
This tutorial has been designed to guide the user through a few examples using *PolySNAP 2* with typical data that might be encountered in general use. It is not intended as a replacement for the full program manual, but as a basic introduction to actually using the program. It should therefore be read in conjunction with the manual itself for a more detailed explanation where necessary.

The tutorial requires the user to have already installed *PolySNAP 2*, and be familiar with Windows-based interfaces. The data files used in the tutorial are installed along with the software, and can usually be found in the *tutorial* folder in *C:\Program Files\PolySNAP2*.

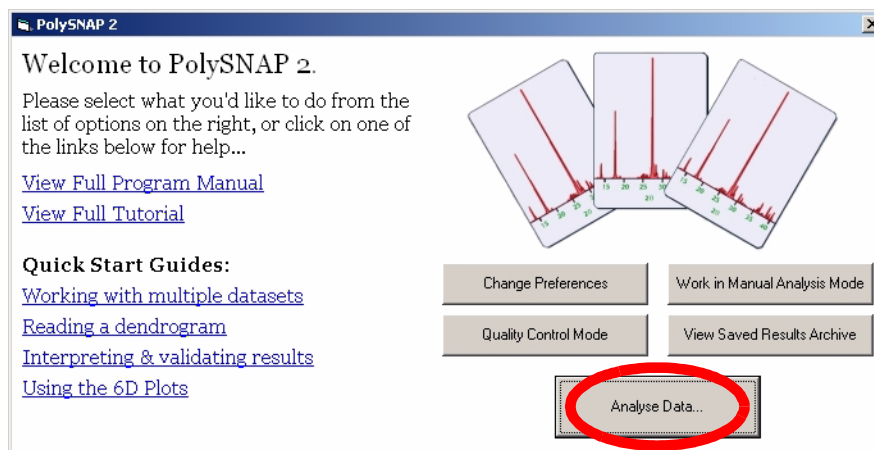
There are three main sections to the tutorial: getting used to working with a single dataset, working with multiple datasets, and working in manual mode with smaller datasets.

7.1 Automatic Analysis on a single dataset

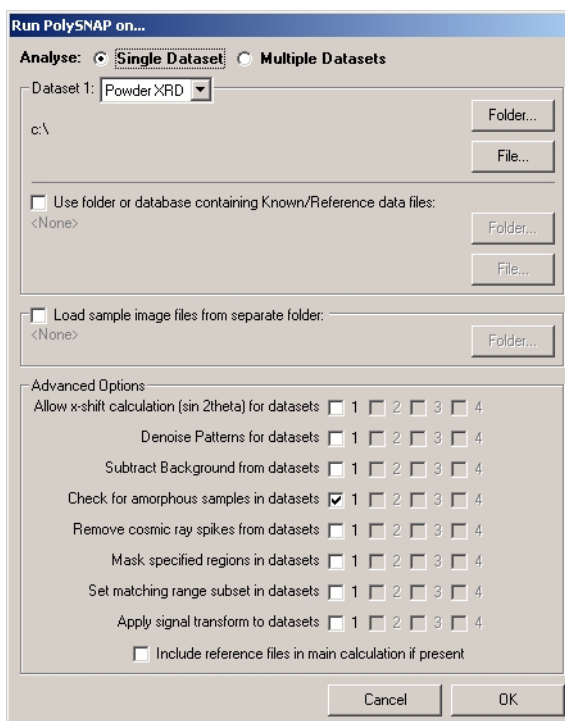
To gain experience with *PolySNAP* in automatic mode, a walkthrough of a simple run using 22 X-ray powder diffraction patterns is presented. The example assumes the program defaults are used.

To begin, launch *PolySNAP 2* from either the icon on the Windows desktop or *via* the shortcut in the Windows *Start* menu.

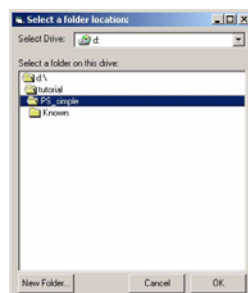
Once PolySNAP has been launched the user is presented with the *PolySNAP* welcome window.



From this, the first step is to define the input folder containing the necessary data. To do this click on the *Analyse Data* button. Alternatively, the same can be done from the *File* menu, by selecting *Automatic Analysis* and then the *Analyse Data...* option. The following dialog box should appear.



1. To specify the input folder containing the sample data files, click the *Folder* button in the top section of the dialog box. A folder-selection dialog box will appear.

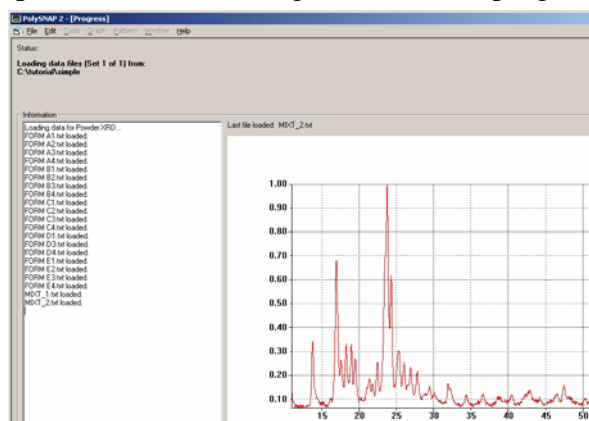


2. Select the drive containing the tutorial data from the pop-up menu at the top. In this case, select C:.
3. Navigate to the folder *C:\Program Files\PolySNAP2\tutorial\simple*. Ensure that the last folder of the desired path (*i.e. simple*) is selected by double clicking it (the folder icon should appear 'open').
4. Click *OK*. The selected path should be displayed in the upper portion of the dialog box. Check it is correct; if it is not, repeat the previous step.
5. Ensure that the other input sections has no defined folder, indicated by the word '<None>'. If this is not the case, uncheck the checkboxes labelled *Use Folder or Database containing Known/Reference data files* and *Load sample image files from another folder*. The use of these settings will be explained later.
6. Leave the Advanced Options settings to their default values (all should be off except *Check for amorphous samples*).
7. Finally, click *OK* in the bottom of the window.

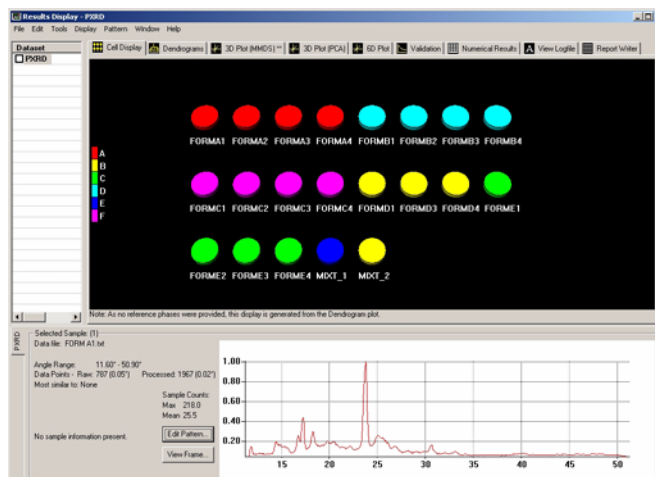
The pattern files from the specified input folder will now be loaded into *PolySNAP 2*.

If you get an error after clicking OK about no data files being found, double-check the data location specified is correct.

PolySNAP now proceeds by reading the sample files from the specified input folder, and loading these into the program.



Each pattern is loaded in turn, with the pattern profiles shown as they are read in. When all 22 patterns have been loaded, the program examines the input files, matches the data files against one another, and then performs cluster analysis. When complete the main results window will appear; this should only take a few seconds for this small dataset.

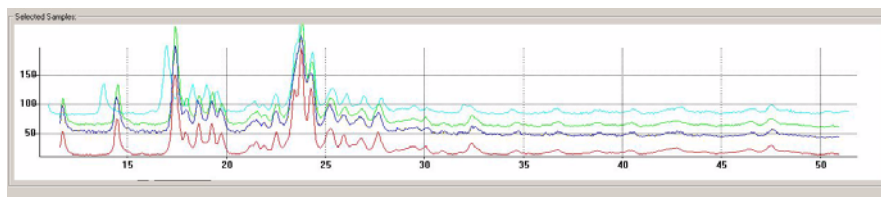


N.B. Depending on the size of your screen, the layout of the results window may not be exactly as shown in the screenshots.

There are now a number of ways to view and examine the *PolySNAP* results. The first of these - the default view displayed - is the *Cell Display*, which visually represents the contents of each sample. Each cell (shown as a disc) represents a different pattern from the input folder, with colour being used to denote the suggested grouping of compounds. In other words, similar samples are given the same colour. In the lower part of the window the sample information and pattern display of the selected cell is available. The first sample cell is selected automatically when the window is first opened.

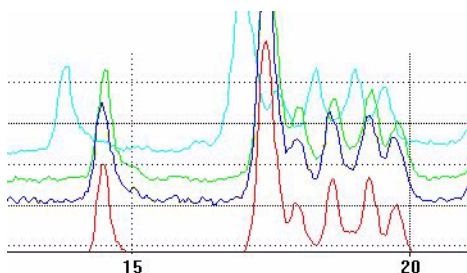
The program can be set to display a certain number of cells per row, so that the display represents, for example, the layout of a 96-well sample plate.

1. In the *Cell Display* click on cell *FORMD1*. The associated full pattern profile of pattern D1 is now displayed in the lower region of the window beside the sample information. In this particular case the information available is limited, so many of the fields are blank.
2. Hold down the *shift* key and click on cell *FORMD4*. The cells from D1 to D4 are now selected, and the sample information pane now disappears, leaving only an extended view of the profiles of patterns D1, D3 and D4 overlaid to allow a visual comparison. From this view they are all obviously the same compound.
3. From the *Display* menu select *Offset Overlaid Pattern Profiles* and click *on y-axis*. This will now display the multiple patterns with an offset along the y-axis of the plot, when any further cells are selected.
4. Hold down the *control* key and click on the other cell coloured yellow, *MIXT2*. The display now includes this new fourth pattern, and to allow an easy comparison are displayed overlaid with an offset along the y-axis.



5. Zoom into the area between around 15° and 20° by holding down the left mouse button and dragging a box over the desired area. When the mouse button is released, the graph region is redrawn to show just the selected area. Zooming in and out smoothly can also be accomplished by clicking in the display and then holding down the *control* key while moving the scroll wheel of the mouse (if available).
6. To move around the area and position the view more accurately, hold down the *Alt* key and *left mouse button* simultaneously while dragging with the mouse - the display updates in real time.

These zoom and movement functions are the same for all graphical displays within the *PolySNAP* program.



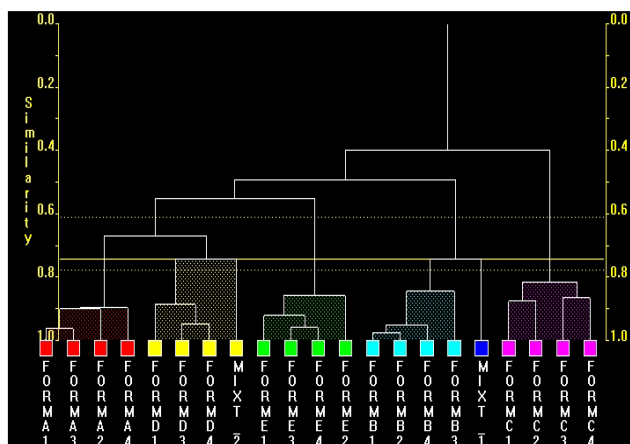
To check which of the plotted profiles correspond to which sample, 'hover' the mouse pointer over part of the plotted line; a tooltip appears with the sample filename in it.

This closer view of the overlaid patterns makes it easier to see that the top pattern (MIXT2) is noticeably different from the other patterns - note the extra peak around 14° for example - and, despite being coloured the same, may not actually belong to this group of compounds.

7. Reset the view with a *right click* of the mouse in the graph pane and selecting *Reset View* from the pop-up menu.

There are a series of tabs just below the menu bar, each of which display a different view of the data. Select the tab labelled *Dendrogram*.

The partitioning of the data into groups that were displayed in the coloured cells is carried out by the cluster analysis of the sample data. In *PolySNAP* there are five different methods of clustering available, each of which tend to give slightly different results. The program computes the best, in the sense of the most internally consistent, dendrogram method and displays the results from that (although the user can choose to view the results from the other methods and overrule the selected one if required).



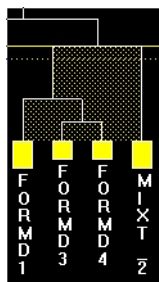
A dendrogram provides a visual display of the results from the hierarchical method of data classification using cluster analysis. The dendrogram itself takes the form of a tree-diagram in which each terminal branch (coloured box) is representative of a single pattern sample.

The higher up the similarity scale two samples are connected by a horizontal line (called a 'tie-bar'), the less similar they are. Hence samples FORMA1 and A2, with a similarity value of around 0.9 are very similar, whereas samples FORMA1 and FORMC1, which are only joined much further up the tree by a horizontal line with a similarity value of around 0.4, are quite different.

In this dendrogram there are 6 separate clusters, each distinguished by its own colour. These are the same colours displayed earlier in the cell display and throughout most of the other *PolySNAP* displays. The number of clusters are defined by the yellow cut-line which in this case was initially set to 0.708. The calculation of this level is *via* a number of statistics. The confidence levels on this choice of cut position are shown by the yellow dotted lines either side. Selecting a cut-line for a dendrogram is a difficult procedure, and the results must be treated with caution. The program-calculated level therefore should always be carefully examined by the user to see if looks sensible.

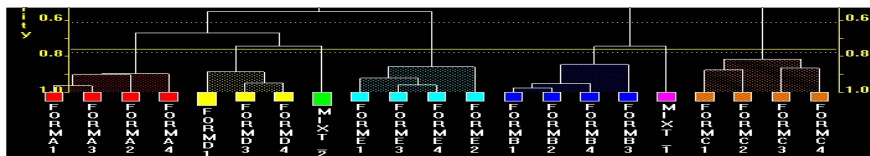
Adjusting the cut level upwards creates fewer separate clusters, and effectively reduces the discrimination between differences; adjusting the cut-level downwards creates more separate clusters.

1. In the *dendrogram* click on the yellow square cell *FORMD1*. The pattern profile of the sample is displayed.
2. Using the *control* key, also select cell *MIXT2*. These are the two patterns that appeared different from one another when they were overlaid in the cell display.

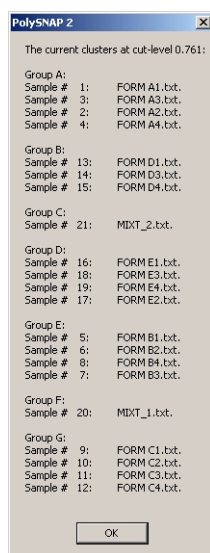


Looking at the position of the cut-line it is seen to be only very slightly above the similarity line between cell *MIXT2* and the other yellow cells. With the confidence levels indicated by the yellow lines being between 4 and 8 clusters, it is possible that the present level is not ideal.

- To manually adjust the cut-line, either click in the Dendrogram area and use the *scroll wheel* or hold down *control* and the *left mouse button* while dragging up or down. Move the cut-line down slightly so that it is still above the lower confidence line but so there are now 7 different clusters. Notice that the assigned colours change, and that *MIXT2* is now in clusters of its own.



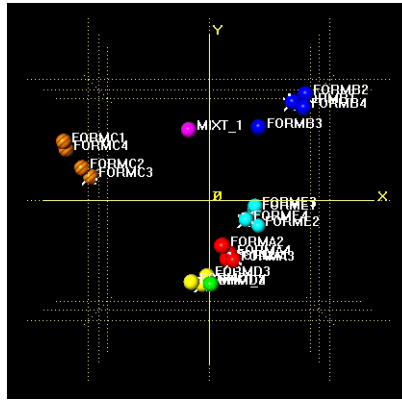
- In the dendrogram area click the *right mouse button* and select *Save Modified Trees....*, to ensure that the changes are retained.
- From the *Tools* menu select *List Pattern Cluster Members....* A dialog box will appear containing a list of the clusters.



This dialog box displays the different clusters with their sample numbers and actual file names. From the file names used in this demonstration example it can be seen that the samples are now all properly grouped, with the two mixtures separate from the rest. Notice in the dendrogram display that these two mixture patterns are quite dissimilar to anything else, having a low similarity connection value to their neighbours.

The differences between samples can also be viewed by making use of the distances derived from the correlation matrix to give a representation of the data in three dimensions. There are two methods used for calculating the resulting 3-D plots: these are *Metric Multi-Dimensional Scaling* (MMDS) and *Principle Component Analysis* (PCA). They both give different views of the samples because of the differences in calculation and will therefore give slightly different results. The control of each plot is the same so only the MMDS will be described here in detail.

Select the tab *3D Plot (MMDS)* (or PCA if you wish).



The initial view shows only the x and y axis, while the z axis lies in projection. Each point represents a sample. The position on the plot is taken from the MMDS calculation. The colour of each sample is taken from the dendrogram display to allow easy comparison of the results from these different methods. Allowing the mouse to hover over a sample displays the sample label in a tooltip popup.

Samples that are similar are plotted close to one another, so are seen to clump together in groups. Note that different coloured samples can also be close, this shows that they also have similarities. This can be seen by the yellow group being very close to the green sample, which is pattern MIXT2. When comparing with the dendrogram display, the MIXT2 sample is only separated from the yellow group by the current level of the cut-line.

Also notice the number plotted at the top-left of the display, in this case, 0.96. This is a correlation coefficient called the goodness of fit, that measures the quality of the 3D representation to the original data. The closer to 1.0 it is, the greater the reliability of the results. This value tends to decrease when larger data sets are used.

Use the following methods for exploring the 3-D plot.

Action	Control
Rotate the 3-D plot	Drag while holding shift key and left mouse button
Move plot laterally	Drag while holding alt key and left mouse button
Alter size of spheres	Drag up or down while holding control key and left mouse button
Zoom on centre	Click in area, hold shift key and use mouse scroll wheel

Zoom on area	Hold left mouse button and draw box over area
Select a sample	Click on sphere
Select multiple cells	Hold control key and select additional cell
Alter rendering quality	Press F2 and adjust scale (lower values are better for slower graphics cards)
Centre view	Right click mouse and select <i>Centre Selection</i>
Reset view	Right click mouse and select <i>Reset View</i>

The 3D plot is useful to spot patterns that are quite different from the others, as they tend to stand out on their own and are not easily grouped. Also, cases where the colours (from the dendrogram) and the positions (from the MMDS) of the samples appear to contradict each other are the samples that should be looked at manually in more detail.

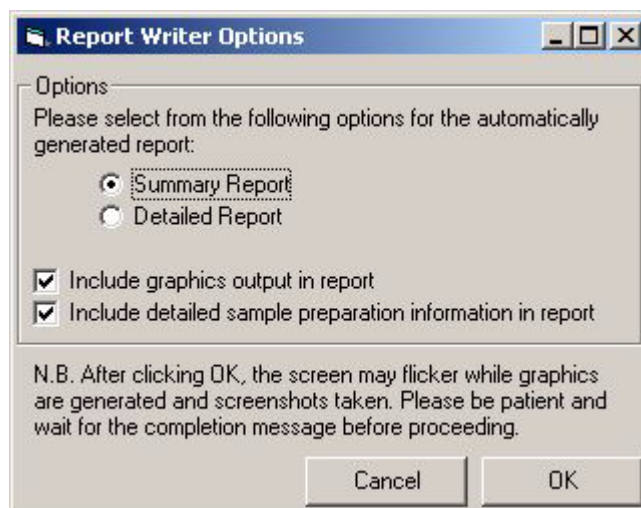
Finally, notice that some of the spheres have white 'spikes' protruding from them; these show which samples are the Most Representative Patterns of that cluster. Clicking on one brings up a dialog box showing the mean pattern-pattern distance for that cluster. The smaller the distance, the tighter the cluster, and the more similar the samples are within it.

7.1.1 Creating a Report describing the results

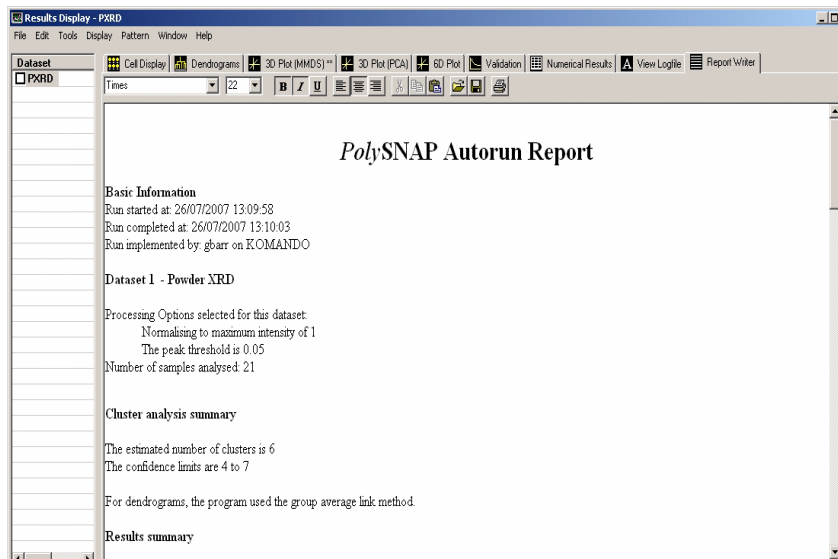
A basic report of the calculated results can be automatically generated using PolySNAP. This contains the information gained from the data and optionally some of the graphical views of the results that have been obtained.

1. Select the *Report Writer* tab.

- From the *Tools* menu select *Generate Report....* A dialog box will appear.



- Select *Summary Report*. Deselect the checkbox labelled *Include detailed sample preparation information*, and then click *OK*. The report will now be generated and screenshots taken. This may take a few moments.
- A Dialog box will appear confirming that "Report generation is now complete." Click *OK*.



The report begins by detailing when the calculations were done, who was the user and on which computer. It then continues with the settings used for the analysis and a summary of the results, including the cell display and dendrogram. If the *Detailed Report* option had been chosen, a more detailed report is presented, including output from all of the main graphic displays. The report may be edited, or additional information can also be added to the report manually, as in

a standard word processor, by using the simple formatting tools supplied.

For example, to add one of the profile plots to the report, switch to the Cell Display, and select say Form A1. In the profile pane at the bottom, right-click in the graph area and choose *Copy* from the pop-up menu. Then select the Report Writer tab, click in the report at the point where you want the image to be inserted, right-click and select *Paste* from the pop-up menu.

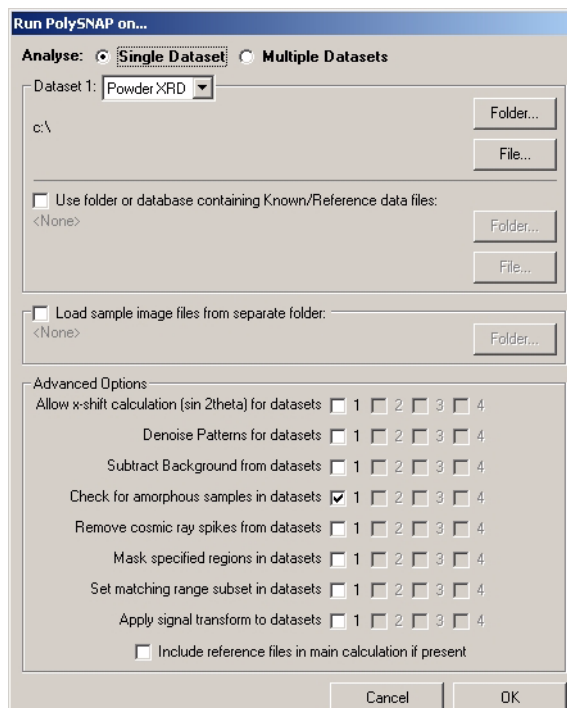
5. To save the report, click the *right mouse* button and select *Save as...*
6. A dialog box will appear. Use this to navigate to a suitable folder where the data can be saved and select a file name. *e.g. C:\PolySNAP Report.rtf.*

The report is now saved as an RTF file and can be opened and edited in any standard word processing package.

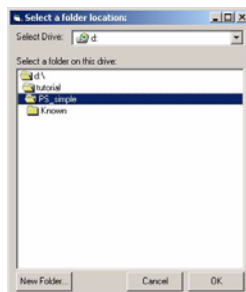
To end this session select *Close Window* from the *File* menu. If prompted to save changes to the report, select *No* because this has already been done. A second window will open asking if the results from the current run are to be saved. This would save the current results as a *PolySNAP* archive file (*.psnaparchive*) that could be accessed later, but this is not necessary at the moment. Select *No*.

Instead we will now run this dataset a second time, but this time there will be extra information and options used to give a more advanced understanding of the dataset. To begin, click on the white PolySNAP window background to bring up the welcome screen, and click *Analyse Data*, or select *Analyse Data...* from the *File -> Automatic Mode* menu.

The following dialog box should appear.



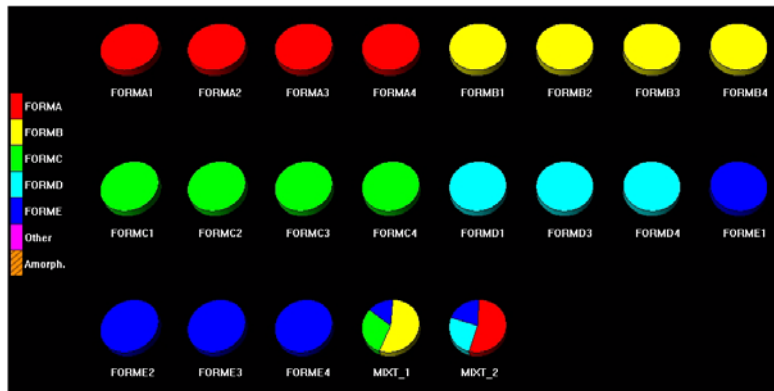
1. To define the input folder containing the sample data files, click the *Folder* button in the top section of the dialog box. A folder-selection dialog box will appear.



2. Select the drive containing the tutorial data from the pop-up menu at the top. In this case, select C:.
3. Navigate to the folder *C:\Program Files\PolySNAP2\tutorial\simple*. Ensure that the last folder of the desired path (i.e. *simple*) is selected by double clicking it (the folder icon should appear 'open').
4. Click *OK*. The selected path should be displayed in the upper portion of the dialog box. Check it is correct; if it is not, repeat the previous step.
5. Click on the checkbox next to the option *Use folder or database containing Known/Reference data files*. The *Folder* button can then be selected. Navigate to the *C:\Program Files\PolySNAP2\tutorial\simple\Known* and double-click to select this folder.

6. The *Load sample image files from another folder* section should still be empty and unchecked.
7. Finally, click *OK* in the bottom of the window.

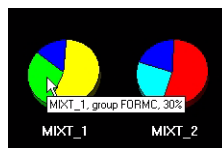
After the analysis has been performed the results open with the new cell display:



By selecting an extra option and including known reference files three important changes have been made to the cell display.

The reference files that were provided to the program are known patterns of pure phases that the samples in the dataset are to be compared to. Because of this the cells in the cell display are now coloured according to the known phases that they match and not according to the results from the dendrogram.

The presence of reference files has a more profound affect on the mixture samples.



Instead of being assigned to a particular cluster they are now analysed using the known phases as a reference, and how much of each known, pure phase present in each mixture is calculated. The cell is now displayed as a pie chart, divided up and coloured according to the percentage of the mixture that each known phase is thought to account for. Holding the mouse cursor over a slice of the pie chart will open a tooltip box displaying the phase, and percentage that the slice represents.

Finally, there are two extra keys that have been added onto the cell display legend.



Each known phase has its own entry, however there is now also a key called *Other*. Any pattern that did not match any of the known phases provided above a set similarity cut-off would be assigned here.

NB. - the *Other* group is not a group of similar patterns, but a collection of patterns that do not match the standard reference phases. However none of the samples in our dataset are in this group as they all match one of the phases, apart from the mixtures.

Switch to the *Logfile* tab. Scroll down through the information to the section headed *Quantitative Analysis*. The numeric output from the analysis can be found here.

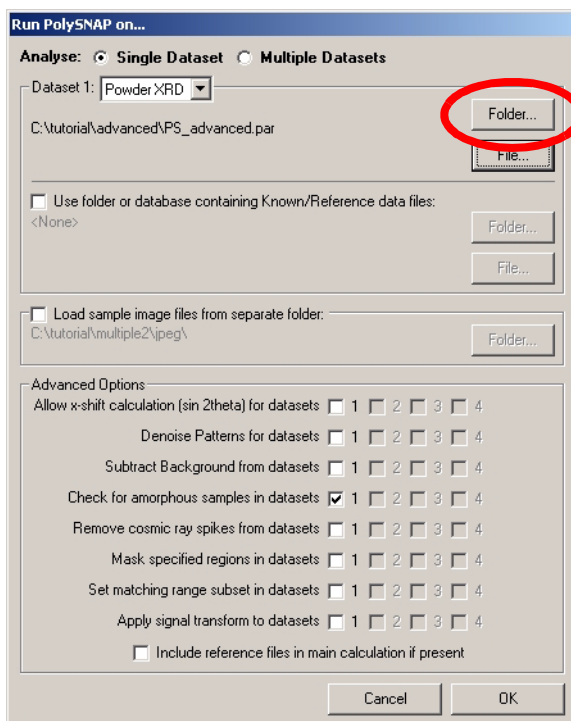
You can now close the Results window, as the first section of the tutorial is now complete.

7.2 Amorphous and Pattern Shifts

This part will cover some of the more advanced options available that include pattern matching while allowing for a 2θ -shift, and automatic identification of non-crystalline samples.

7.2.1 Loading Pattern Data from a Database

From the *File* menu, select *Automatic Analysis* and the *Analyse Data...* option. The following dialog box will appear.



1. To define the input database containing the sample data, click the *Folder* button. A file-selection dialog box will appear.
2. Select the drive containing the tutorial data, in this case *C:*.
3. Navigate to the folder *C:\Program Files\PolySNAP2\tutorial\advanced*.
4. Ensure that no known phases are used by making sure the relevant checkbox is turned off.
5. Leave the Advanced Options settings at their default values.
6. Click *OK*.

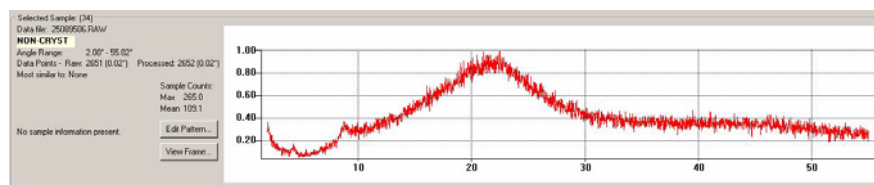
The pattern files in the specified folder are now loaded into PolySNAP, where it checks for amorphous samples, matches the patterns with one another, and performs a cluster analysis. This may take a little time to complete, but once finished the results window will appear with an initial view of the *Cell Display*.

7.2.2 Analysis of the Results

The 35 samples contained within the database are now presented in the *Cell Display*.

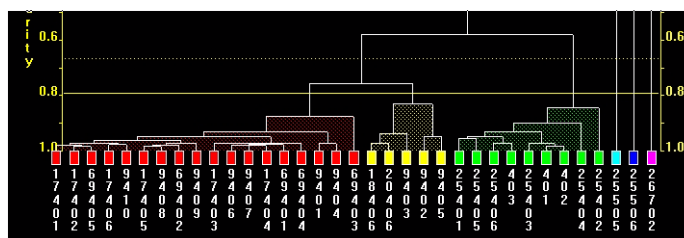


The program is using the last few digits of the filename of each sample to label them. Selecting pattern number 25505, the pattern information display shows that it has been labelled as a non-crystalline sample, which seems reasonable given its profile.

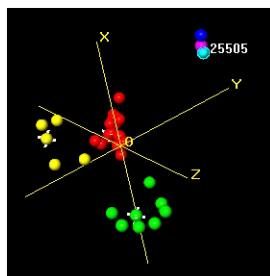


The other two patterns, 25506 and 26702 are similarly labelled for the same reason. Identification of such amorphous samples is done on the basis of checking to see if any signal (corresponding to peaks) would be left after subtraction of the entire amorphous hump. The method tends to err on the side of caution.

Looking at the *Dendrogram* tab:

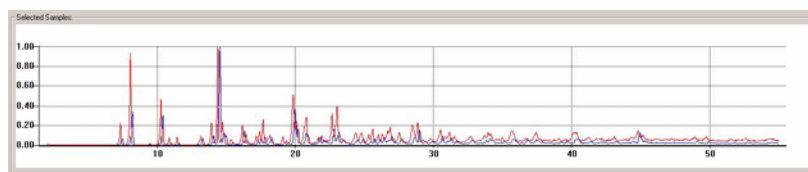


It is seen that three non-crystalline labelled patterns are placed on the far right of the diagram with a zero similarity to the rest. This is deliberate, in order to remove them from the main clusters. Compare this to the 3D (MMDS) plot:



Again, the three non-crystalline samples are quite separate from the rest of the patterns. Both the dendrogram and 3D plots suggest a loose grouping of the rest of the patterns, suggesting there may be some differences between them.

Using the *control* key, click two patterns which are on opposite sides of the dendrogram, for example patterns 17401 and 25402. Examining their profiles, it appears that in addition to some preferred orientation issues, there seems to be a noticeable 2θ -shift between the otherwise relatively similar profiles:



Generate a full detailed report of the data including the graphics by selecting the *Generate Report* option from the *Tools* menu.

This completes the initial analysis of the data, but the 2θ -shift in some of the samples could be examined in more detail. From the *File* menu select *Close Window*, and do not save the results.

7.2.3 Reprocessing the Data allowing for an x-shift

When collecting powder diffraction data from a diffractometer the sample or instrument alignment can result in linear or non-linear shifts along the x-axis of the resulting pattern. This can especially be a problem if the sample height varies from sample to sample, giving rise to systematic errors in the pattern matching unless it is accounted for. However, to allow for this is a time consuming process and should therefore not be used unless such a shift is suspected - it is switched off by default.

A general expression for the shift is:

$$\Delta(2\theta) = a_0 + a_1 \sin \theta \quad (7.1)$$

where the a_0 coefficient corresponds to a linear (zero-point) shift described earlier, and the a_1 coefficient a non-linear component [Zevin & Kimmel, 1995]. The requirement then is to find values of a_0 and a_1 that results in a maximum matching correlation result between two patterns.

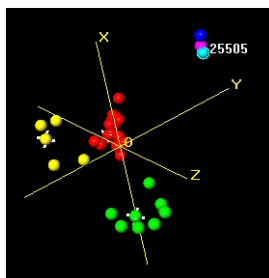
The same data as in the previous run will now be examined again. Unlike the last time, the program will vary the x-offset parameters to attempt to maximise the match result.

Define the input database, known phases and output folder as before:

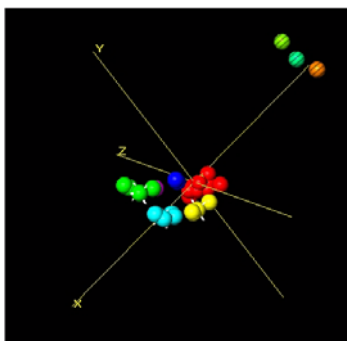
1. From the *File* menu, select *Automatic Analysis* and the *Analyse Data...* option. Define the same input data folder as used before in the previous run of PolySNAP.
`C:\Program Files\PolySNAP2\tutorial\advanced\`
2. Select *None* for the known phases directory.
3. In the Advanced Options area, turn on the *Allow x-shift calculation (sin theta)* checkbox.
4. In the *Run PolySNAP on...* dialog box click *OK*.

The same data as before is now run allowing x-shifts on the patterns. The time required to process this will take longer than a normal run of PolySNAP.

Now look at the 3D MMDS plot. Previously, it looked like this:

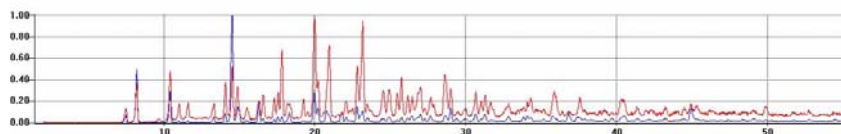


Now, with the option to calculate the best-offset value for each pattern turned on, it looks like this:



The three non-crystalline patterns are still quite separate, but the rest of the patterns have condensed together as a result of allowing for the 2 θ -shift, showing that a large part of the differences between the pattern profiles was due to variation in sample heights during data collection. The program still separates them out within this grouping due to the preferred orientation problems.

Similarly with the dendrogram display, the similarity values between the patterns are much improved. The remaining differences appear to arise from preferred orientation effects, which are quite noticeable in some cases. For example, overlay the profiles of samples 401 and 69402:



To see how much the patterns have been shifted as part of the calculation, switch to the Numerical Results tab. Locate 401.txt in the list on the left hand side, and read along to find where this row crosses the 69402.txt column:

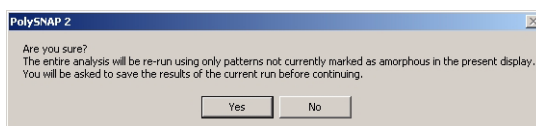
	Cell Display	Dendrograms	3D Plot (MMDS)	3D Plot (PCA)	SD Plot	Validation	Numerical Results
Rank:	26702.txt	401.txt	402.txt	403.txt	69401.txt	69402.txt	6940
17401.txt	0.4489	0.9309	0.9404	0.8988	0.9606	0.9773	0.9
17402.txt	0.4664	0.9203	0.9309	0.8735	0.9677	0.9773	0.9
17403.txt	0.4786	0.9045	0.913	0.8046	0.9718	0.952	0.9
17404.txt	0.4767	0.9298	0.9333	0.892	0.9731	0.9789	0.
17405.txt	0.4562	0.9259	0.9299	0.8894	0.9519	0.9783	0.
17406.txt	0.4231	0.9154	0.9181	0.8849	0.9367	0.9569	0.9
18406.txt	0.2878	0.8174	0.8103	0.8911	0.7261	0.7386	0.7
20406.txt	0.3303	0.8472	0.8431	0.9104	0.7823	0.7881	0.7
25401.txt	0.2852	0.9106	0.9036	0.9565	0.756	0.7902	0.7
25402.txt	0.4405	0.963	0.9634	0.9121	0.9381	0.9442	0.9
25403.txt	0.391	0.9825	0.9783	0.9503	0.932	0.926	0.9
25404.txt	0.2977	0.9167	0.9124	0.9779	0.8151	0.8137	0.7
25405.txt	0.3293	0.9689	0.9605	0.982	0.8568	0.8732	0.8
25406.txt	0.2944	0.9313	0.9261	0.972	0.7896	0.8119	0.7
25505.txt	0.9359	0.4014	0.4018	0.3509	0.4543	0.4738	0.4
25506.txt	0.9224	0.331	0.3318	0.2844	0.4584	0.4503	0.4
26702.txt	1	0.3747	0.374	0.3261	0.454	0.454	0.4
401.txt	0.374	1	0.9627	0.856	0.856	0.856	0.8
402.txt	0.3261	0.9638	1	0.8573	1	0.8573	0.8
69401.txt	0.491	0.8854	0.8952	0.8573	1	0.963	0.9
69402.txt	0.4756	0.9137	0.9135	0.8624	0.963	1	0.9

Clicking on the rank value shows the two profiles overlaid in the bottom pane; hovering the mouse over this value shows a tooltip with the calculated values of a_0 and a_1 for this pair of patterns.

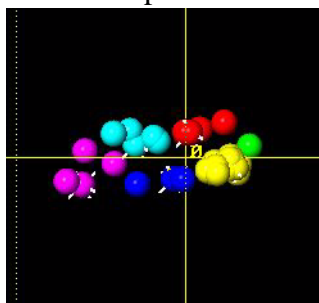
Create an auto-report on the run. This report will not have detailed any manual changes that might have been made to the dendrogram cut level. Therefore any information that has been changed can be included by adding them manually or copying from the *Logfile* pane, which details all the calculations and changes made during the PolySNAP session. Note that the use of an x-offset calculation has been noted in the output, as has the identification of the non-crystalline samples.

In cases where there are a larger proportion of amorphous samples in a dataset, it may be helpful to remove these samples and re-run the analysis without them to give a clearer idea of what is going on. This can be done automatically by selecting the *Tools* menu item *Rerun Analysis*, and from the sub-menu selecting *Ignoring All Non-crystalline Samples*.

A dialog warning box appears.



Select *Yes*. After being prompted to save the current results, a new analysis is performed, and a new results display window is opened as before. Note that this time, there are no amorphous samples visible in either the dendrogram or the 3D plots.

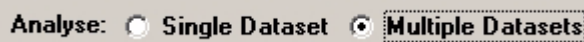


7.3 Multiple samples - Dataset 1

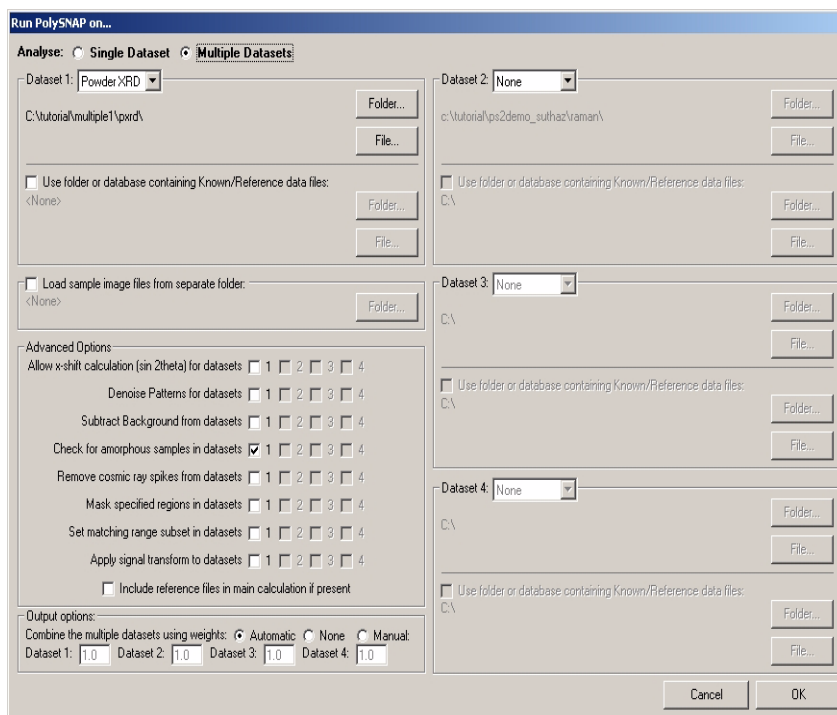
This tutorial involves the input and comparison of multiple datasets at the same time. This will demonstrate how combining results from different techniques can be useful and how looking at multiple datasets can yield extra information not available from just one.

As before, first select *Analyse Data* from the welcome window:

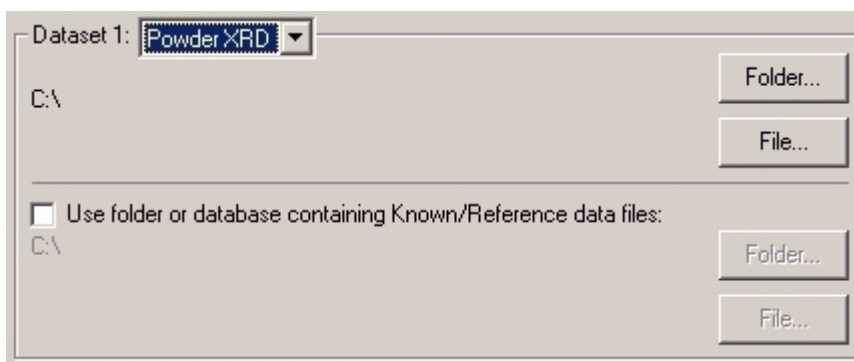
This opens the data input window. When first opened there is only one section for inputting a single dataset. To input multiple datasets at the same time select the *Multiple Datasets* option from the top of the window:



This expands the data input window:

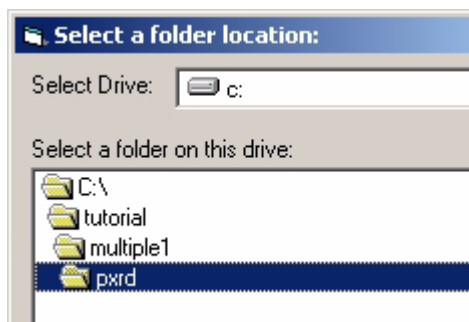


It now has four sections for entering up to four datasets; we will be using just two in this example. Each input section is identical:



There is a pull-down menu near the top that allows selection of the type of dataset that is going to be used. To the right there are buttons at the side that allow the location of the dataset to be specified. There is also the option to input known or reference files can be entered however these will not be used in this example.

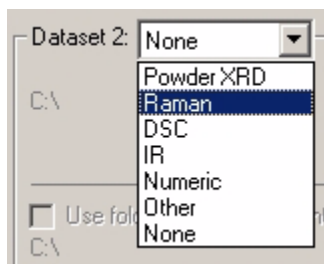
For this example the first dataset is a dataset of powder X-ray diffraction data and the dataset is contained inside a folder. Ensure that the dataset type is set to *Powder XRD* and click on the *Folder...* button. This opens the following window:



1. Select the drive containing the tutorial data from the pop-up menu at the top. In this case select *C:*.
2. Navigate to the folder *C:\Program Files\PolySNAP2\tutorial\multiple1* and open the folder called *pxrd*. This contains the Powder X-Ray diffraction data. Ensure that the last folder of the desired path (*i.e. pxrd*) is selected by double-clicking it. The folder icon should appear open, as in the screenshot above.
3. Click *OK*. The selected path (*C:\Program Files\PolySNAP2\tutorial\multiple1\pxrd*) should now be displayed in the upper portion of the section for Dataset 1. Check it is correct. If it is not, repeat the previous steps.

If we were to click *OK* in the main input window now, it would start an analysis run using the single dataset that has just been entered. However for this run two different datasets are going to be used.

The second dataset is a folder of Raman data. In the input section for *Dataset 2* select *Raman* from the pull-down menu.



Click on *Folder...* for *Dataset 2* and select the *Raman* folder located in same location as the *pxrd* folder and click *OK*. The filepath of the second dataset should now be displayed in the main window (*i.e. C:\Program Files\PolySNAP2\tutorial\multiple1\raman*). Now both datasets have been entered and both will be used in the analysis.

Below the input sections there is a section of advanced options. These are all the processing that can be applied to the data as it is being input and before it is fully analysed.

Advanced Options

Allow x-shift calculation (sin 2theta) for datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

Denoise Patterns for datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

Subtract Background from datasets ☒ 1 ☒ 2 ☐ 3 ☐ 4

Check for amorphous samples in datasets ☒ 1 ☐ 2 ☐ 3 ☐ 4

Remove cosmic ray spikes from datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

Mask specified regions in datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

Set matching range subset in datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

Apply signal transform to datasets ☐ 1 ☐ 2 ☐ 3 ☐ 4

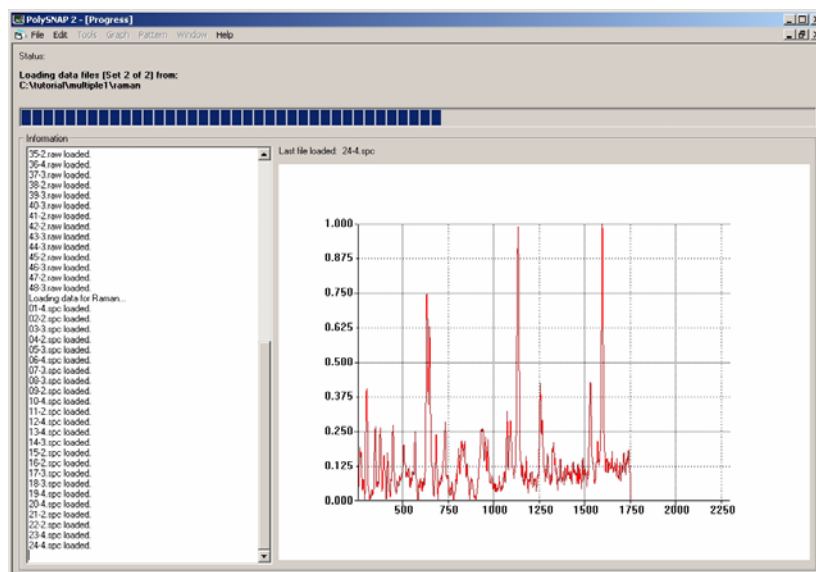
☐ Include reference files in main calculation if present

These options can be applied to individual datasets selectively. In this case only *Datasets 1* and *2* are being used, so only these two datasets have selectable checkboxes. The checkboxes for *Datasets 3* and *4*, which are both currently empty, are grey and unselectable.

Select to Subtract Background from both datasets 1 and 2.

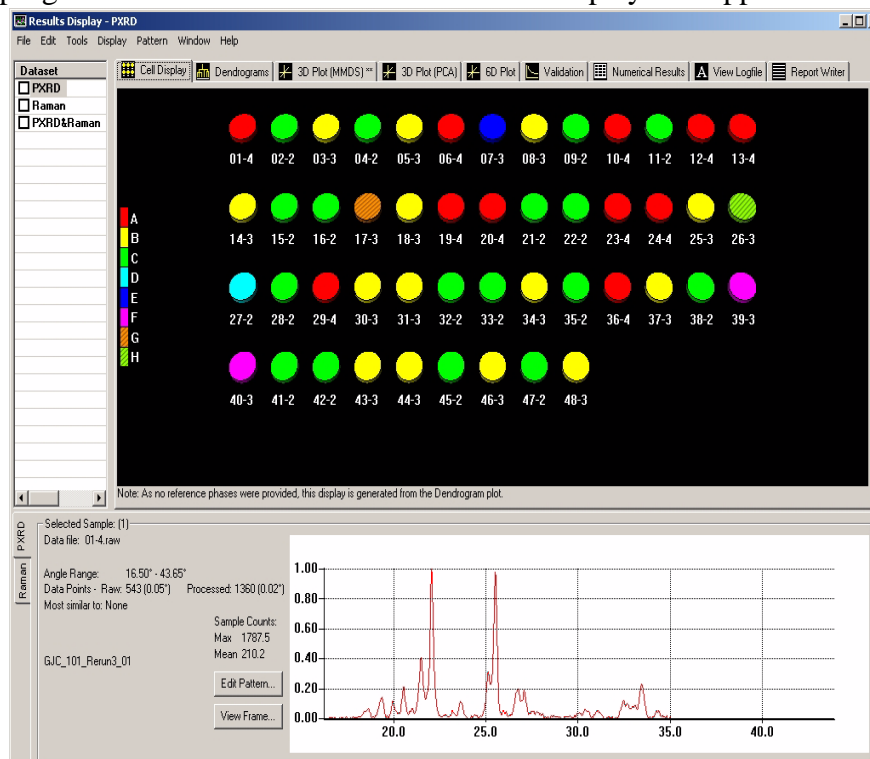
All of the relevant instructions have now been provided to the program. Click *OK* in the main input window to start the analysis.

A progress window will open as *PolySNAP* reads in the two datasets in turn, before performing analysis on both of them.

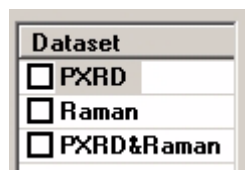


After analysis is performed on each dataset separately, a third analysis is carried out to create the combined results using

information from both datasets. Once this analysis is complete the progress window will close and the results display will appear:



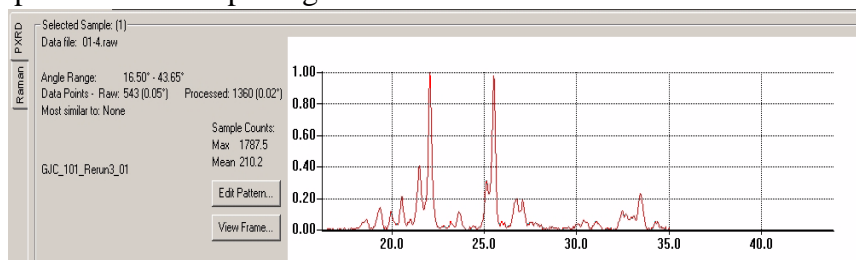
By default it will display the cell display of *Dataset 1* which is the Powder X-Ray dataset. There is a list of all the available datasets in a menu running along the left hand side of the display.



In this case there are three available datasets - the original datasets that were input (*PXRD* and *Raman*) and the combined dataset that was obtained using both the Power X-Ray and Raman data (*PXRD&Raman*).

Click on the **label name** *e.g.* **Raman** (not the checkbox) to switch to the results display for that dataset. The tab bar along the top can be used to switch between the display screens (Cell Display, Dendrogram, 3D plot *etc.*) within each datatype as before.

At the bottom of the results display is the display pane that shows the spectra for the sample patterns as well as any image files that were specified when inputting the data.

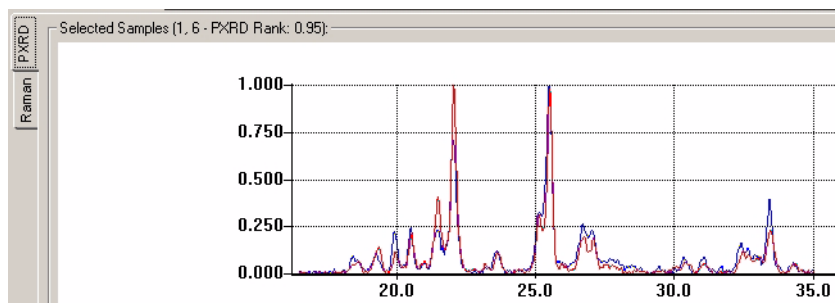


There is a tab bar running vertically along the side of this display that allows the spectra being displayed to be changed between showing either the original PXRD or Raman spectra for the currently selected pattern.



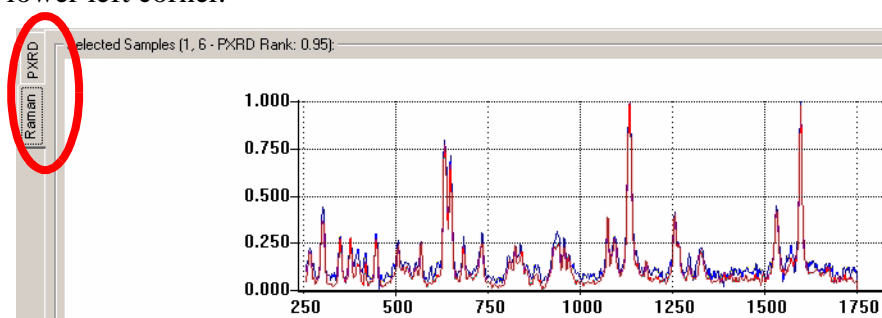
Spectra are only available for the two original datasets and these can be alternated between at anytime by using this tab. Changing the type of pattern shown has no effect on the main display.

Switch to the PXRD results display, by clicking on the label **PXRD** in the list in the top-left. Now, using the control-key, select two patterns of the same colour in the cell display - for example, *01-4* and *06-4*. The two PXRD profiles are displayed overlaid in the bottom portion of the window:



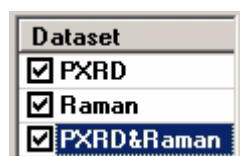
Based on the profiles, it looks reasonable that these have been clustered together. Now, switch to viewing the Raman profiles of the

selected samples, by choosing *Raman* from the vertical tab-bar on the lower left corner.

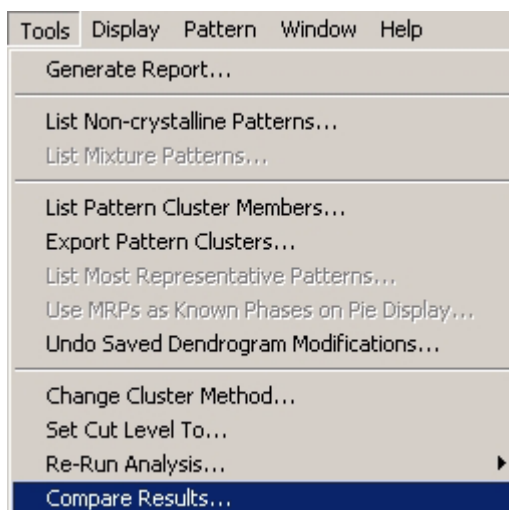


By comparing the profiles from the two separate methods, and seeing they are in agreement, we can have greater confidence that the analysis has been correct.

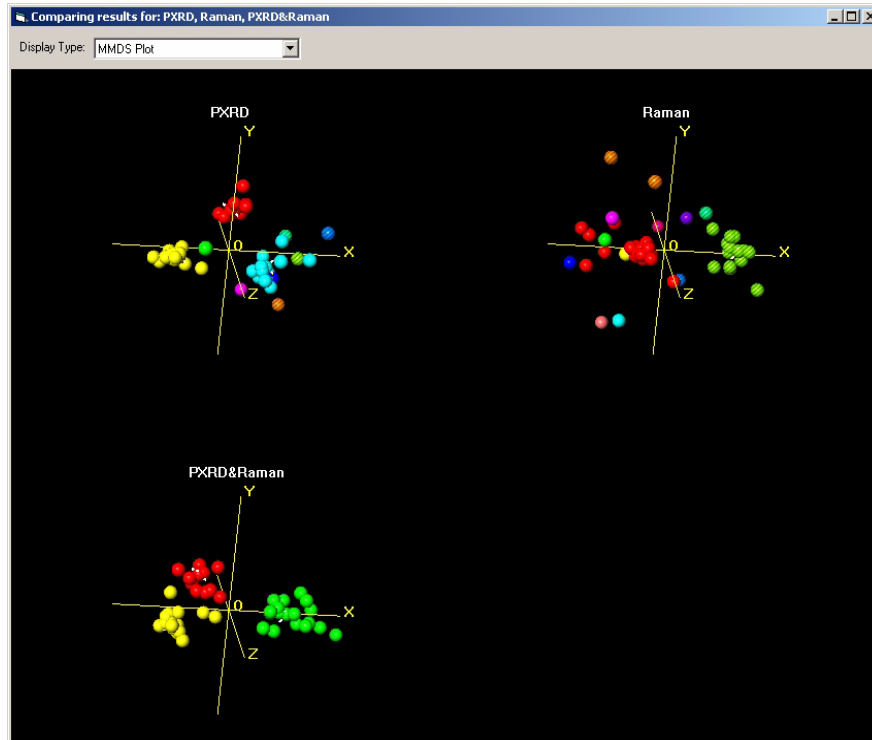
As well as switching between different displays, the results can also be compared directly in a new window. To do this select all of the checkboxes in the dataset list.



Select *Compare results...* from the *Tools* menu.

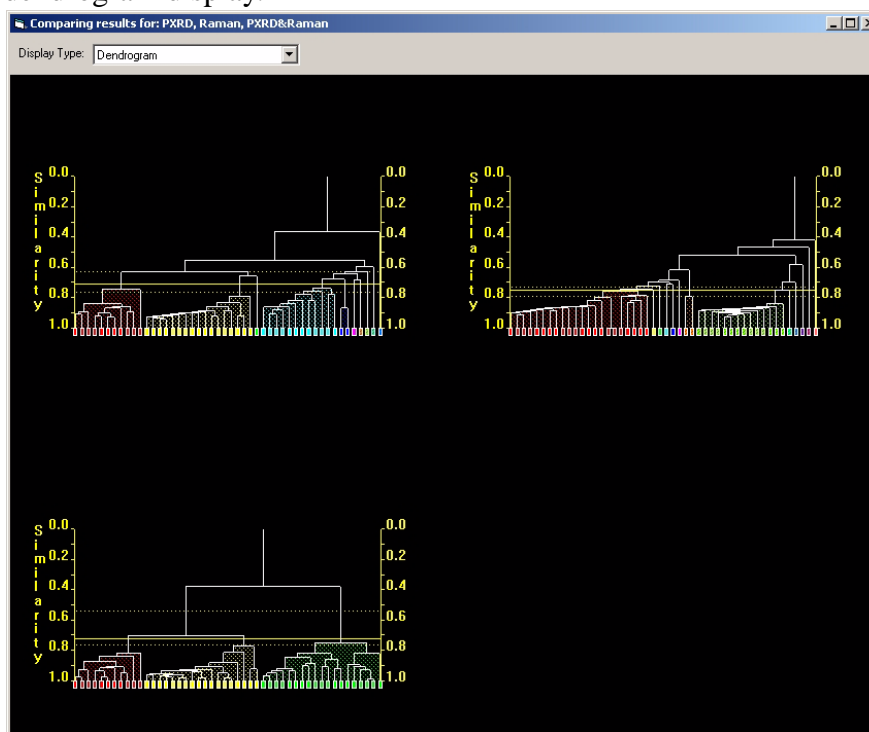


This opens a new window where the visual results from the three selected datasets are displayed next to each other. By default it opens with the 3D (MMDs) plot.



In this way the features of the different 3D plots can be directly compared, as the display can be rotated and zoomed into like a normal 3D plot. Here, we can easily see that the clusters for the combined PXRD&Raman dataset seem to be better than for either of the individual original datasets.

There is a pull-down menu in the top-left of the window that allows the 3D plot (MMDS) to be changed either to the 3D plot (PCA) or the dendrogram display.

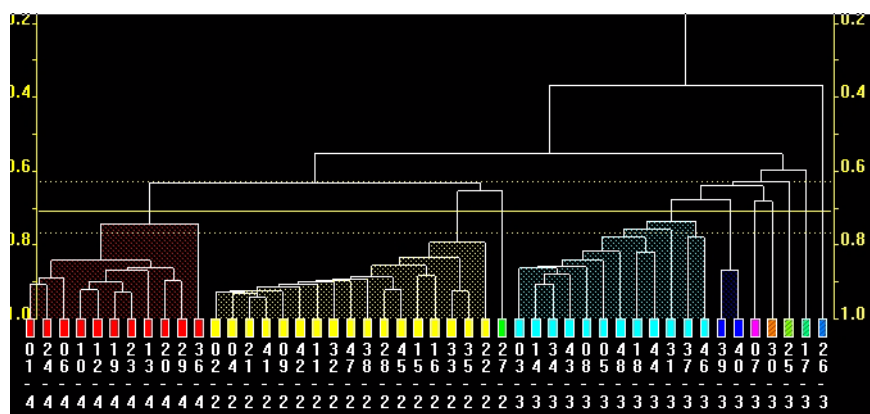


Now, let's look in more detail at the dendrograms of each of the datasets. Close the *Compare Results* window, and switch to the PXRD dataset view, then select the *Dendrogram* tab.

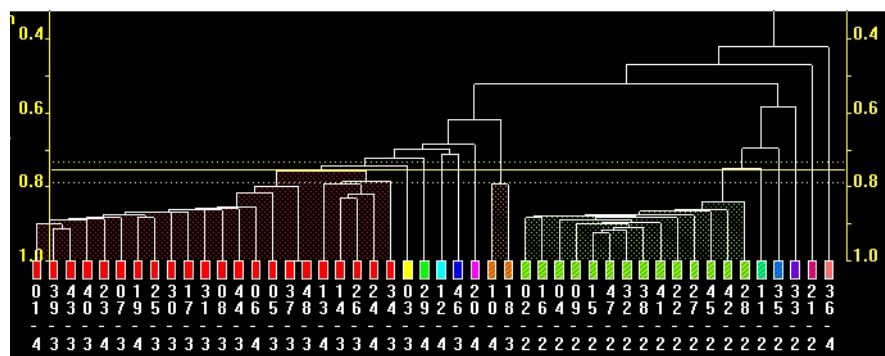
This dataset consists of three forms of a particular compound. The last character of the pattern labels show if a given sample should be Form 2, 3 or 4.

From the initial default cut-level of the PXRD dendrogram, it can be seen that using PXRD data alone has done quite a good job of partitioning the data into clusters corresponding with the different forms; with the only problem being the Form 3 samples have been

split into different places on the dendrogram (blue cluster, and group of patterns at the far right):

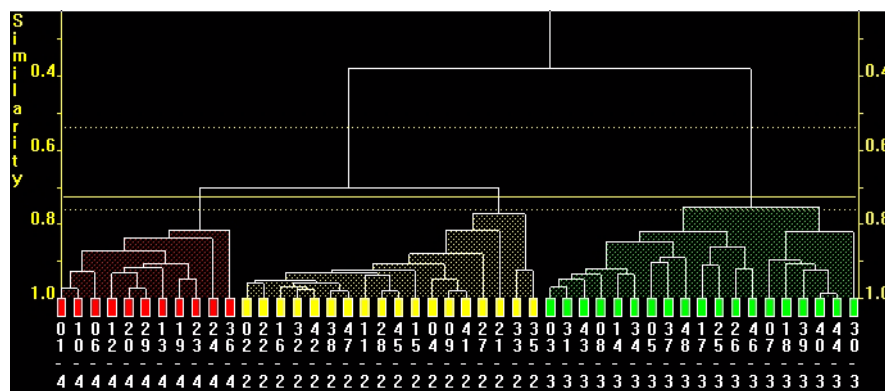


Now, switch to looking at the Raman results, by clicking on *Raman* in the list on the upper left. The display will update to show the dendrogram based on purely the Raman spectra provided:



Clearly, Raman by itself has not done as good a job as the PXRD in this example; the red cluster has forms 3 and 4 all mixed up, with the Raman unable to distinguish between them.

Now switch to the dendrogram for the combined PXRD&Raman results:



This has identified the 'correct' cluster for every pattern in the dataset, showing the value of combining results from multiple techniques.

7.4 Multiple samples - Dataset 2

Close any open results windows, and click in the white background of the main PolySNAP window to bring up the welcome screen. Select *Analyse Data*, and select the following options:

Dataset 1: Powder XRD

Data folder: *C:\Program Files\PolySNAP2\tutorial\multiple2\pxrd*

Background subtraction on.

Set Matching Range Subset to 5 - 30 degrees.

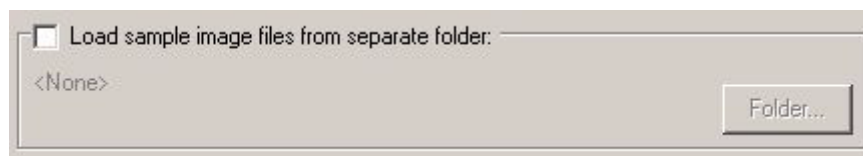
Dataset 2: Raman

Data folder: *C:\Program Files\PolySNAP2\tutorial\multiple2\raman*

Background subtraction on.

Set Matching Range Subset to 250 - 1750.

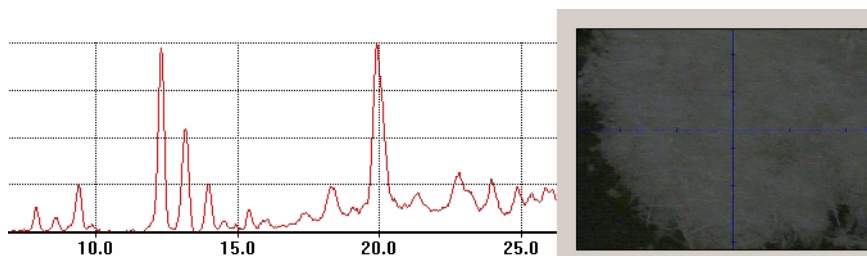
Finally there is an option that allows image files of the samples to be included and displayed along with the results.



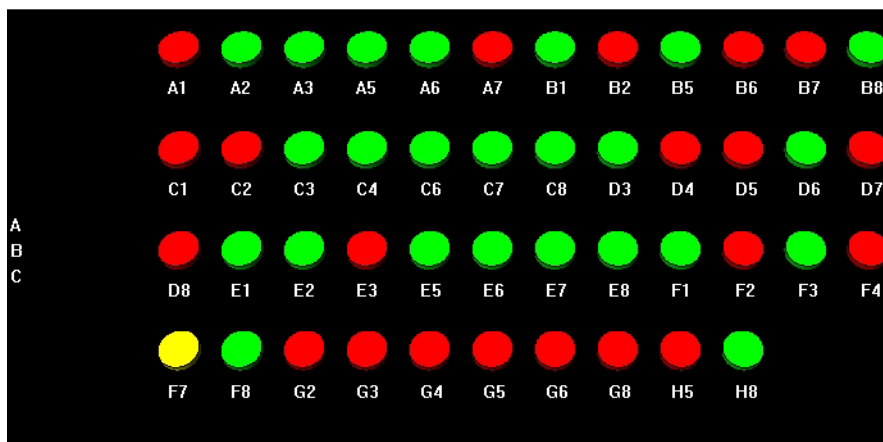
Click on the checkbox to activate this option then click on the *Folder...* button. Select the *jpeg* folder from the same location as before and click *OK*. The filepath (*C:\Program Files\PolySNAP2\tutorial\multiple2\jpeg*) should now be displayed.

Click *OK* to begin the analysis.

The results window appears, including in the bottom-right, a thumbnail view of the sample image corresponding to the currently selected sample in the display. To see this image full-size in a separate window, click on the thumbnail view.

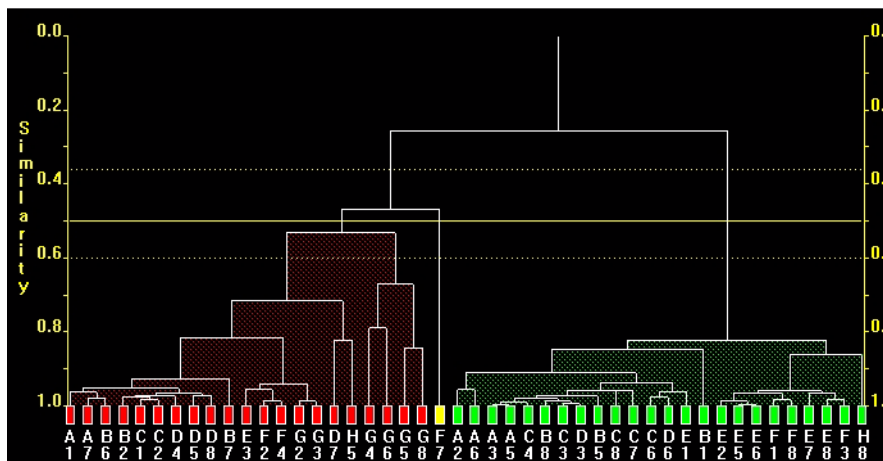


Look at the cell display for the combined results (*PXRD&Raman*).



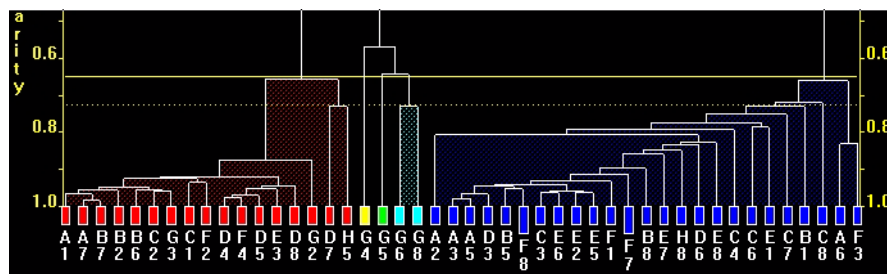
Most of the patterns have either been grouped into the red cluster or the green cluster; interestingly the yellow pattern *F7* has been assigned to a cluster of its own.

Switching to the dendrogram it can be seen that *F7* is quite isolated and highlighted as markedly unusual compared to everything else:

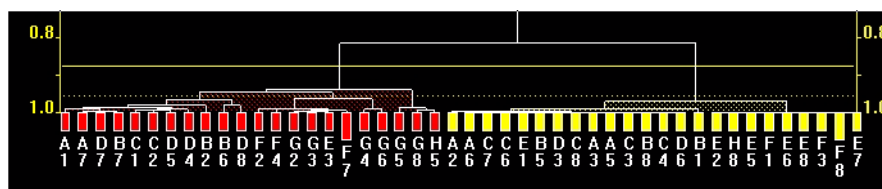


There is a very high tie-bar between it and the any other pattern. The combined results have identified this as pattern of interest. The results from the dendrogram are very similar to those 3D (MMDS) plot where *F7* is plotted noticeably far from the other samples.

To understand this, switch to the dendrogram of *Dataset 1 - PXRd*. In the X-Ray data results there appears to be nothing unusual about pattern *F7*. It is in the same cluster as a similar sample, *F8*.

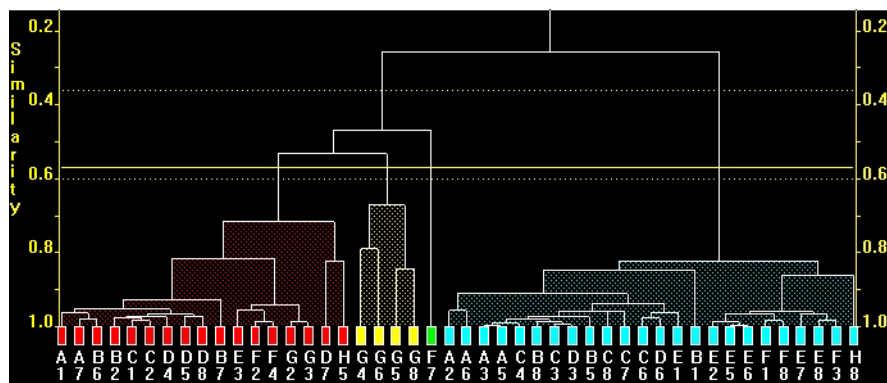


Switch to the raman dataset dendrogram.

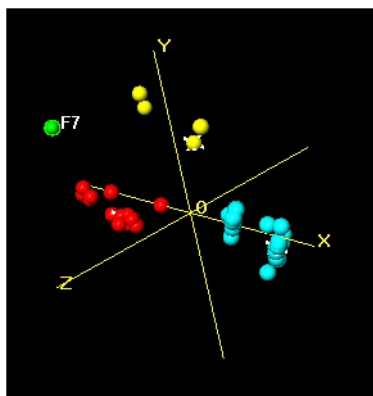


According to the Raman data, these two samples are in completely different clusters, contradicting the PXRD results.

Switch back to the dendrogram for the Combined results. Lower the cut-level to around 0.6, separating out G4, G5, G6 and G8.



Switch to the 3D MMDS plot, and see that the cut-level chosen describes the clusters shown there as well.



We now have a result that well describes the dataset; two large clusters of pure material A and B, four mixtures of A&B (the yellow cluster), and the one odd outlier.

Looking at each individual dataset display alone there would be no indication that there is anything unusual about *F7*. However by combining the results, it becomes clear there is a contradiction between the two methods, and this results in this particular pattern being highlighted for further investigation by the user.

This shows the value of having different techniques available to examine a sample, and the worth of combining the results from each. With a single dataset this unusual sample that needs to be brought to the user's attention can end up hidden amongst the rest of the results. It is only when the results are combined that the discrepancy becomes clear, drawing the attention of the user to that sample and allowing it to be investigated in more detail.

7.5 6-Dimensional Plots

There are a number of methods for analysing groups of patterns, some of which were detailed in the previous examples. However, these methods only concerned the powder diffraction pattern itself and not any other associated data that may be available. For example, this additional information can include sample preparation details such as the solvent used, concentration, pH, volume, temperature, *etc.*, and up to three of these can be incorporated into the 6D-plot at one time.

To begin, start with only the empty PolySNAP window containing the menu bar.

From the *File* menu, select *Automatic Analysis* and the *Analyse Data...* option.

Input Folder: *C:\Program Files\PolySNAP 2\tutorial\6D-plot*

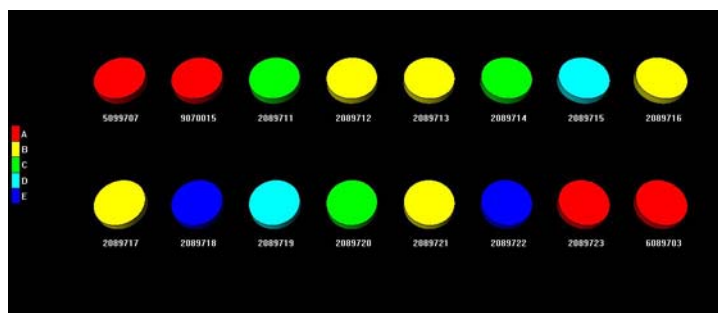
Known Data Files: *<None>*

Datatype: Powder XRD

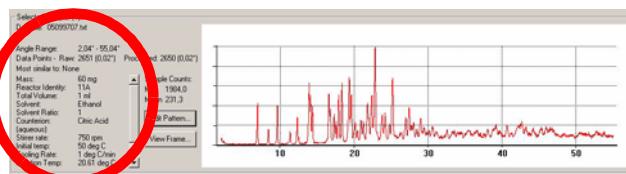
Click *OK*. The pattern files in the specified folder are now loaded into PolySNAP, which carries out pattern matching and cluster analysis of the data as before. Once complete, the results window will appear with a view of the *Cell Display*.

7.5.1 Analysing Results using 6-D Plots

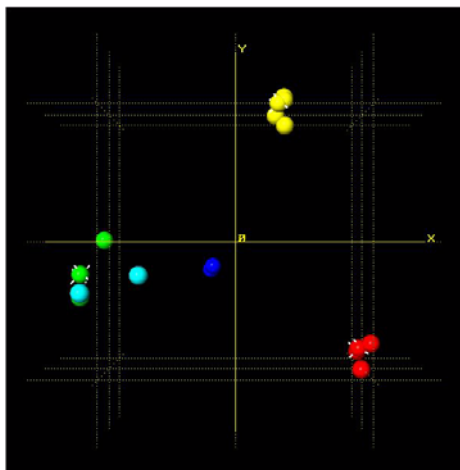
The 16 patterns samples contained within the input folder are presented in the *Cell Display*, where it can be seen that there seem to be five different types of pattern.



When an individual pattern is selected it can now be seen that not only is there a view of the pattern profile, but there is also textual pattern information (on the left).



The sample information shown is embedded in the pattern data file, Select the tab labelled *3D Plot (MMDS)*. The following display appears:



This view displays several distinct main clusters with the cells spread out along the x-axis. By selecting the cells individually, and examining the sample information pane, it can be seen that there appears to be a correlation between the solvent used in preparing the samples, and the different groupings of patterns. For example, all of the red samples are ones that were prepared with ethanol as solvent.

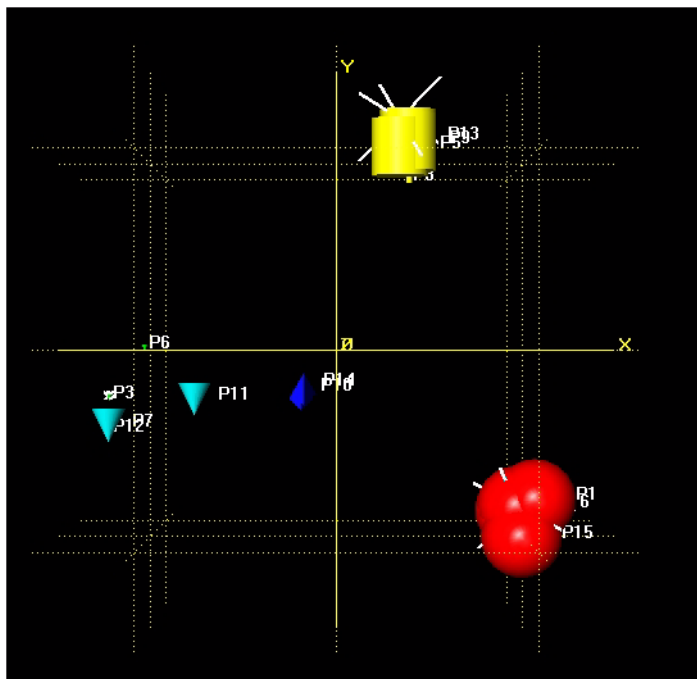
This information can be viewed more easily in a graphical manner, by enhancing the 3D-plot by plotting additional information on it, in the form of extra dimensions. For example, we can plot the different types of solvent used in terms of different shapes, and different reaction times in terms of different sizes of each plotted point.

1. From the main toolbar select *6D Plot...* At the top of the frame several options are presented.



2. Set the *Plot* option to *MMDS*.
3. Make sure the *Ball Size* option is checked, and select to plot *Reaction Time*.
4. For the *Ball Colour* option, select *Dendrogram Colours*.
5. Ensure *Ball Shape* is selected and set to *Solvent*.

6. Click *Apply*.

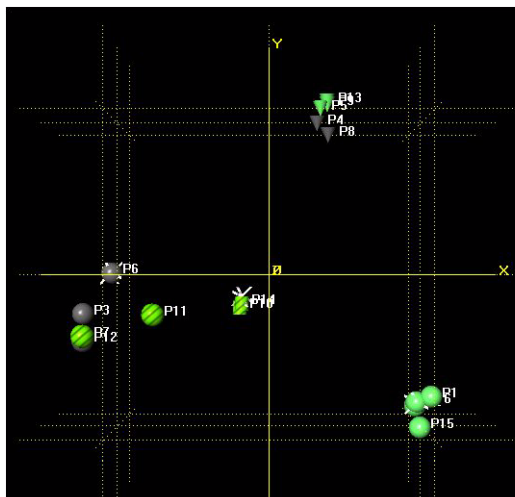


It can be seen that while the plotted points are in the same positions and colours as before, there are now several different sizes and shapes on the display. Note that the different shapes are clustered together - spheres are next to spheres, cylinders are next to cylinders, *etc.* Select one of the sphere-shaped patterns by clicking on it. Note that the solvent used in all of the patterns plotted as red spheres is ethanol; whereas all the patterns plotted as cylinders, for example, used methanol. While this information was available before, it is easier to spot trends in the data when viewed in this manner.

From the different sizes of shapes plotted, it can be seen that there were three different reaction times used in this set of experiments. This may be seen more clearly by replotting the display, using colour instead of size to represent the different amounts.

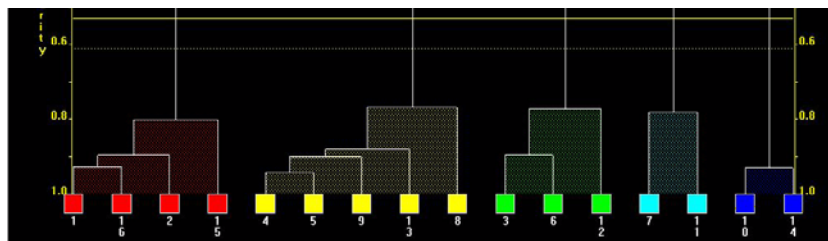
Deselect the *Ball Size* option, turn on *Ball Colour* and set it to plot *Reaction Time*, and ensure *Ball Shape* is still set to *Solvent*. Click *Apply*.

The new plot now looks like this:

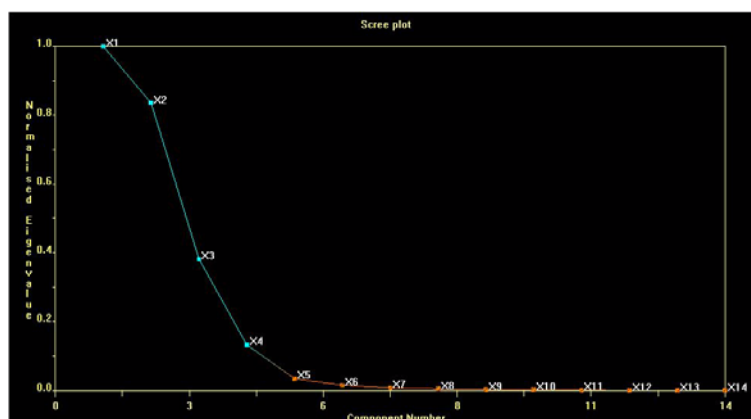


The different colours - green, grey and striped, represent the differing reaction times used. It is interesting to note that the two patterns 4 and 8 come up as separate colours - and hence times - from the rest of the patterns they are similar to.

Switch to the *Dendrogram* tab. From the *Display* menu, uncheck the option *Show Well Identity If Available*. Notice that pattern 8, in yellow, seems more separated than any of the rest of the yellow patterns.



We need to consider if the cut-line is positioned correctly. Switch to the *Validation* tab and check that *Scree Plot* is selected in the pop-up menu in the upper left corner.



The scree plot is one of the many methods used for estimating the cut-level of the dendrogram, and can be a useful visual aid for the user. It is derived from the eigenvalues of the pattern correlation matrix. To interpret the scree plot the gradient between the values (number of clusters) need to be analysed. Where the eigenvalues are very different from one another there will be a steep gradient between clusters. This is one of the methods used to select the number of clusters.

For this dataset the number of clusters suggested from this is between 4 and 5, because it is here that the plotted line changes colour, and the gradient starts to level out. 5 clusters corresponds to the current cut level; 4 would correspond to raising the cut line such that the green and blue patterns were considered part of a single cluster.

In cases such as this where the graphic displays may not give a clear-cut solution, it can often be useful to examine the detailed statistical output from the analysis. This is shown in the *Logfile* tab. Switch to it, and scroll to the very bottom of the output. Scroll back up slowly, until you see the output from the cluster analysis. One section of it reads as follows:

```
Summary of the estimates of cluster numbers:

From principal components analysis (non transformed matrix): 5
From principal components analysis (transformed matrix):      4
From multidimensional metric scaling:                          4
From the gamma statistic using single linkage:                 5
From the Calinski-Harabasz statistic using single linkage:    5
From the C-statistic using single linkage:                     5
From the gamma statistic using group averages:                 5
From the Calinski-Harabasz statistic using group averages:    5
From the C-statistic using group averages:                     5
From the gamma statistic using the Ward method:               5
From the Calinski-Harabasz statistic using the Ward method:   5
From the C-statistic using the Ward method:                   5
From the gamma statistic using complete linkage:              5
From the Calinski-Harabasz statistic using complete linkage:  5
From the C-statistic using complete linkage:                   5

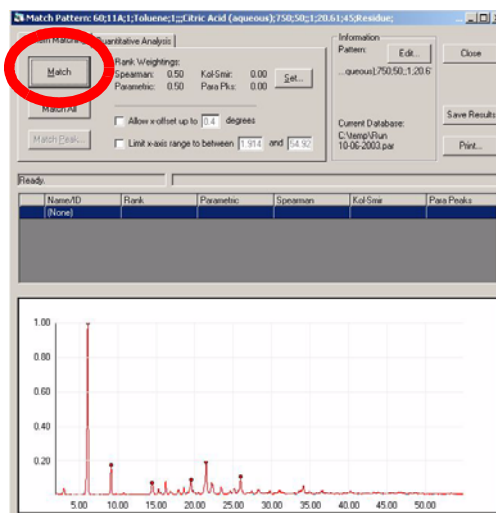
Maximum estimate is 5
Minimum estimate is 4
Combined weighted estimate of the number of clusters is 5
The median value is 5
```

This suggests that 5 is indeed the correct number of clusters.

Go to the Dendrogram view, and select patterns 8 and another of the yellow patterns. Consider their overlaid profiles - the similarity between them suggests that the program was correct in its initial placement of the cut line, and the reason 8 is separated somewhat from the rest appears to be down to an x-shift.

7.5.2 Manual Matching

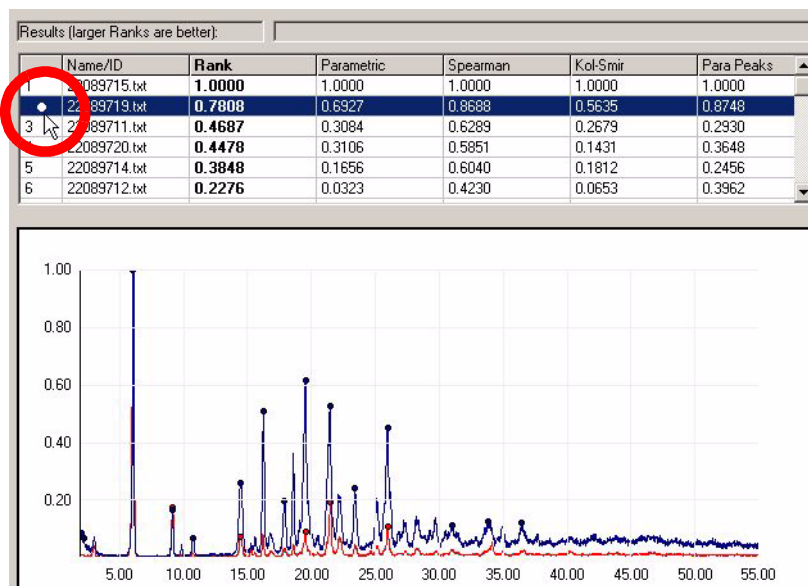
Select pattern 7. To get a better idea of what is going on with particular patterns, sometimes it is useful to compare them manually to the rest of the patterns. From the *Pattern* menu, select *Manual Match/Quantitative*, and from the resulting sub-menu, select *against Sample database*. A new window will appear.



This window, the manual match window, has the selected pattern shown in the bottom half. The upper half has the controls for the manual matching. Click the *Match* button in the top left, and the selected pattern is compared to each of the other patterns in the database in turn.

The results are then displayed in a table in the centre part of the screen. The most similar pattern to the selected pattern is that pattern itself, which appears at the top of the list, with perfect correlation scores of 1.0 in each of the 4 tests displayed. The results are sorted in descending order of Rank value, which a user-controllable combination of some of the individual matching tests. By default, it is the average of the Spearman and Parametric tests.

Click on the number **2** in the first column of the second row. The corresponding profile for that pattern is shown overlaid on the original to allow visual comparison.

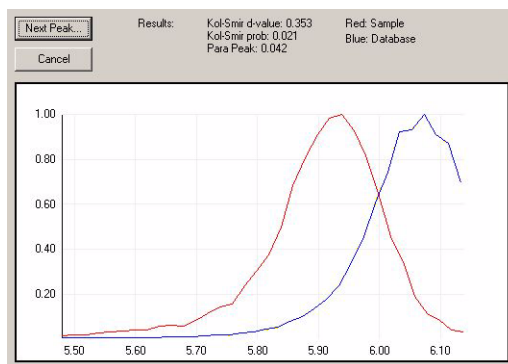


This pattern is the most similar to the selected profile, but has quite a few obvious differences. This is reflected in the relatively low scores that the pattern gets in each of the matching tests. This pattern corresponds to pattern 11 on the dendrogram earlier, which that method also identified as being most similar. It seems there are preferred orientation problems here.

Click again at the same place in the first column to make the overlaid pattern disappear. Now click on the number **3** to see what the next most similar pattern is. This corresponds to pattern 3 on the dendrogram display, part of the green cluster.

Notice that the numerical results are much lower than for the previous pattern, which supports the choice of the program to place these patterns in separate clusters.

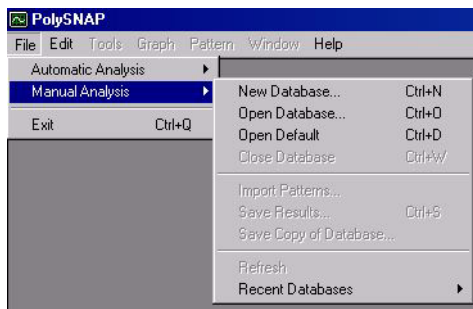
To examine the differences between them on an individual peak-by-peak basis, click the *Match Peaks* button.



A small window appears showing the first two coinciding peaks, and individual matching tests results between them. Clicking the *Next Peak* button repeatedly moves through the marked peaks on the selected patterns. Click *Cancel* to dismiss the window.

7.6 Manual Matching Options

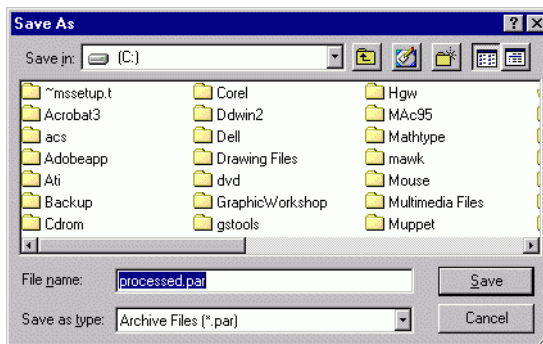
The rest of the examples in this tutorial demonstrate the manual analysis section of PolySNAP. This mode can be accessed from the *Manual Analysis* section of the *File* menu.



The first thing to do is to create a new database of patterns with which to work.

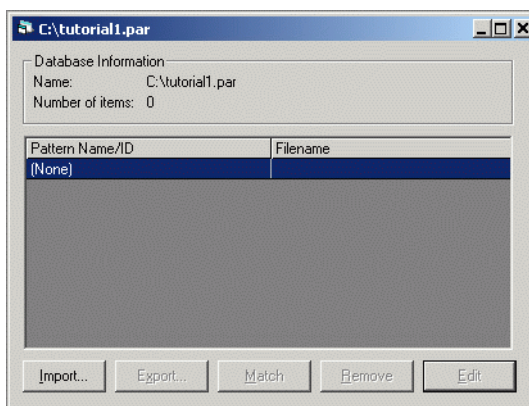
We first need to create a new database, which will initially be empty. We will then load some powder patterns into it, and work with them.

Choose *New Database* from the *Manual Analysis* section of the *File* menu. A standard Windows file dialog box appears, and you are invited to choose a name and location for the new database.



Enter *tutorial1* as the database name, select a location (C:/ for example), and click *Save*. The dialog box disappears, and a new window opens inside the PolySNAP workspace.

This is our new empty database; its filename is displayed in the window title bar. Like any other window, it can be moved around the screen by clicking and dragging on the title bar, and can be re-sized by dragging at the edges of the window.



The filename and your selected location appear in the Database Information section of this new window.

The number of patterns in this database is also displayed - as expected, it is currently zero.

Several buttons are arranged along the bottom of this window. Most of these are unavailable as there are no data files in the database. The only one that is available for use is the *Import* button.

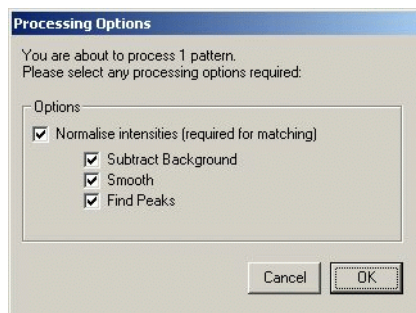
We now need to bring some pattern data files into the database.

Click the *Import* button. We need to navigate to where the data is stored. Locate the folder

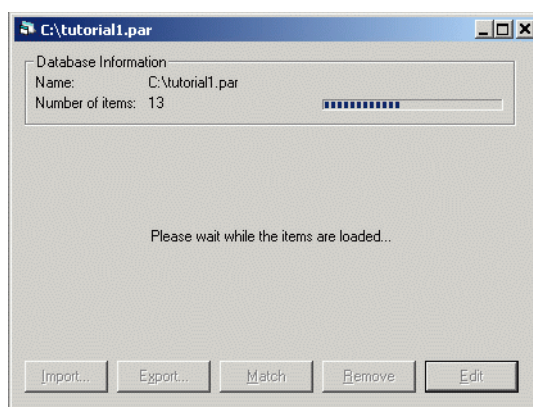
C:\Program Files\PolySNAP2\tutorial>manual

Once in the correct location, a list of different files in the folder that the program recognises is listed.

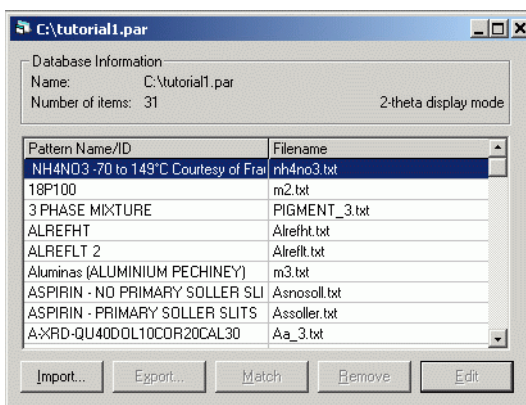
We want to open all of the files in this particular folder, so hit *control-A* on the keyboard to select them all, and then click *Open*. A new window appears with the pattern processing options.



These will be discussed in more detail later, so leave them with their default values for the moment and click OK. A progress bar appears at the top of the database window to allow progress to be monitored. As there are only a few patterns in the folder, this process should only take a couple of seconds.



Once importing has finished, the patterns will be listed in the main part of the database window:



If the window is too small to see all of the patterns, the scroll bar on the right hand side can be used to view the rest, or you can drag the window border to enlarge the window.

Each entry in the list represents one pattern file that has been imported.

Initially, the patterns are listed in the order they were imported. The list can be sorted by either *Pattern Name/ID*, or *Filename*, by clicking on the headers at the top of the list.

If some of the chemical names here look very similar to filenames this is because the ASCII files we have imported do not contain any chemical name information, so the filename is used instead. The actual chemical name, if known, can be added manually later.

You can select a particular pattern in the list by clicking once on it. When you do so, the other buttons along the bottom of the window become activated.

The *Import* button has already been used. The *Export* button lets you make a copy of a particular database entry to a separate file, in either ASCII text or PolySNAP pattern format.

Remove deletes the selected pattern or patterns from the database.

Experiment with these options by exporting a pattern to a separate ASCII text file under the name *exporttest.txt*. Make sure to note where you saved it! Then try importing the file you have just saved. Finally, select the *exporttest.txt* pattern in the list and delete it from the database using the *Remove* option.

Removing patterns from the database does not in any way affect the original data files which are left intact and unchanged throughout all

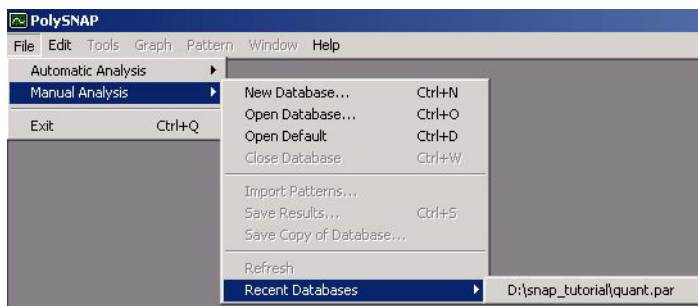
operations performed using PolySNAP - the program works on a copy of the data only.

The *Match* and *Edit* options will be dealt with in following sections, so leave them be for the moment.

Now quit the program by choosing *Exit* from the *File* menu. All windows will close. Note that any changes made to a database are automatically saved, so there is no need to manually save any changes before closing.

Relaunch the program again in the same manner as before. We want to return to the database we have just been working with. You could select *Open Database* from the *Manual Analysis* section of the *File* menu, navigate to where the database was saved, select it and click *Open*, but there is a much easier way.

The program keeps a record of the last four databases used in the *Recent Databases* submenu of the *Manual Analysis* section of the *File* menu. Open this now; one of the entries should be the

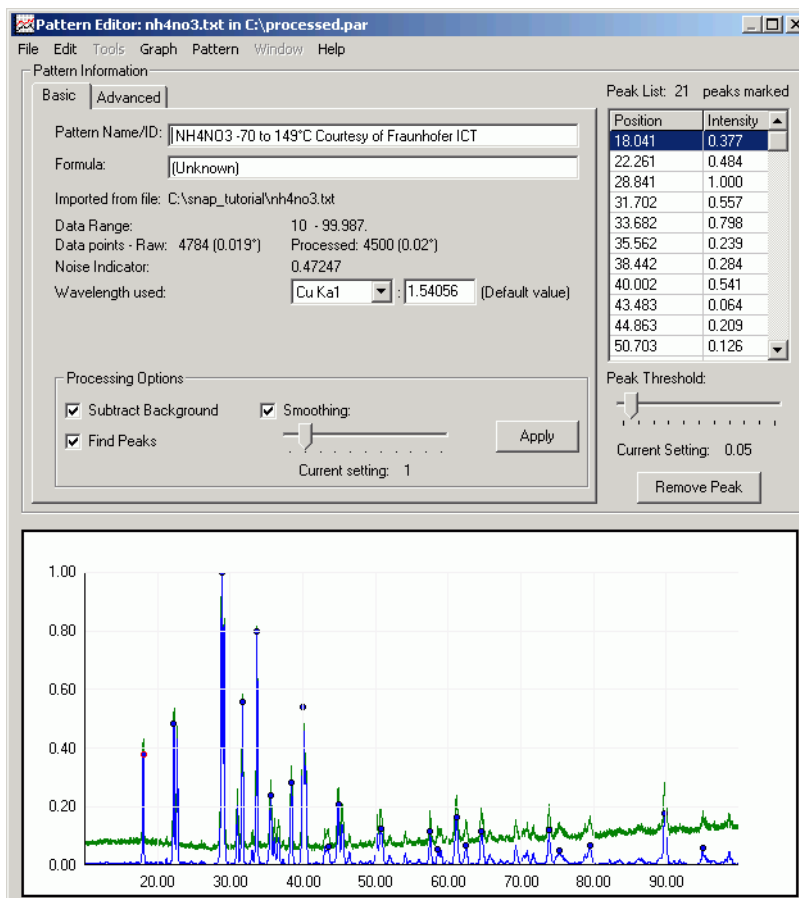


tutorial1.par database we were using earlier. Select it from the menu, and it should start to open automatically.

Next, we want to examine some of the patterns we have just imported in more detail.

7.6.1 Editing Patterns

Select the pattern *nh4no3.txt* from the list in the database by clicking once on it. Click the *Edit* button, and the Pattern Editor window appears.

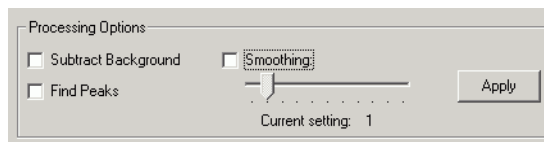


Several useful pieces of information about the pattern you have selected are displayed in this window, and a plot of the pattern itself is shown at the bottom.

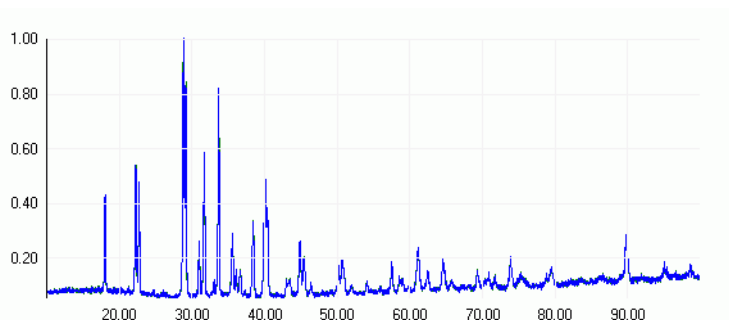
The chemical name or pattern ID can be changed here by editing the current name in the white text box.

Details such as the pattern filename, its start and end angles, and the number of data points are also displayed.

There are three check boxes labelled Processing Options just above the region where the pattern is displayed. To begin with, click once in each of the checkboxes, so that they are all turned off:



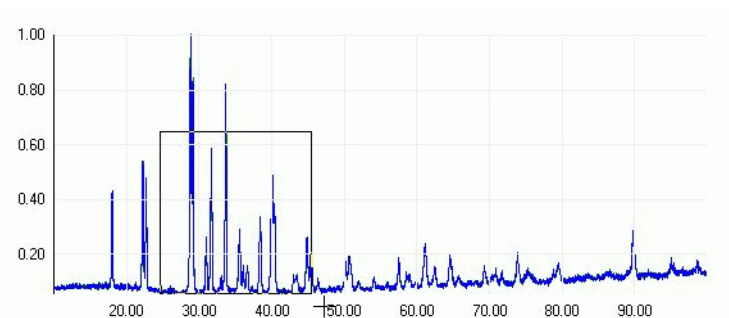
Then click *Apply*. The graphing region should now look like this:



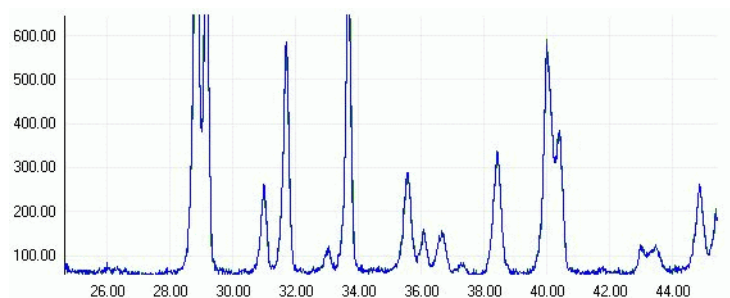
As you can see, the data has been scaled along the y-axis to run from 0.0 to 1.0. This operation is performed for all data being imported into PolySNAP in order to allow for suitable scaling between different data sets.

The raw data has also been interpolated from its original resolution, to the PolySNAP standard 0.02 degrees. Again, this is for consistency between patterns.

You can zoom into the pattern display, by clicking and dragging a rectangle on the region you wish to see more closely:

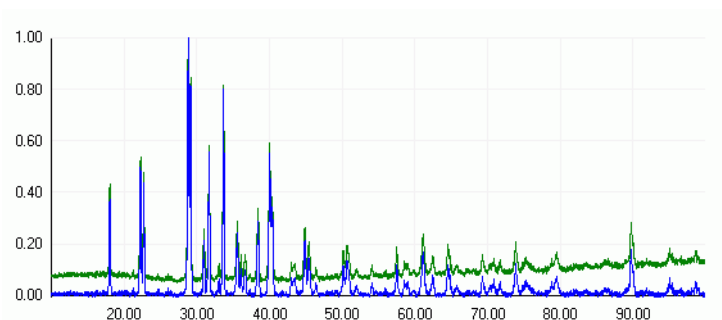


The view will then change to just the selected region:



To return to original view, right-click anywhere on the pattern display, and select *Reset View* from the resulting pop-up menu. Multiple zooms are possible by repeating the process.

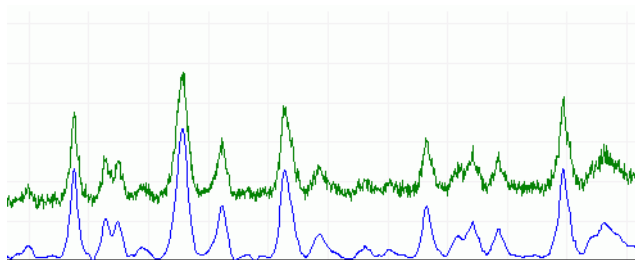
Now click the *Subtract Background* checkbox, and then click *Apply*. The pattern display should change:



The green line represents the raw pattern data. The blue line shows the same pattern after the background level has been subtracted. To see what the program considered to be background before subtractions, select *Show Background Curve* from the *Pattern* menu. A new window will open showing the subtracted background curve where the blue line now represents what is subtracted as background. Once you are finished looking, click *Close* to close this window and return to the standard view.

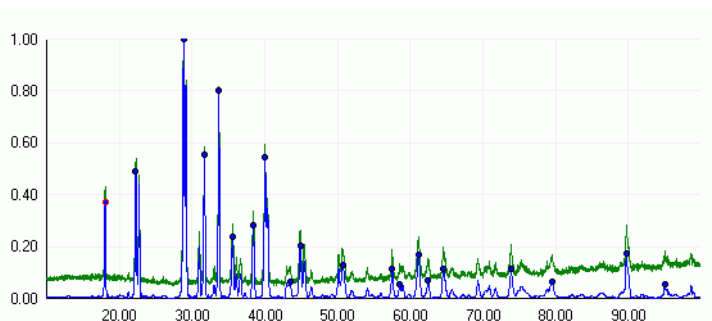
Now click the *Smoothing* checkbox, so both it and the *Background Subtraction* boxes are checked, and click *Apply* again.

The graph pane will be updated, and any noise in the pattern will be smoothed out (shown here zoomed in).



Finally, check the *Find Peaks* option, and click *Apply*.

The pattern display will update, and several small blue circles should appear on the top of the larger peaks:



These mark the location of what the program considers peak regions. There is a minimum peak height below which any peaks are ignored. This is set as a default to 0.05, which is why the smallest peaks are unmarked.

A list of the peaks the program has found is shown in the upper right corner of the pattern editor:

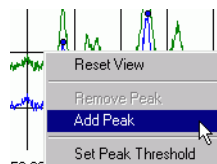
Peak List: 20 peaks marked

Angle	Intensity
18.041	0.372
22.261	0.479
28.841	1.000
31.702	0.554
33.682	0.794
35.562	0.237
38.422	0.284
40.022	0.536
43.463	0.064
44.863	0.205
50.703	0.123

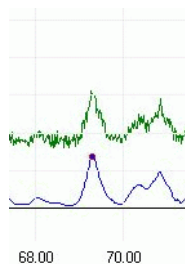
This lists the peaks found in order of increasing angle, and displays the corresponding intensity for each.

If the program has missed a peak you believe should be included, it is easy to add manually. Note that peak locations are added to the blue, processed pattern line, not the green, raw data line on the graph profile.

For example, if we wish to add the small peak that is located at around 54°, just right-click once at the point on the graph where you judge the peak maximum to be:

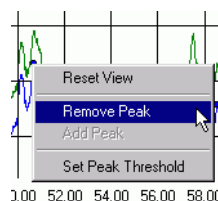


Then select *Add Peak* from the resulting pop-up menu. A round blue peak marker should appear at the top of the peak, and an entry for it should be added to the peak list:

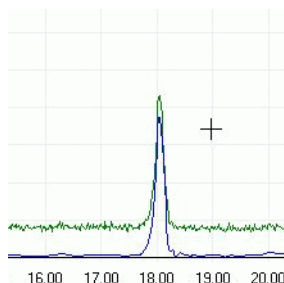


If this does not appear, try clicking as close as possible to the top of the peak. It may help to zoom in to the area of interest.

It is also possible to remove peaks you believe to be incorrect. For example, say we wish to delete the marker from the peak around 18°:

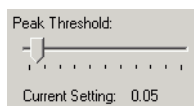


Right click as close to the peak marker as possible, and select *Remove Peak* from the pop-up menu that should appear. The peak marker, and its corresponding entry in the peak list, should vanish.



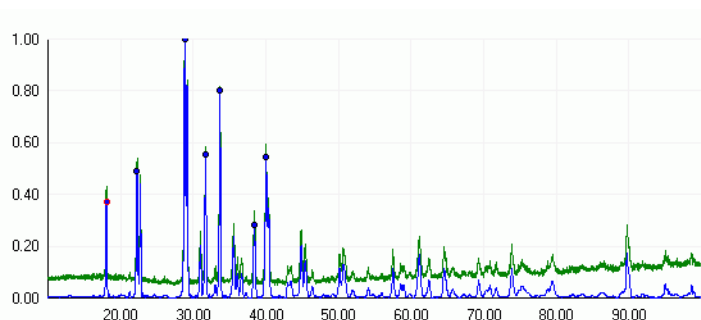
If the *Remove Peak* option is unavailable from the pop-up menu, you have not clicked close enough to the marker. Zoom in to make this easier, and try again.

Finally, it is possible to adjust the minimum peak height threshold to include less of the smaller peaks. Look at the slider just below the peak list - it should be set to 0.05.



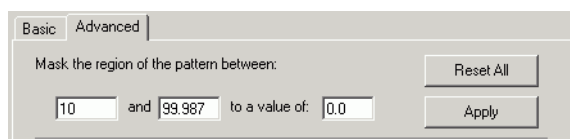
Drag the handle on the slider until it is set to about 0.25.

Several additional peaks that were marked before should now show up without peak markers, as they are now below the new minimum threshold of intensity.



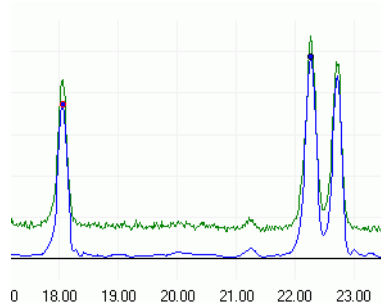
Finally, return the slider to 0.05. The peaks that previously became unmarked should show up again.

Now click on the *Advanced* tab in the top-left of the window.

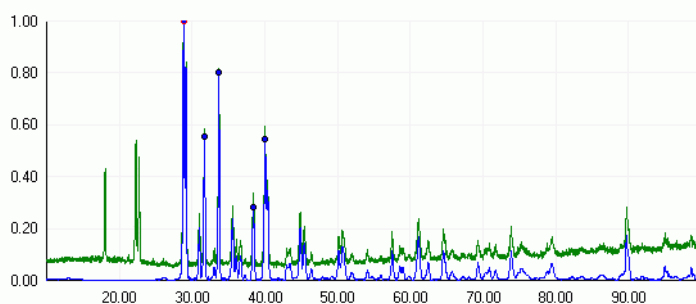


The top section revealed allows us to mask selected regions of the pattern to be ignored during matching or analysis processes. This may be useful where a particular peak is negatively affecting results. Say we wish to mask the peaks in the region of the pattern between 15° and 25°. Enter the start and end angles of the region to be masked in the relevant text boxes.

Examine the blue processed pattern in this region to determine its average background level – in this case, a value of zero should suffice, so enter 0.0 in the level text box, and click *Apply*.



The change in the pattern should be quite noticeable – while the green line of the raw data is untouched, the blue processed data line no longer has any peaks or features between the ranges we have entered:



Now repeat the process and mask the peak at approximately 90° using the same method.

Once finished, return the pattern to its initial state by clicking *Reset All*.

The rest of the *Advanced* tab options will be discussed in the Analysing Mixtures section later in this tutorial.

Choose *Close Window* from the *File* menu to dismiss the editor window. The program will check if you wish to retain any of the changes made in the editor. At present you do not want to keep any changes, so click *No*. You are returned to the main database window.

Feel free to look at some of the other patterns in the database in the editor window, and investigate the various processing options.

7.6.2 Matching Patterns

Assume that you have just obtained a new, unknown, powder pattern.

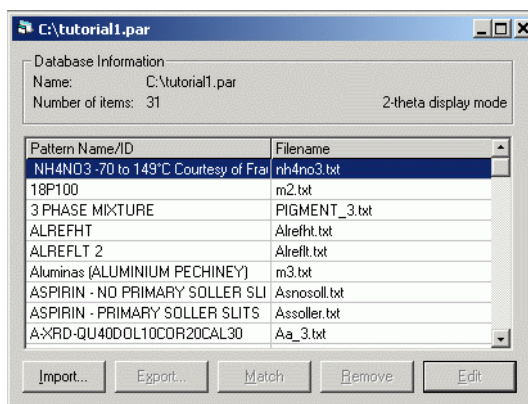
You want to find out what pattern in your existing database it is most similar to - and hence possibly identify what substance it may represent.

First, close any currently open database or editor windows.

Now, go to the *Manual Analysis* section of the *File* menu, and select *Open...*

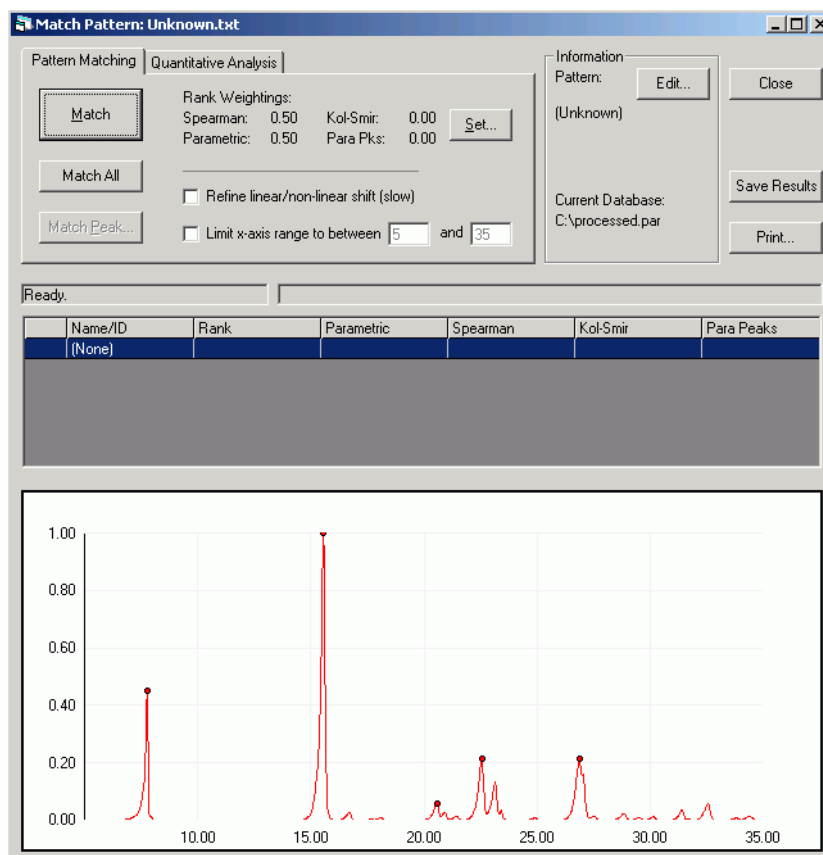
Open the pattern database you created earlier, *c:\tutorial1.par*.

A new database window will open, and the pattern data will be loaded into it.



The first step would normally be to import the unknown pattern into this existing database in order to examine it. In our case however, it is already loaded in the database.

Locate the pattern *Unknown.txt* in the scrolling list of patterns in the database, click once to select it, and then click on the *Match* button. The main program match window will appear.



Our unknown pattern is displayed in the graph pane in the bottom half of the window. Note that the pattern name and the database we are using are listed at the top right. On the top left of the window are the matching and analysis controls contained within two tabs – *Pattern Matching* and *Quantitative Analysis*. The default tab is *Pattern Matching*, which is what we wish to do first.

Click on the *Match* button. A progress bar appears as the program runs through various tests in order to compare our selected pattern to every other pattern in the database. Once matching is complete, the



centre section of the window fills with the numerical results, sorted by the column in bold type, Rank.

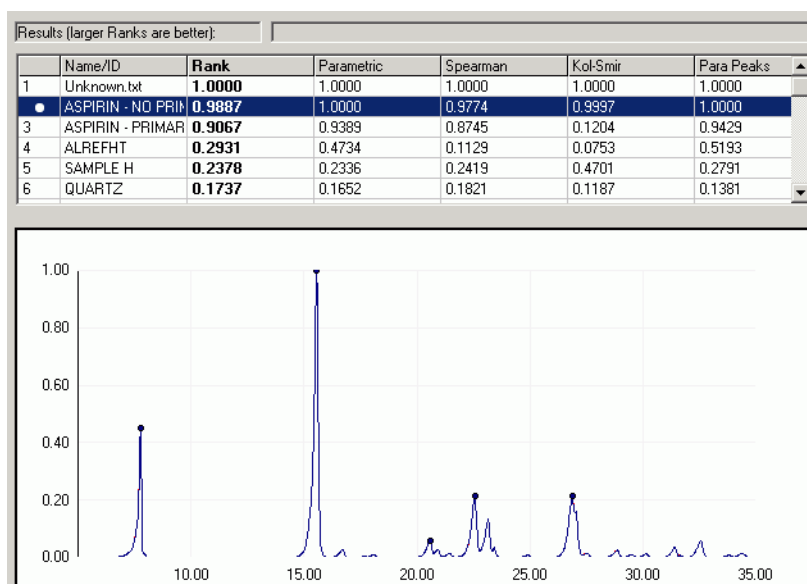
Results (larger Ranks are better):						
	Name/ID	Rank	Parametric	Spearman	Kol-Smir	Para Peaks
1	Unknown.txt	1.0000	1.0000	1.0000	1.0000	1.0000
2	ASPIRIN - NO PRIM	0.9887	1.0000	0.9774	0.9997	1.0000
3	ASPIRIN - PRIMAR	0.9067	0.9389	0.8745	0.1204	0.9429
4	ALREFHT	0.2931	0.4734	0.1129	0.0753	0.5193
5	SAMPLE H	0.2378	0.2336	0.2419	0.4701	0.2791
6	QUARTZ	0.1737	0.1652	0.1821	0.1187	0.1381

To sort the results by a different column, click once on that column's header. To change the sort order, *e.g.* to sort a column ascending instead of descending, or *vice versa*, click the header again. For the moment, click on the Rank header to re-sort by this column, and ensure the largest value (normally 1.0) is at the top of the list.

Look at the pattern associated with this value. This is what the program considers the 'best match' to our unknown. If it looks familiar, that is because it is – it is the *Unknown.txt* pattern itself. Because it is in the database, it is compared to itself. If this does not result in perfect match scores, there is a problem somewhere, so this is a useful check.

More interesting is the next entry down in the list. The next best match has a rank value of above 0.9, quite close to the perfect score of 1.0. The individual test scores reading along the row reflect this: all close to 1.0. The rank value is calculated by default by summing the Spearman and Parametric test scores and dividing by 2.0.

To see these scores reflected visually, we can overlay the best-match pattern with the unknown pattern. Click once in the left-most column of the pattern you wish to overlay (the one with '2' in it in this case).



The '2' is replaced by a coloured dot, and a pattern in this colour

appears in the graph pane. It is obvious the two patterns are very similar. Click on the coloured dot in the left-hand column once again to remove the extra graph.

This suggests our unknown sample is aspirin.

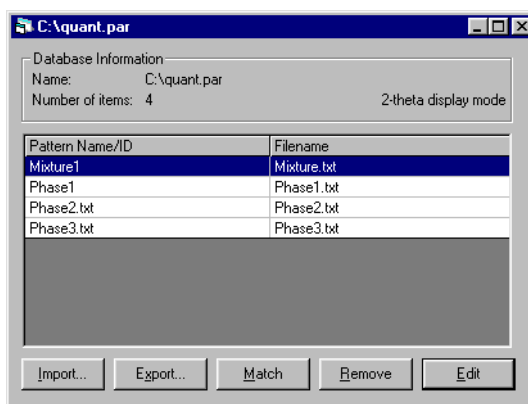
Experiment with overlaying different graphs on the unknown; this helps to get a feel for how the pattern scores correlate to similarity between patterns. Several patterns can all be overlaid at once on the graph; to clear them all in one go select *Clear Overlaid Graphs* from the *Graph* menu.

7.6.3 Analysing Mixtures Manually

Using PolySNAP, open the database file *quant.par* in the tutorial data folder. (Normally found in

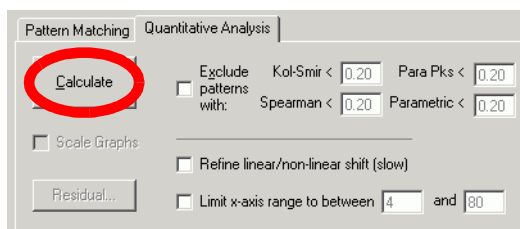
C:\Program Files\PolySNAP2\tutorial\manual)

It contains four patterns – one mixture, and the three pure component phases.



We will use the program to quantify the amounts of each phase in the mixture. Select the mixture from the list of patterns, and click *Match* to bring up the match window.

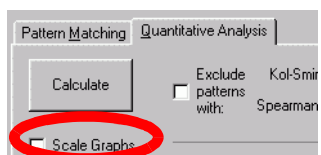
Go to the *Quantitative Analysis* tab, and click *Calculate*. After a few seconds when a progress bar is displayed, the program should display its answers.



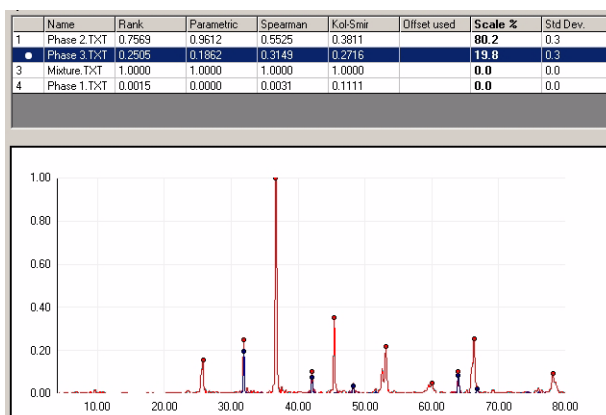
These are in a similar form as before, except that there are two additional columns – Scale % and Std Dev – the amount of each phase in the mixture, and the error on the calculation.

Analysis Results:							
	Name	Rank	Parametric	Spearman	Kol-Smir	Offset used	Scale %
1	Phase 2.TXT	0.7569	0.9612	0.5525	0.3811		80.2
2	Phase 3.TXT	0.2505	0.1862	0.3149	0.2716		19.8
3	Mixture.TXT	1.0000	1.0000	1.0000	1.0000		0.0
4	Phase 1.TXT	0.0015	0.0000	0.0031	0.1111		0.0

There are various ways available to see if the programs suggested percentages are sensible. First, check the *Scale Graphs* box on the upper left.

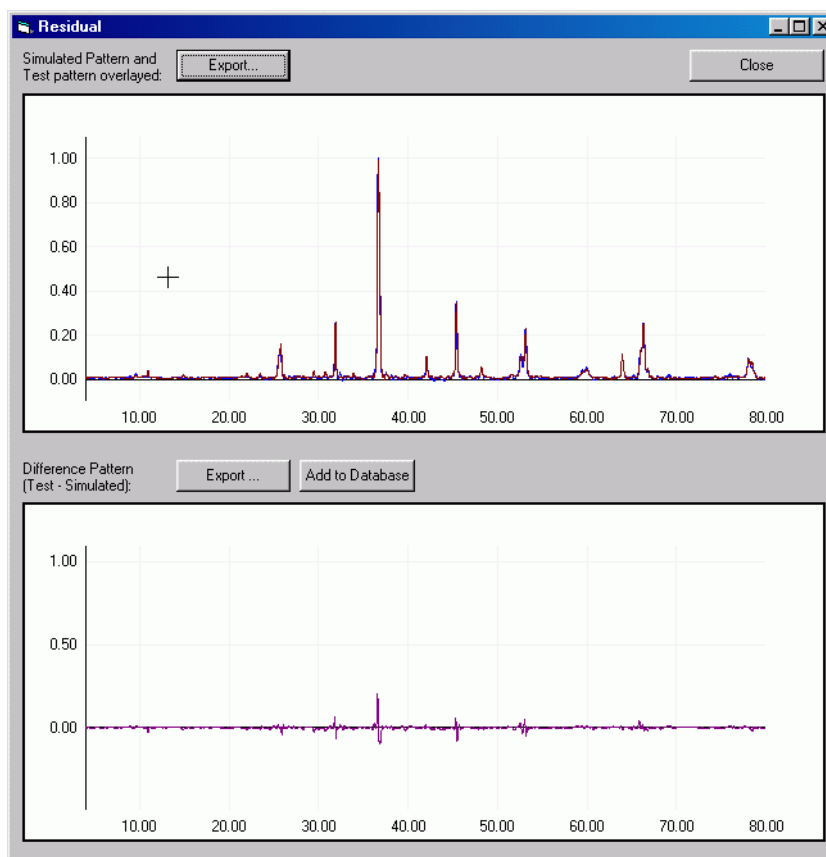


Then click on the left-most column of the results table for one of the phases.



As before, this overlays the selected pattern over the mixture, but is now scaled to the percentage intensity suggested by the programs analysis. The pattern should hopefully look sensible.

Another option is to click the *Residual* button, which brings up a window with two graph panes.

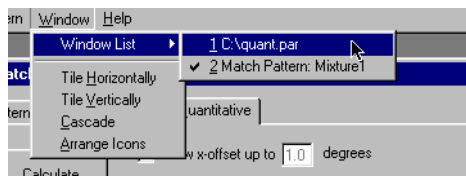


The upper one superimposes a simulated mixture pattern, made up of combining the pure phases in the amounts suggested by the program, on top of the original mixture. The bottom panel is then the difference between the two. This difference plot helps show up either missing phases or extra intensity which is not part of the mixture. If required, it can be output to a file to be imported as a new pattern at a later date. In this case, the small amount of residual intensity suggests the calculated answer is quite good.

Click *Close* to dismiss the window.

So far, we have quantified the mixture in terms of how much of each of the pure phase *patterns* are required to make it up. A more useful number in the real world would be in terms of the weight of each phase used in the mixture.

In order to calculate this, we need to add some additional information to each of the pure phase patterns. To do this, first select the database *quant.par* from the list of open windows in the *Window* menu.



This brings it to the front. Now select the first pure phase pattern: Phase 1. Click *Edit* to bring up the pattern editor window, and click on the *Advanced* tab.

 A dialog box titled 'Please enter additional known information:'. It contains input fields for 'Unit cell contents' (Al2 O3), 'a' (4.7592), 'b' (4.27592), 'c' (12.992), 'Z' (6), 'alpha' (90), 'beta' (90), and 'gamma' (120). Below these fields, it says 'Using the above information, the following have been calculated:' followed by a list of calculated values: 'Unit cell volume: 228.97', 'Formula Mass: 101.96', 'Density (g cm3): 4.44', 'Linear Abs. Coeff. (cm-1): 23.37', and 'Mass Abs. Coeff. (cm2/g): 5.26'. There are two buttons: 'Update' and 'Enter manually'.

We need to enter information into the lower region of this window, in particular the chemical formula of the phase, its unit cell dimensions, and the number of formula units per unit cell.

Phase	Formula	a	b	c	α	β	γ	Z
Phase 1	Al2 O3	4.7592	4.7592	12.992	90	90	120	6
Phase 2	Ca F2	5.4649	5.4649	5.4649	90	90	90	4
Phase 3	Zn O	3.2501	3.2501	5.2071	90	90	120	2

Using the information in the table, enter the relevant details for this phase, and click *Update*. The fields showing molecular weight and absorption coefficients and so on for this phase should update. If not, be sure to enter the formula in the form

<Atomic Symbol><No. of atoms> <AtomicSymbol><No. of atoms> etc.

(The allowed formats for formula entry are discussed in more detail in section 3.3.11, Additional Pattern Information in the main program manual).

Repeat for the other two phases.

Click on the mixture pattern database entry to select it, before clicking the *Match* button to bring it up in the main match window.

Return to Match window, and in the *Quantitative Analysis* tab, click *Calculate* again.

This time, the results table should be slightly different – the results should be headed Weight % rather than Scale %. If not, go back and check that all of the three pure phases has had the extra information added successfully.

Analysis Results:							
	Name	Rank	Parametric	Spearman	Kol-Smir	Offset used	Weight %
1	Phase 2.T\T	0.7569	0.9612	0.5525	0.3811		67.6
2	Phase 3.T\T	0.2505	0.1862	0.3149	0.2716		32.4
3	Mixture.T\T	1.0000	1.0000	1.0000	1.0000		0.0
4	Phase 1.T\T	0.0015	0.0000	0.0031	0.1111		0.0

7.7 Conclusion

This completes the basic tutorial. There are many other features and options in the program that should allow much more complex problems to be examined. Each of these are described in the full program manual, which should be consulted for more information.

