CHAPTER 5

PolySNAP M Tutorial

5.1 Introduction

This tutorial has been designed to guide the user through a few examples using PolySNAP M 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 M, 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\PolySNAP M.*

5.1.1 Simple Automatic Pattern Matching

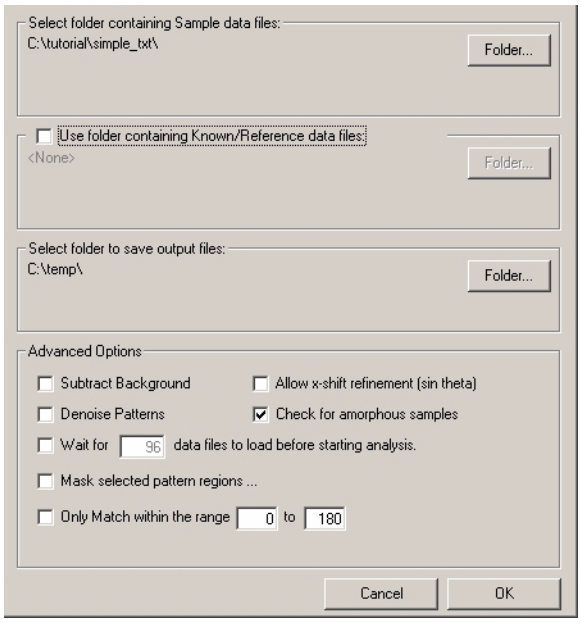
To gain experience with PolySNAP M, a simple run using 21 X-ray powder diffraction patterns is performed. The example assumes the program defaults are used, and any more advanced options are unnecessary.

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

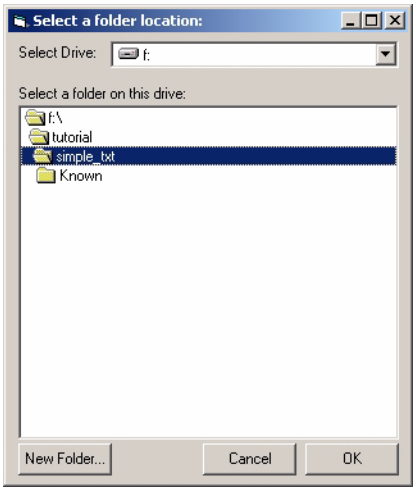
5.1.2 Selecting Input and Output Folders

Once PolySNAP M has been launched the user is presented with an empty window that contains only a menu bar. From this, the first step is to define the input folder containing the necessary patterns and the output folder where the results are to be placed.

From the *File* menu, select *Automatic Analysis* and the *Run on...* option. 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 *D:*.
3. Navigate to the folder *C:\Program Files\PolySNAP M\tutorial\simple_txt*. Ensure that the last folder of the desired path (*i.e. simple_txt*) 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. In the *Run PolySNAP M on...* dialog box again, ensure that the central section of the dialog box has no defined folder, indicated

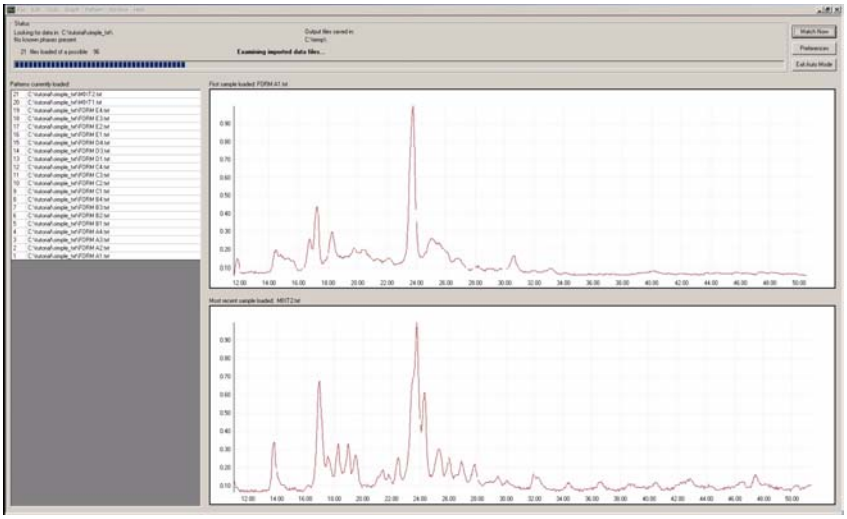
by the word '<None>'. If this is not the case, uncheck the check-box labelled *Use Folder containing Known/Reference data files*. The use of this setting will be explained later.

6. The output folder where the result files are to be placed must be defined. Click the third *Folder* button in the next section of the dialog box. A folder selection dialog box will appear as before.
7. Select the local hard disk on your machine, which in this case is *C:*. (Alternatively any other drive you wish to use for the output data may be selected.)
8. Navigate to the folder you wish to place the output files. e.g. *C:\output*. If required use the *New Folder* button to create a new folder. Ensure that the last folder of the path is selected by double clicking it (the folder icon should appear 'open'). Note that you **cannot** use the same directory for both input and output.
9. Click *OK*, and check that the desired path is displayed correctly.
10. Leave the Advanced Options settings to their default values (all should be off except for *Check for amorphous samples*).
11. Finally, in the *Run PolySNAP M on...* dialog box click *OK*.

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

5.1.3 Pattern Matching the Different Powder Patterns

PolySNAP M is now in automatic mode, and proceeds by reading the sample files from the specified input folder, and loading these into a database located in the output folder.

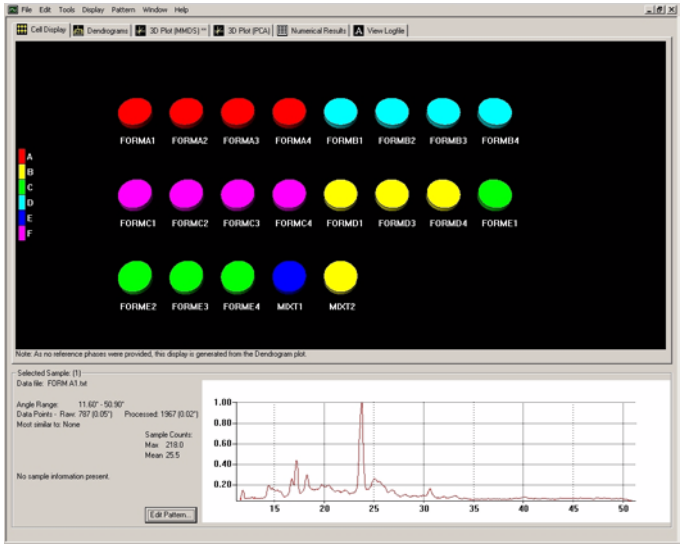


Each pattern is loaded in turn, with the first pattern profile shown on the top right and the most recently loaded profile below it. When all 21 patterns have been loaded the process will stop.

The program now examines the input files, checking for amorphous samples, matches the data files against one another, and then performs cluster analysis. When complete the results window will appear.

5.1.4 Viewing the Results in the Cell Display

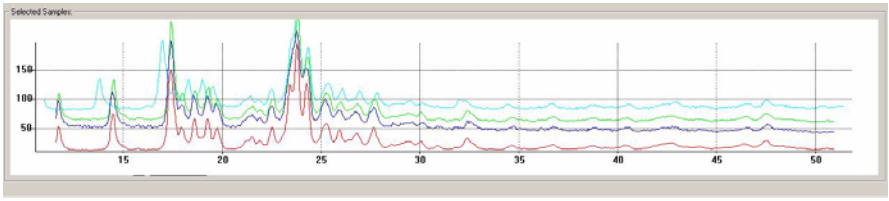
There are now a number of ways to view and examine the PolySNAP M results. The first of these, and default view displayed is the *Cell Display*, which visually represents the contents of each sample cell.



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.

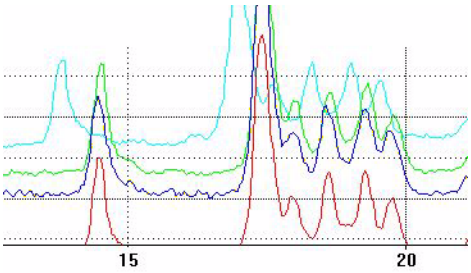
1. In the *Cell Display* click on cell *FORMD1*. The associated full pattern profile of pattern D1 is now displayed in the lower region of the window beside the sample information. In this particular case the information available is limited, so many of the fields are blank.
2. Hold down the *shift* 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 cell *MIXT_2*. 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 M program.



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

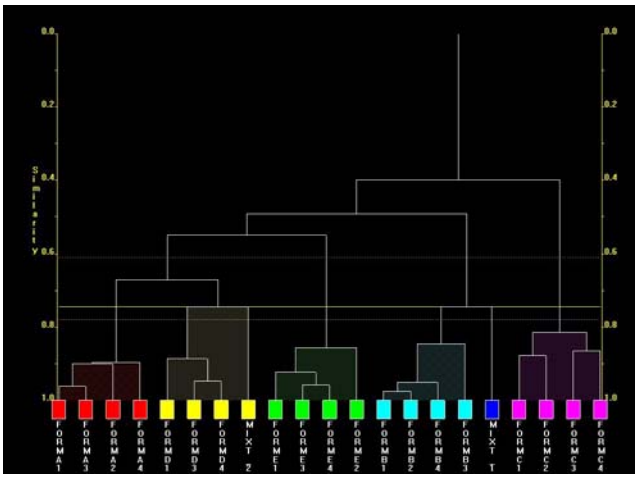
This closer view of the overlaid patterns makes it easier to see that the top pattern (MIXT_2) is noticeably different from the other patterns - note the extra peak around 14° for example - and may not actually belong to this group of compounds.

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

5.1.5 Viewing the Results by Dendrogram

In the PolySNAP M window 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 M 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).



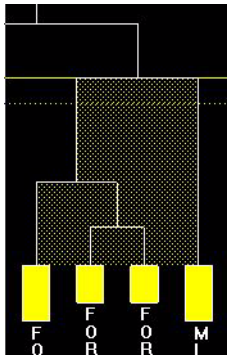
A dendrogram provides a visual display of the results from the hierarchical method of data classification using cluster analysis. The dendrogram itself takes the form of a tree-diagram in which each terminal branch (coloured box) is representative of a single pattern sample.

The higher up the similarity scale two samples are connected by a horizontal line, the less similar they are. Hence samples FORMA1 and A3, 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 M displays. The number of clusters are defined by the yellow cut-line which in this case was initially set to 0.744. 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.

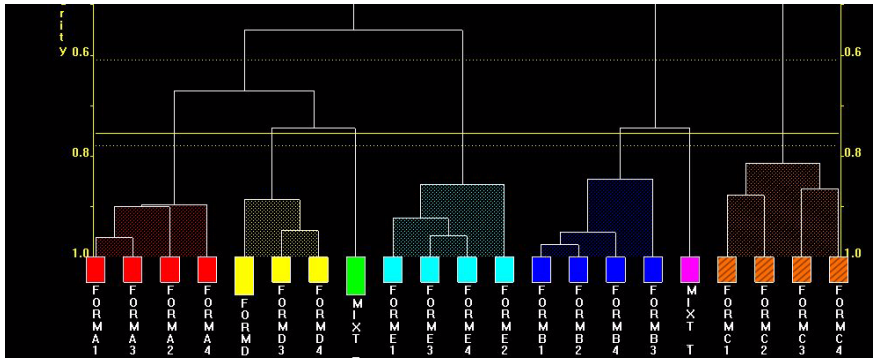
Adjusting the cut level upwards creates fewer separate clusters, and effectively reduces the discrimination between differences; adjusting the cut-level downwards creates more separate clusters.

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 *MIXT_2*. These are the two patterns that appeared different from one another when they were overlaid in the cell display.



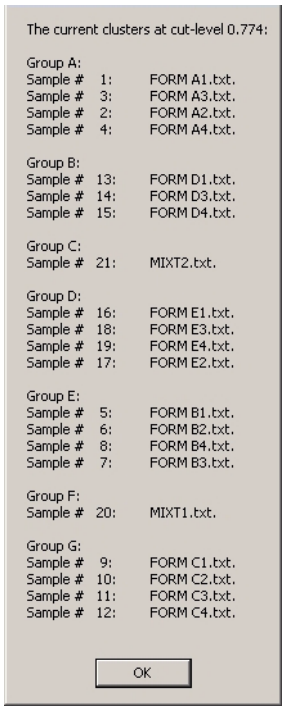
Looking at the position of the cut-line it is seen to be only very slightly above the similarity line between cell MIXT_2 and the other yellow cells. It can also be seen that another group (the dark blue and light blue samples) has only just been split by it. With the confidence levels indicated by the yellow lines being between 4 and 7 clusters, it is possible that the present level is not ideal.

3. 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 cell 20 is now in a cluster of its own.



4. In the dendrogram area click the *right mouse button* and select *Save Modified Trees....*, to ensure that the changes are retained.

5. From the *Tools* menu select *List Pattern Cluster Members....* A dialog box will appear containing a list of the clusters.

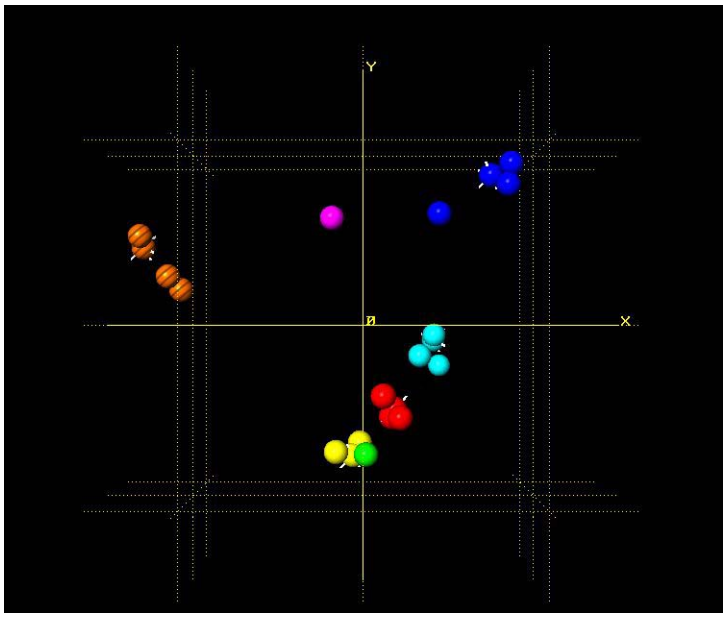


This dialog box displays the different clusters with their sample numbers and actual file names. From the file names used in this demonstration example it can be seen that the samples are now all properly grouped, with the two mixtures separate from the rest. Notice in the dendrogram display that these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

5.1.6 Viewing the Results by 3-D Plot

The samples can also be viewed by making use of the distances between objects 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 MIXT_2 and, when comparing to the dendrogram display, 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 measuring 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 *Centre Selection*
Reset view	Right click mouse and select *Reset View*

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.

To end this session select *Close Window* from the *File* menu.

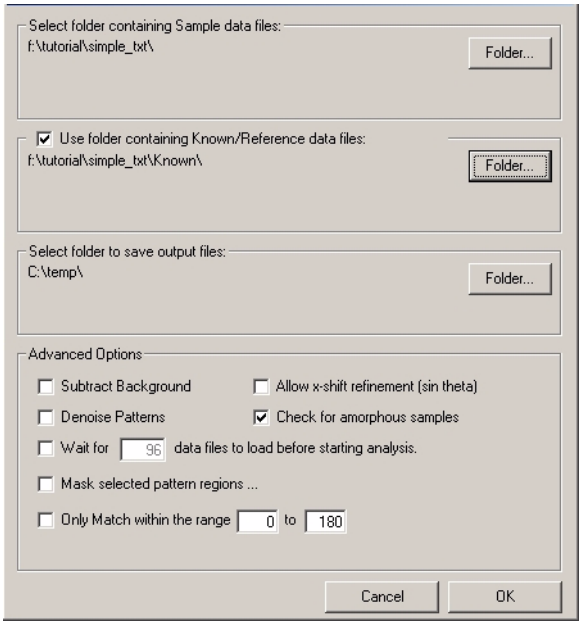
Automatic Quantitative Pattern Matching using Known Phases

A more sophisticated pattern match might include a database of known phases for comparison purposes. These are patterns that have been previously identified and can be used to decide whether the sample is a previously known substance, a mixture of known materials, or something completely different.

5.1.7 Selecting the Input, Output and Known Phase Folders

To begin, the user should start with the empty PolySNAP M window containing the menu bar, as in the previous example.

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



Set the input, output and known phase folders as before, except this time we will additionally select a folder containing some known phases prepared earlier.

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 *D:*.
3. Navigate to the folder *C:\Program Files\PolySNAP M\tutorial\simple_txt*. Ensure that the last folder of the path (*simple_txt*) 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. The known phase folder containing the known sample data files must be defined. Turn **on** the checkbox labelled *Use Folder or Database containing Known/Reference data files*. Click the *Folder* button in the that section of the dialog box.
6. Select the drive containing the tutorial data, in this case *D:*.
7. Navigate to the folder *C:\Program Files\PolySNAP M\tutorial\simple_txt\known*. Ensure that the last folder of the path (*i.e. known*) is selected by double clicking it (the folder icon should appear 'open'), and click *OK*. This folder contains an example pattern of each of Form A, B, C *etc.*

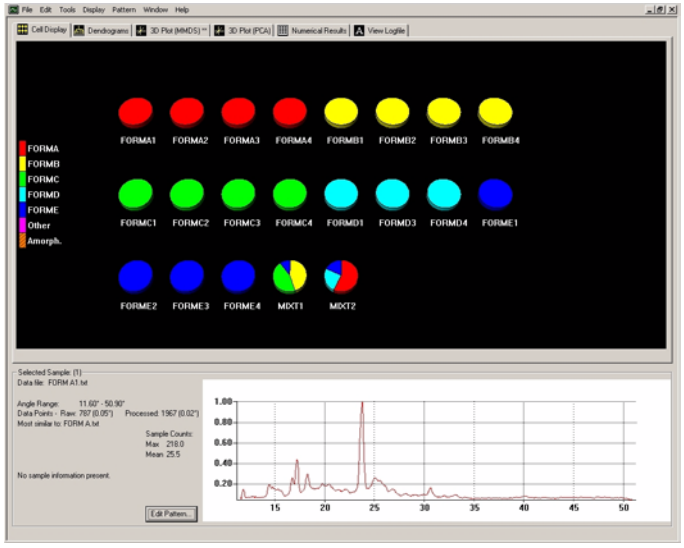
8. The output folder where the result files are to be placed must be defined. Click the third *Folder* button in the bottom section of the dialog box. A folder selection dialog box will appear as before.
9. Select the local hard disk on your machine, which in this case is *C:*. (Alternatively any other drive you wish to place the output data may be selected.)
10. Navigate to the folder you wish to place the output files. e.g. *C:\output*. If required use the *New Folder* button to create new folders. Ensure that the last folder of the path is selected by double clicking it (the folder icon should appear 'open').
11. Click *OK*, and check that the desired path is displayed correctly.
12. Finally, in the *Run PolySNAP M on...* dialog box click *OK*.

The pattern files in the specified input folder are now loaded into PolySNAP M, which proceeds to the matching and analysis phase automatically after loading the 21 datafiles.

When complete the results window will appear, with an initial view of the *Cell Display* as before.

5.1.8 Viewing the Results

The input files are now displayed in the *Cell Display* as in the first tutorial example. However, it now looks a little different with cells MIXT_1 and MIXT_2 being shown as pie charts.

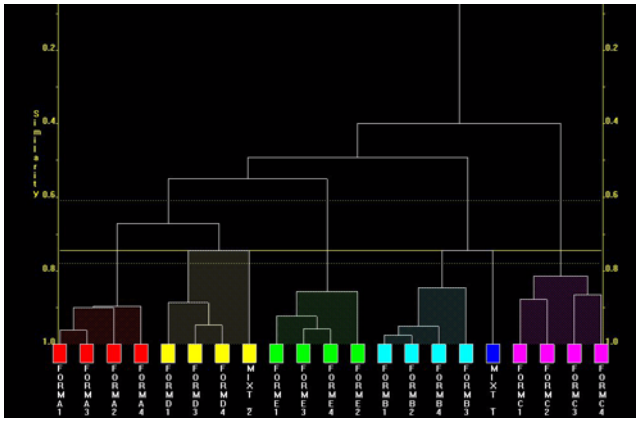


Unlike the previous run, the colours here are **not** taken from the dendrogram plot, but from comparing each sample in turn to the provided known phases. Samples such as MIXT_1 and 2 that do not give high matching correlation coefficients to any of the known reference patterns are tested to see if they are mixtures, and if so, quantitative analysis is performed. The results from this are

displayed in pie-chart form, the colours of each slice corresponding to the pure phases listed in the key on the left hand side.

Another change to the view is the sample key - it now shows the known phases FORM A to FORM E. By clicking on the colours of the key all the samples matching the designated phase are selected. For example, clicking on the red box labelled FORMA, cells A1 to A4, as well as MIXT_2 are selected, as they are all of the samples thought to either be, or contain, Form A. The key titles are taken from the reference phase filenames.

Select the tab labelled *Dendrogram*. This view is exactly the same as in the first example, including cell 20 being contained within the yellow group.



The reason for this difference between views is that the cluster analysis does not take the known phases into account. Only the results shown in the *Cell Display* are directly affected by them. Because of the way they work, the dendrograms and other methods merely show the similarity between samples without taking account of the known phase information.

With the quality of data available for this sample the predicted mixture composition is not very good, and the quantitative results should therefore be taken as a very approximate measure, certainly not accurate to any more than around 10% at best. This approximate value could however be useful as a starting point for a more rigorous analysis later.

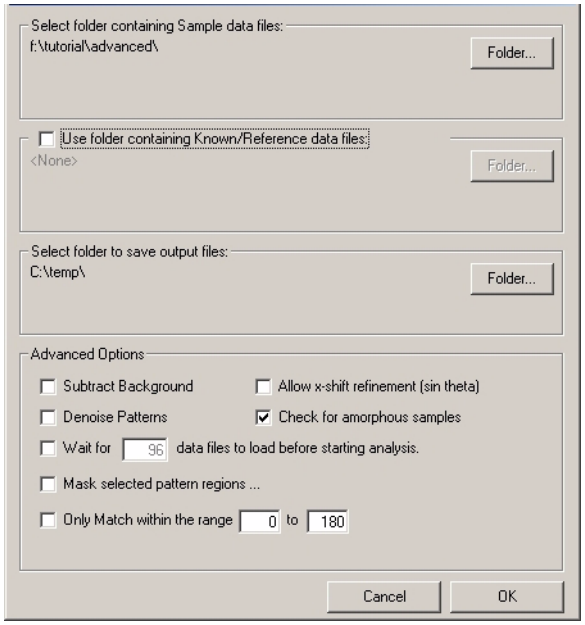
5.2 Identification of Amorphous Samples

The first two tutorial examples of this tutorial have covered simple PolySNAP M procedures. 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.

To begin, the user should start as before with only the empty PolySNAP M window containing only the menu bar as before.

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



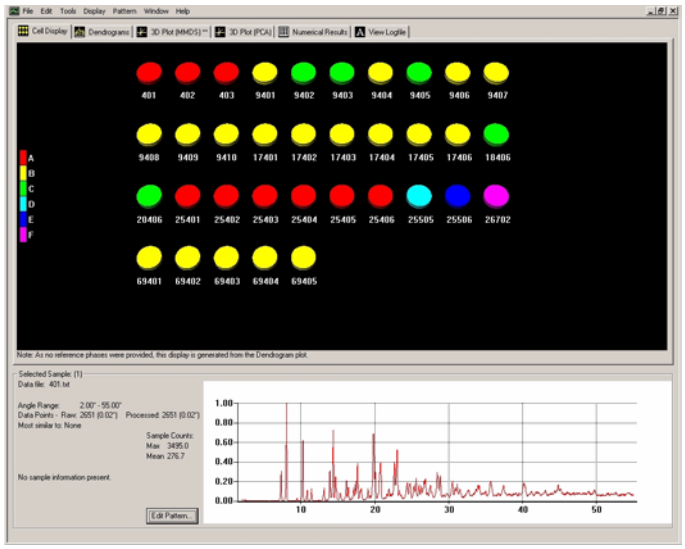
1. To define the input database containing the sample data, click the *Folder* button.
2. Select the drive containing the tutorial data, in this case *D:*.
3. Navigate to the folder*C:\Program Files\PolySNAP M\tutorial\advanced*
4. Ensure that no known phases are used by making sure the relevant checkbox is turned off.
5. Define the output folder where the result files are to be placed. Click the third *Folder* button in the bottom section of the dialog box. A dialog box will appear.
6. Navigate to the folder you wish to place the output files. e.g. *C:\output*. If required, use the *New Folder* button to create new folders, ensuring that the last folder of the path is selected by double-clicking on it (indicated by the folder icon being in an open position). Click *OK*.
7. Leave the Advanced Options settings at their default values.
8. In the *Run PolySNAP M on...* dialog box click *OK*.

The pattern files in the specified database are now loaded into PolySNAP M, where it checks for amorphous samples, matches the patterns with one another, and performs a cluster analysis. This may

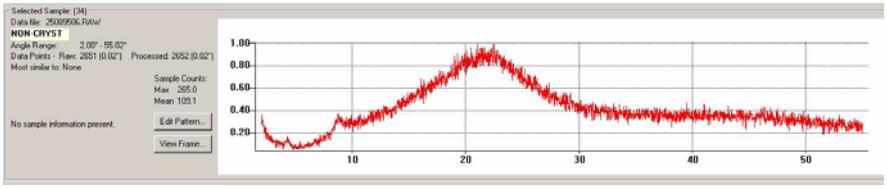
take a little time to complete, but once finished the results window will appear with an initial view of the *Cell Display*.

5.2.1 Analysis of the Results

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

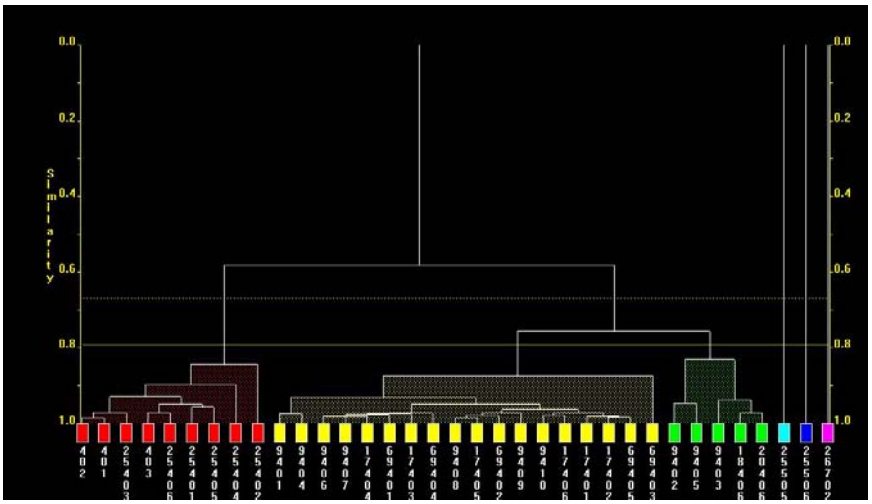


The program is using the last few digits of the filename of each sample to label them. Selecting pattern number 25505, the pattern information display shows that it has been labelled as a non-crystalline sample, which seems reasonable given its profile.

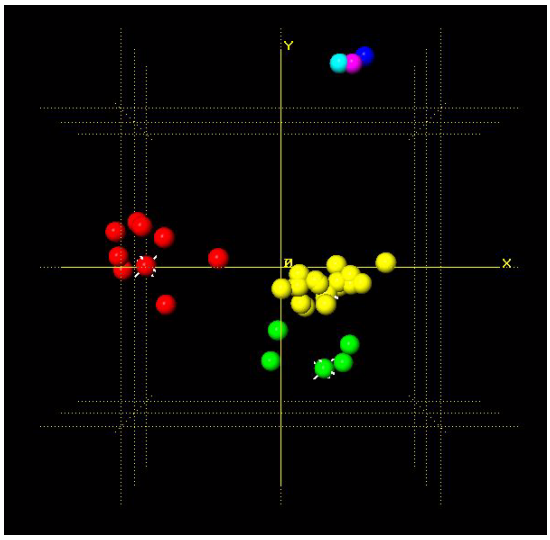


The other two patterns, 25506 and 26702 are similarly labelled for the same reason. Identification of such amorphous samples is done on the basis of checking to see if any signal (corresponding to peaks) would be left after subtraction of the entire amorphous hump. The method tends to err on the side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

Looking at the *Dendrogram* tab:.



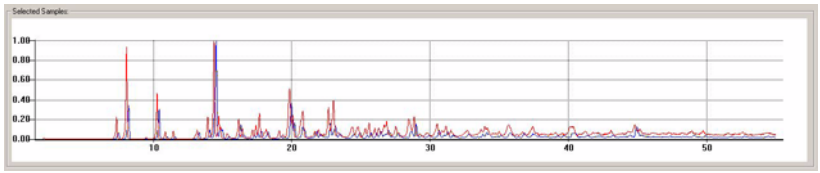
It is seen that three non-crystalline labelled patterns are placed on the far right of the diagram with a zero similarity to the rest. This is deliberate, in order to remove them from the main clusters. There also appears to be one main grouping of the data - within which some patterns seem much more closely grouped than others. 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 main (yellow) group from each other, for example patterns 9401 and 69403. Examining their profiles, it appears that in addition

to some preferred orientation issues, there seems to be a noticable 2θ-shift between the otherwise relatively similar profiles:



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.*

5.2.2 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 with the Bruker GADDS systems if the sample height varies from sample to sample. This can give 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:

$$\Delta(2\theta) = a_0 + a_1 \sin\theta \qquad (5.1)$$

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 default maximal allowed values for a_0 and a_1 are 0.4.

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.

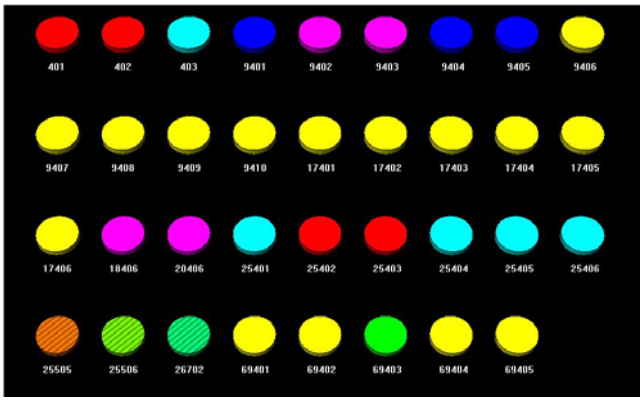
Define the input database, known phases and output folder as before:

1. From the *File* menu, select *Automatic Analysis* and the *Run on...* option. Define the same input folder as used before in the previous run of PolySNAP M. *D:\tutorial\advanced*
2. Select *None* for the known phases directory.
3. Ensure that an appropriate output folder is selected. e.g. *C:\output*.
4. In the Advanced Options area, turn on the *Allow x-shift calculation (sin theta)* checkbox.

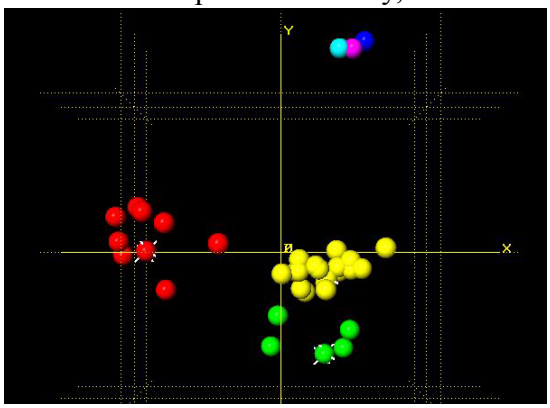
5. In the *Run PolySNAP M on...* dialog box click *OK*.

The same data as before is now run allowing x-shifts on the patterns. The time required to process this will take much longer than a normal run of PolySNAP M. For this data the calculation will take approximately ten times longer than before - so prepare to be patient!

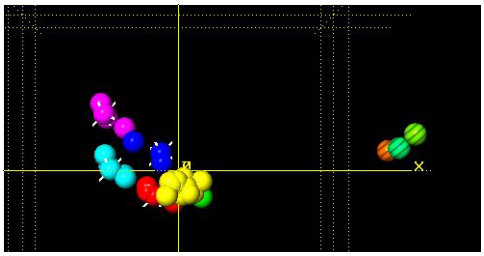
When finished, the initial output should look similar to the previous run:



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

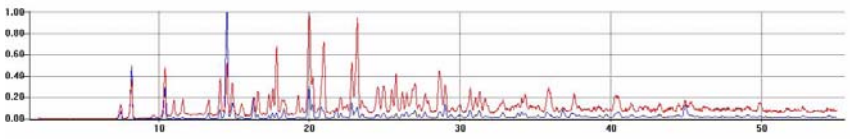


Now, with the option to calculate the best-offset value for each pattern turned on, it looks like this:



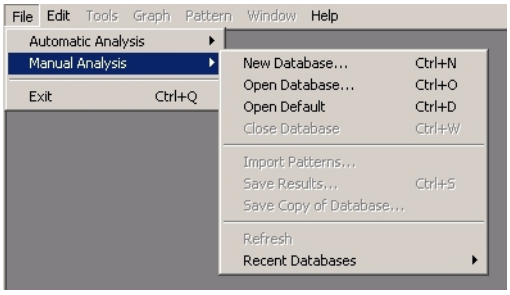
The three non-crystalline patterns are still quite separate, but the rest of the patterns have condensed together as a result of allowing for the 2θ-shift, showing that a large part of the differences between the pattern profiles was due to variation in sample heights during data collection. The program 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 noticable in some cases. For example, overlay the profiles of samples 401 and 69403:



5.3 Manual Matching: Working with Databases

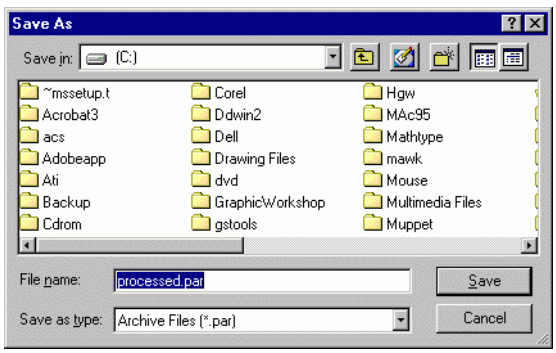
The rest of the examples in this tutorial demonstrate the manual analysis section of PolySNAP M. This mode can be accessed from the *Manual Analysis* section of the *File* menu.



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

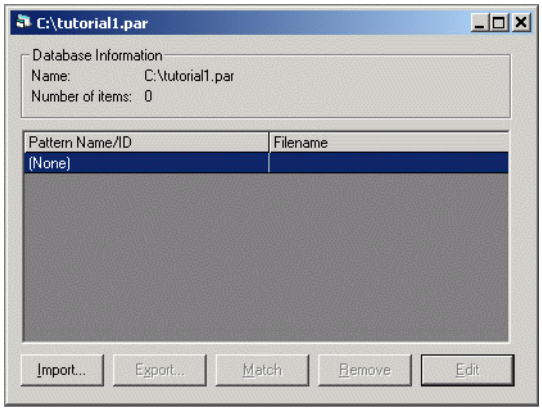
We first need to create a new database, which will initially be empty We will then load some powder patterns into it, and work with them.

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 M 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.

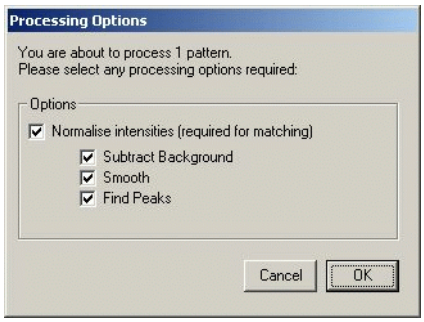
Several buttons are arranged along the bottom of this window. Most of these are unavailable as there are no data files in the database. The only one that is available for use is the *Import* button.

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

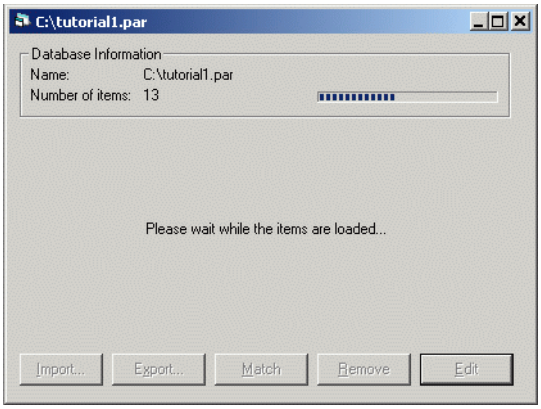
Click the *Import* button. We need to navigate to where the data is stored - *C:\Program Files\PolySNAP M*. The folder *manual* should be found within the '*tutorial*' folder.

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

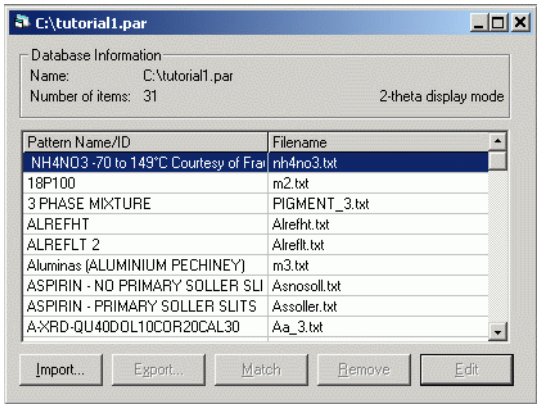
We want to open all of the files in this particular folder, so hit *control-A* on the keyboard to select them all, and then click *Open*. A new window appears with the pattern processing options.



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 M 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.

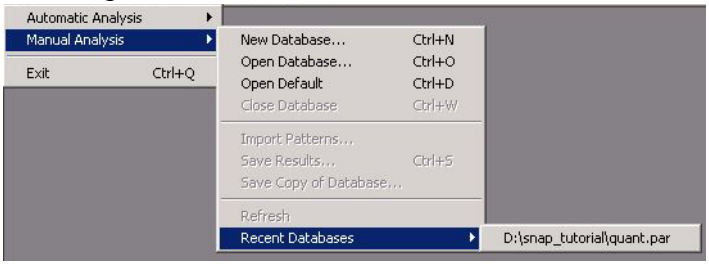
Removing patterns from the database does not in any way affect the original data files which are left intact and unchanged throughout all operations performed using PolySNAP M - the program works on a copy of the data only.

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

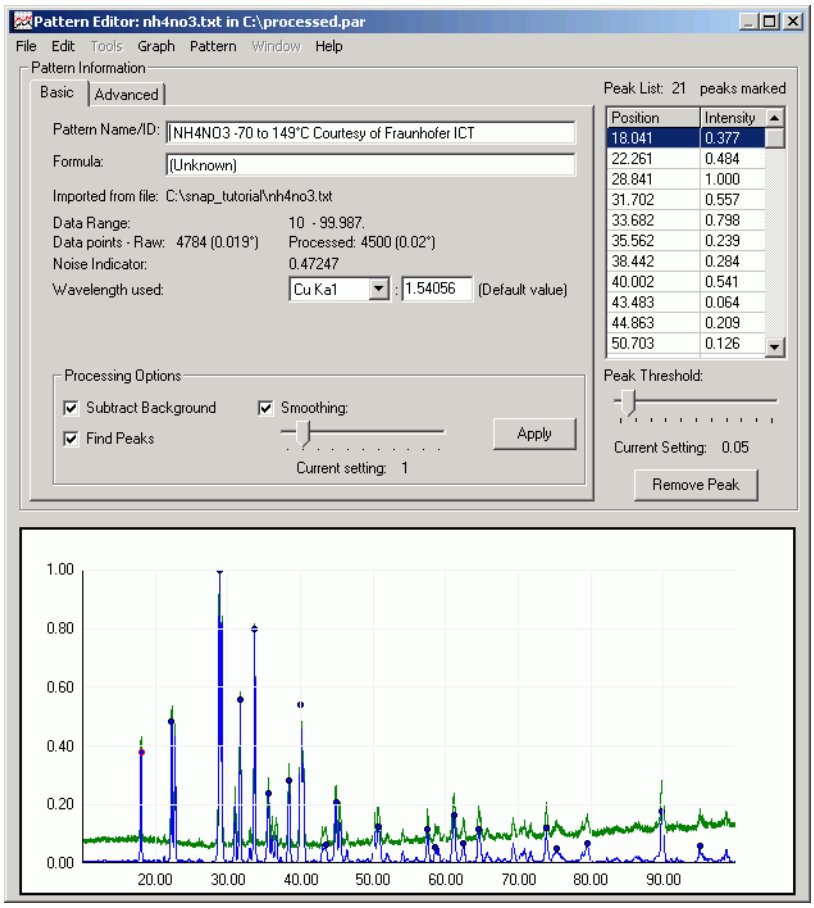


tutorial1.par database we were using earlier. Select it from the menu, and it should start to open automatically.

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

5.4 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.

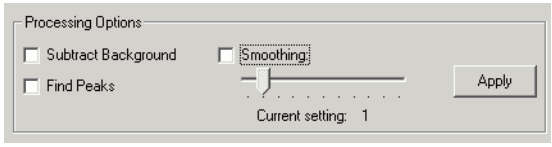


Several useful pieces of information about the pattern you have selected are displayed in this window, and a plot of the pattern itself is shown at the bottom.

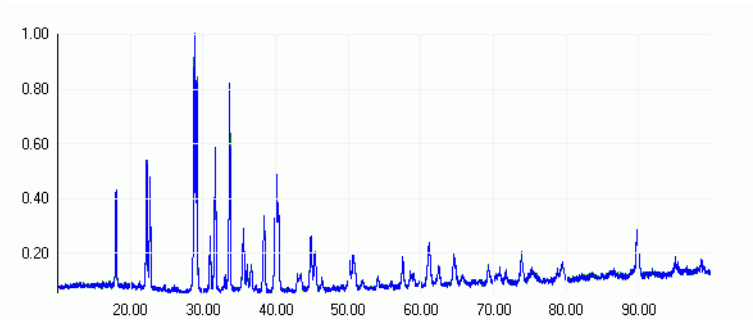
The chemical name or pattern ID can be changed here by editing the current name (*"NH4NO3 -70 to 149°C Courtesy of Fraunhofer ICT"*) in the white text box.

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:



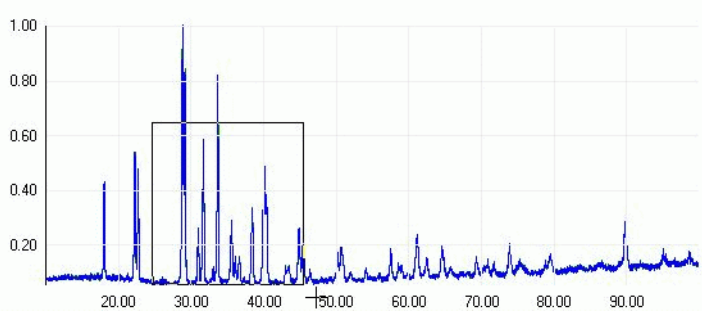
Then click *Apply*. The graphing region should now look like this:



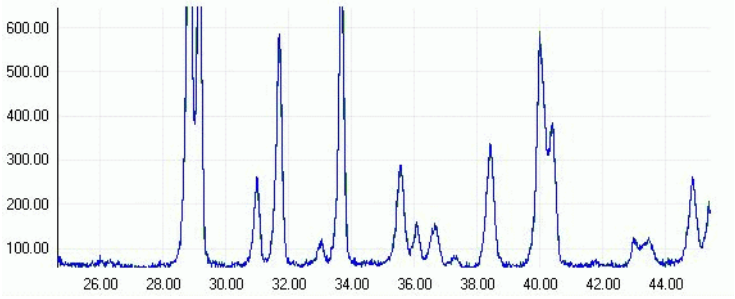
As you can see, the data has been scaled along the y-axis to run from 0.0 to 1.0. This operation is performed for all data being imported into PolySNAP M in order to allow for suitable scaling between different data sets.

The raw data has also been interpolated from its original resolution, to the PolySNAP M 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:

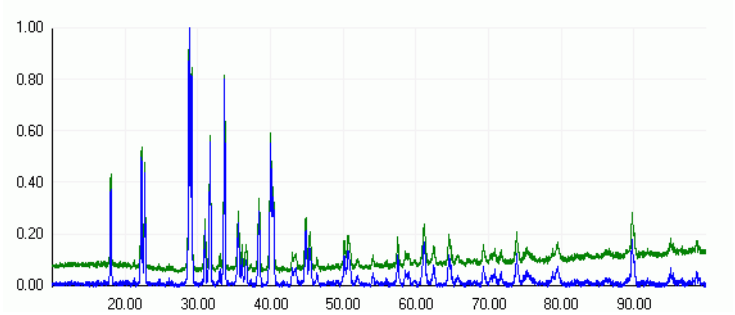


The view will then change to just the selected region:



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.

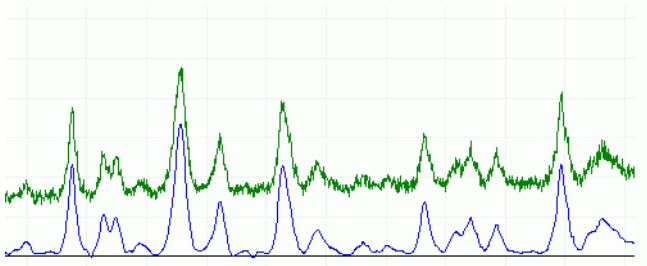
Now click the *Subtract Background* checkbox, and then click *Apply*. The pattern display should change:



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.

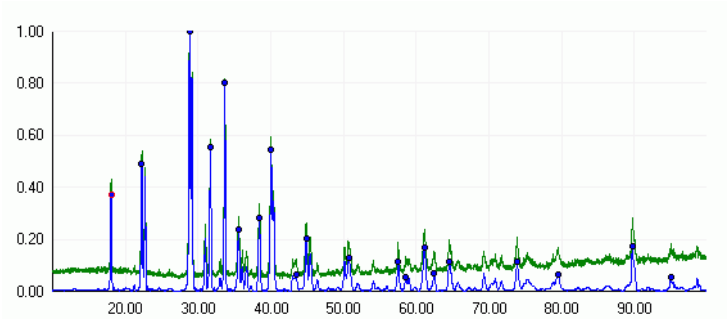
Now click the *Smoothing* checkbox, so both it and the *Background Subtraction* boxes are checked, and click *Apply* again.

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



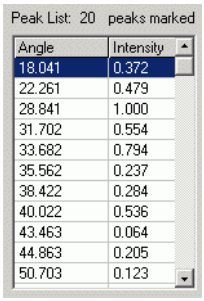
Finally, check the *Find Peaks* option, and click *Apply*.

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 peaks. 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.

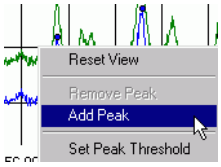
A list of the peaks the program has found is shown in the upper right corner of the pattern editor:



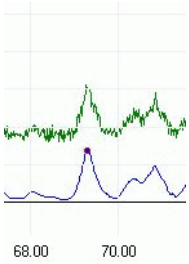
This lists the peaks found in order of increasing Bragg angle, and displays the corresponding intensity for each.

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:

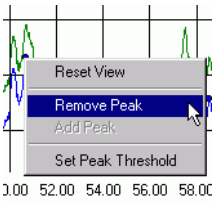


Then select *Add Peak* from the resulting pop-up menu. A round blue peak marker should appear at the top of the peak, and an entry for it should be added to the peak list:

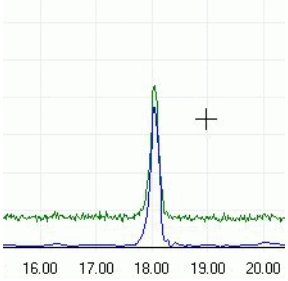


If this does not appear, try clicking as close as possible to the top of the peak. It may help to zoom in to the area of interest.

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°:

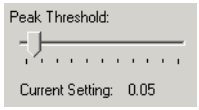


Right click as close to the peak marker as possible, and select *Remove Peak* from the pop-up menu that should appear. The peak marker, and its corresponding entry in the peak list, should vanish.



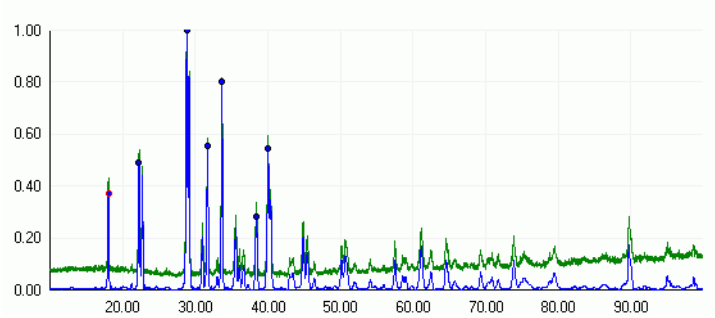
If the *Remove Peak* option is unavailable from the pop-up menu, you have not clicked close enough to the marker. Zoom in to make this easier, and try again.

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.



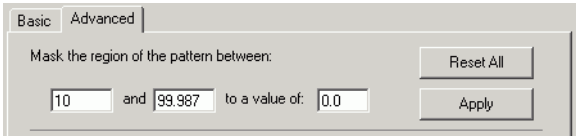
Drag the handle on the slider until it is set to about 0.25.

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.



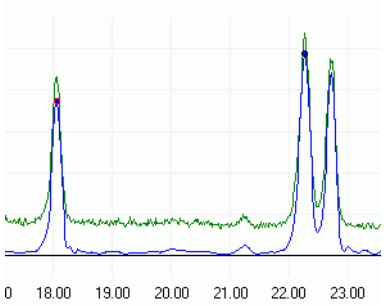
Finally, return the slider to 0.05. The peaks that previously became unmarked should show up again.

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

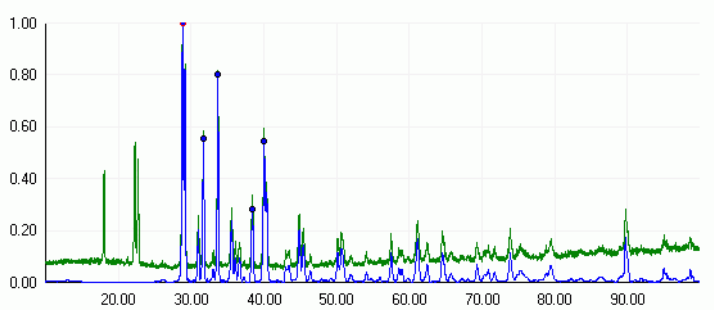


The top section revealed allows us to mask selected regions of the pattern to be ignored during matching or analysis processes. This may be useful where a particular peak is negatively affecting results. Say we wish to mask the peaks in the region of the pattern between 15° and 25°. Enter the start and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

5.5 Matching Patterns

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

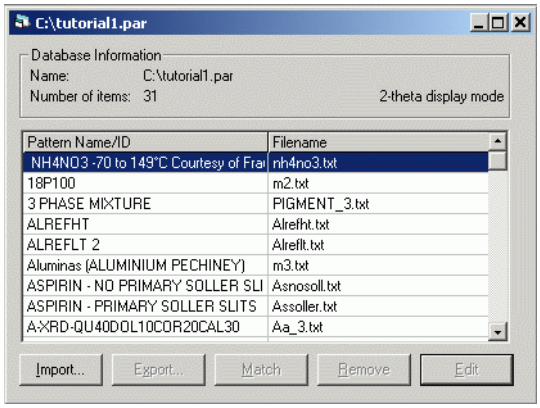
You want to find out what pattern in your existing database it is most similar to - and hence possibly identify what substance it may represent.

First, close any currently open database or editor windows.

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

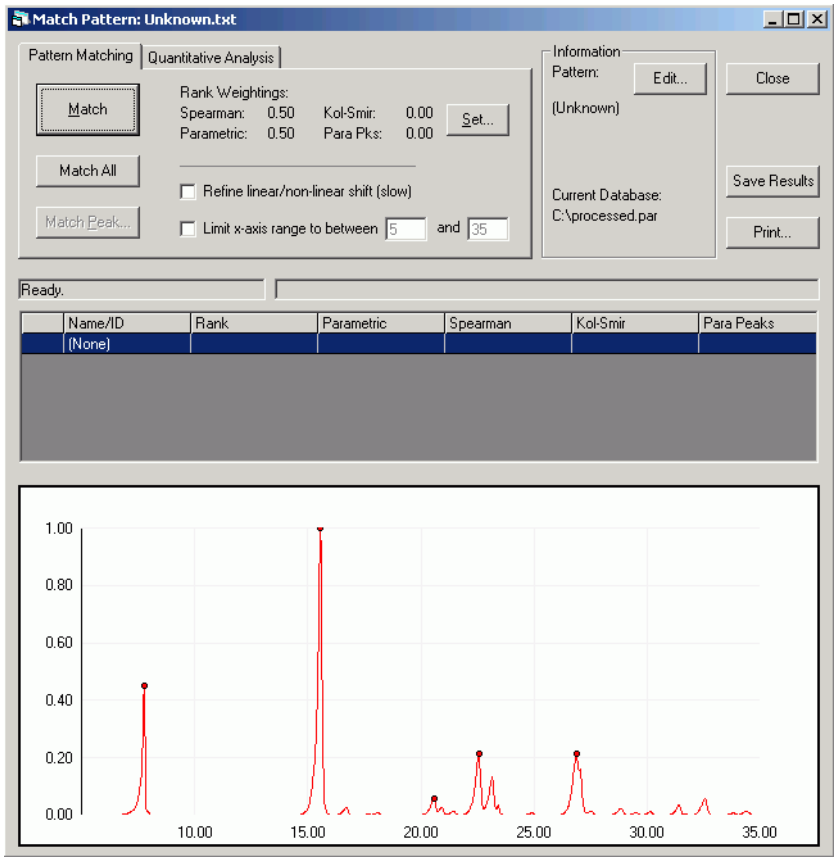
Open the pattern database you created earlier, *c:\tutorial1.par*.

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.

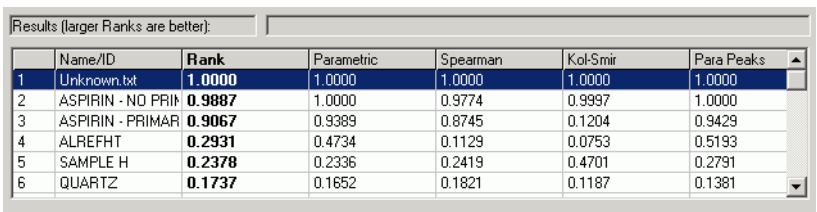
Locate the pattern *Unknown.txt* in the scrolling list of patterns in the database, click once to select it, and then click on the *Match* button. The main program match window will appear.



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.



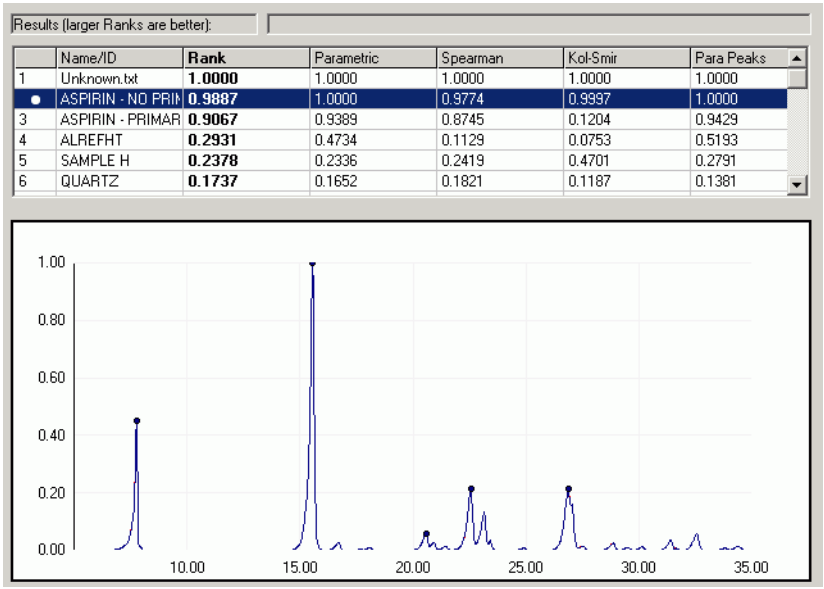
	Name/ID	Rank	Parametric	Spearman	Kol-Smir	Para Peaks
1	Unknown.txt	**1.0000**	1.0000	1.0000	1.0000	1.0000
2	ASPIRIN - NO PRI	**0.9887**	1.0000	0.9774	0.9997	1.0000
3	ASPIRIN - PRIMAR	**0.9067**	0.9389	0.8745	0.1204	0.9429
4	ALREFHT	**0.2931**	0.4734	0.1129	0.0753	0.5193
5	SAMPLE H	**0.2378**	0.2336	0.2419	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.

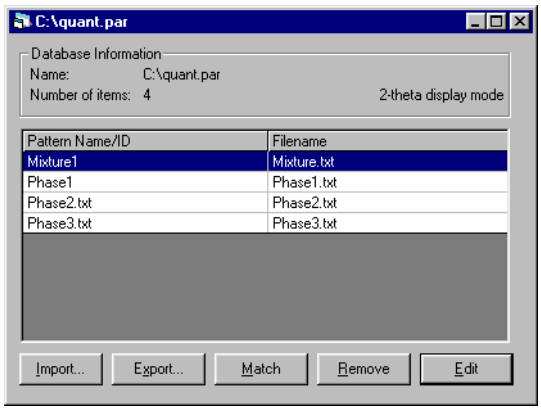
Experiment with overlaying different graphs on the unknown; this helps to get a feel for how the pattern scores correlate to similarity

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.

5.6 Analysing Mixtures Manually

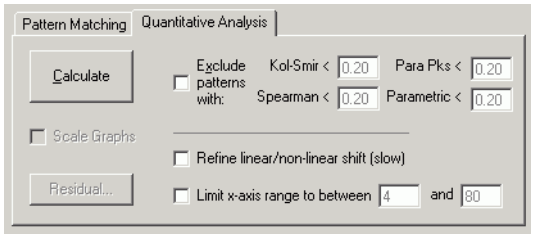
Using PolySNAP M, open the database file *quant.par* in the tutorial data folder. (Normally found on the PolySNAP M CD-ROM in the 'tutorial files', *e.g. tutorial -> manual_data*).

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

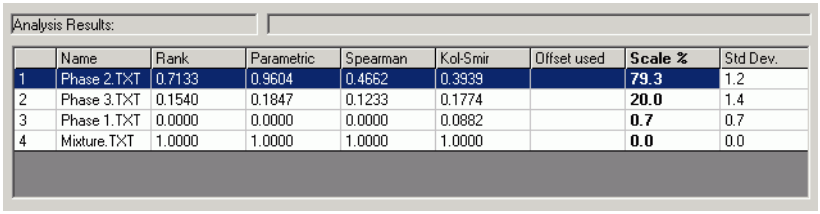


We will use the program to quantify the amounts of each phase in the mixture. Select the mixture from the list of patterns, and click *Match* to bring up the match window.

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.

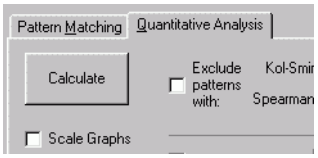


These are in a similar form as before, except that there are two additional columns – Scale % and Std Dev – the amount of each phase in the mixture, and the error on the calculation.

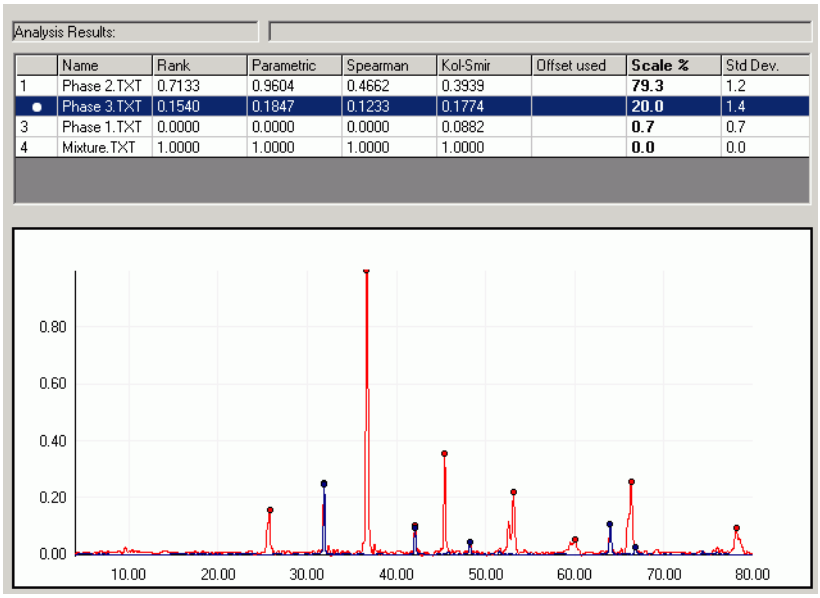


	Name	Rank	Parametric	Spearman	Kol-Smir	Offset used	Scale %	Std Dev.
1	Phase 2.TXT	0.7133	0.9604	0.4662	0.3939		**79.3**	1.2
2	Phase 3.TXT	0.1540	0.1847	0.1233	0.1774		**20.0**	1.4
3	Phase 1.TXT	0.0000	0.0000	0.0000	0.0882		**0.7**	0.7
4	Mixture.TXT	1.0000	1.0000	1.0000	1.0000		**0.0**	0.0

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



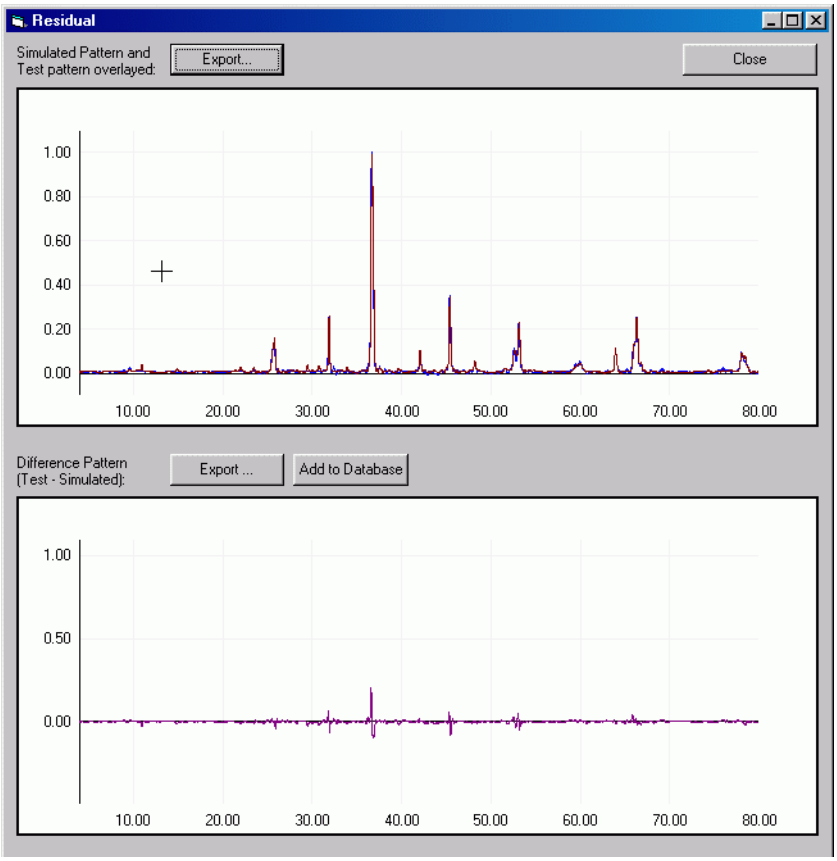
Then click on the left-most column of the results table for one of the phases.



	Name	Rank	Parametric	Spearman	Kol-Smir	Offset used	Scale %	Std Dev.
1	Phase 2.TXT	0.7133	0.9604	0.4662	0.3939		**79.3**	1.2
●	Phase 3.TXT	0.1540	0.1847	0.1233	0.1774		**20.0**	1.4
3	Phase 1.TXT	0.0000	0.0000	0.0000	0.0882		**0.7**	0.7
4	Mixture.TXT	1.0000	1.0000	1.0000	1.0000		**0.0**	0.0

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.

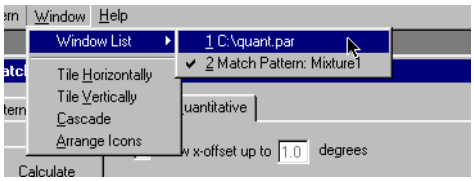


The upper one superimposes a simulated mixture pattern, made up of combining the pure phases in the amounts suggested by the program, on top of the original mixture. The bottom panel is then the difference between the two. This difference plot helps show up either missing phases or extra intensity which 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.

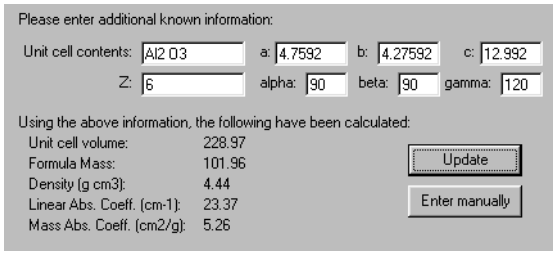
Click *Close* to dismiss the window.

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.



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

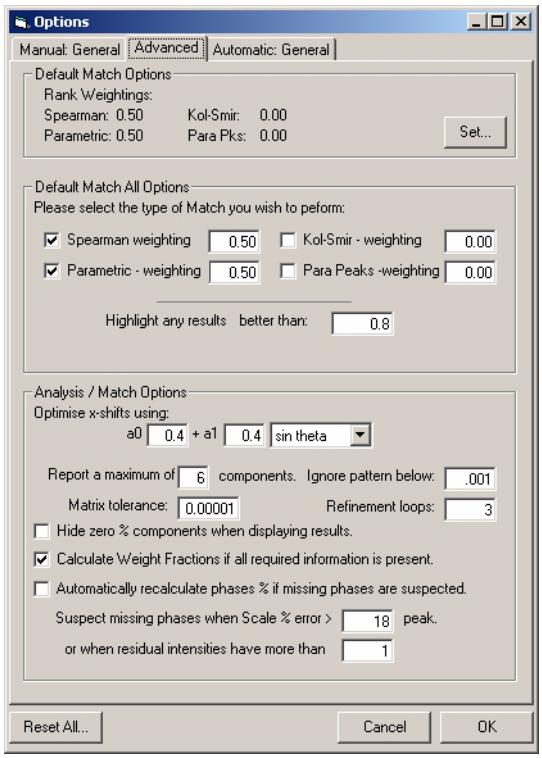
Using the information in the table, enter the relevant details for this phase, and click *Update*. The fields showing molecular weight and absorption coefficients and so on for this phase should update. If not, be sure to enter the formula in the form

<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.

Once all sets of information have been entered and updated, go to the program *Options* dialog box in the *Edit* menu.

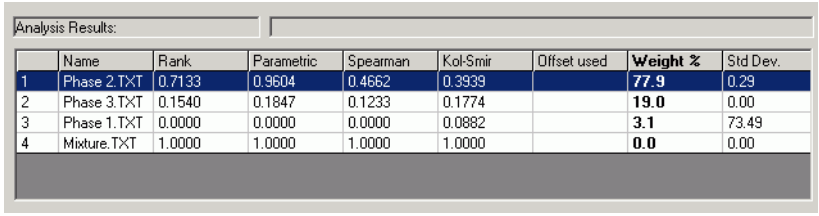


Go to the *Advanced* tab, and make sure that the *Calculate Weight Fractions when all information is present* option is checked. Then close the *Options* dialog by clicking *OK*.

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 %**	Std Dev.
1	Phase 2.TXT	0.7133	0.9604	0.4662	0.3939		**77.9**	0.29
2	Phase 3.TXT	0.1540	0.1847	0.1233	0.1774		**19.0**	0.00
3	Phase 1.TXT	0.0000	0.0000	0.0000	0.0882		**3.1**	73.49
4	Mixture.TXT	1.0000	1.0000	1.0000	1.0000		**0.0**	0.00

If successful, the weight fractions should be close to:

Phase 1 – 3%

Phase 2 – 78%

Phase 3 – 19%

which are the correct answers.

5.7 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.