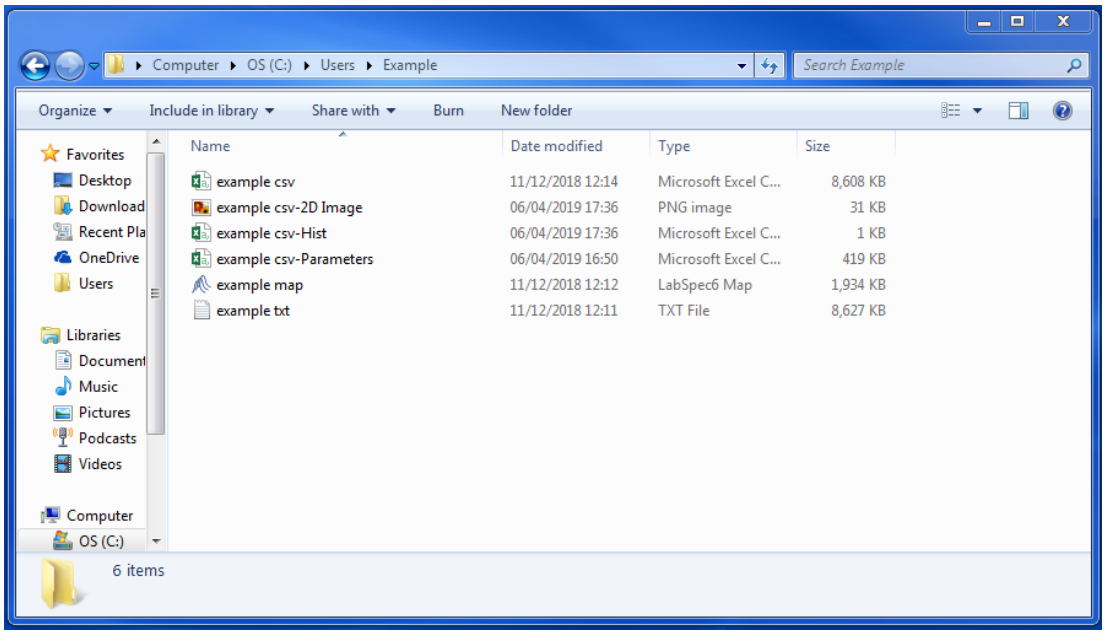
n will restart the histogram plotting, requesting axis limits input.
y will close the program, save the 2D histogram and export the histogram data as .csv files. These are saved in the same directory as the original file.

Output Parameters

The peak parameters are exported as a .csv file that can be opened with excel as soon as the fitting is completed. This means before the histogram plotting is done so any errors or accidental closure of the wrong window will not lose the fitted data.

file_name-Hist = histogram data in form of bins and counts
file_name-2D Image = 2D histogram plot in png format
file_name-Parameters = fitted peak parameters

Further details on the parameters output are included below. Any notation used in the output file are denoted as underlined.



Fitting Model

Fitting is done with Lorentzian line shapes and a 4th order polynomial. The parameters file contains the values of best fit for each parameter and the uncertainty in that value. This is established during the fitting process and is based on the maximum amount by which each parameter can be changed and every other parameter re-optimised to produce a final output within one standard deviation of the true best fit.

Lorentzian Model:

$$f(x, A, \mu, \sigma) = \frac{A}{\pi}\left(\frac{\sigma}{(x-\mu)^2 + \sigma^2}\right)$$

A = amplitude; μ = centre; σ = sigma; 2σ = fwhm; $\frac{0.3182099 \cdot A}{\sigma}$ = height

Notation for each peak:
G = G peak; D = D peak; DD = 2D peak; DG = D+G peak; Dp = D' peak

Ratios are calculated from the height of each peak and the error in the ratio from propagating the error of both heights through the division.

Goodness of Fit

Red Chi2 is the reduced chi squared value. This is a measure of the goodness of fit to data, defined as:

$$\chi_v{}^2 = \frac{\sum_x^N (y_{calc} - y_{meas})^2}{Var(y) \cdot (N - N_{free})}$$

N is the number of data points (x- coordinates); y_{calc} is the calculated value, y_{meas} is the measured value, $Var(y)$ is the variance of the y counts, N_{free} is the number of free variables in the fitting.

Reduced chi-squared is a common metric for goodness of fit. A good fit will have a value around 1, a fit that fails to account for all features will have a higher $\chi_v{}^2$ because the difference between the calculated line and measured data is large. $\chi_v{}^2$ lower than 1 generally indicates overfitted data caused by peaks shapes that are not differentiable from noise. Use caution however because noisy data can produce very promising $\chi_v{}^2$ values since the difference between the polynomial background and measured value will be included in the variance.

SignalNoise is the ratio of signal to noise. The higher this ratio the more pronounced the peaks are above the background; calculated from the height of the G peak divided by the average counts away from the known peak positions.

The background range used for average noise levels and estimating the variance is:
- 1200, 1700 - 2550, 3050 -