Comparison of analysis of Silicone Oil sizing data using a spreadsheet and Van Mierlo Software Consultancy's DAx Data Analysis Software

In ophthalmology, intraocular silicone oil has been suggested for the management of retinal detachment complicated by proliferative vitreoretinopathy (PVR)[1]. For such use, the Molecular Weight distribution of the silicone oil used is important. Components with Mw's smaller than 2400 can lead to complications.

In this article, we will look at the steps needed to perform a molecular weight analysis, using either a spreadsheet program, or Van Mierlo Software Consultancy's DAx Data Analysis software.

[1] Try running a Google search using keywords "ophthalmology silicone oil" to read more about this subject.

The Raw Data

Eight silicone oil samples were analysed using Gel Permeation Chromatography, using a refractometer as detector. The results became available as text files such as the one listed below. The files are essentially tab separated, using a mixture of decimal points and decimal commas.

```
="File Information"    9          18
="Sample Information"  20         27
="Comment"             29         29
="Peak Data"           31         34
="Slice Data"          36         245
="Total Average Molecular Weight"        247        249
="Sectional Average Molecular Weight"    251        252

="File Information"
="Peak Number"      2
="Slice Number"     208
="GPC Method File Name"     "P41102.GMT"
="Class Data File Name"     "IRAM1B.D01"
="LC System Number"         1
="Chromatogram Name"        "IRAM1B.C01"
="Background Chromatogram Name"    ""
="Class Method File Name"   "PATRICIA.MET"
="Date & Time"      "02/11/0518:31:10"

="Sample Information"
="Operator Name"    ""
="Sample Name"      ""
="Sample ID"        ""
="Sample Type"      "U "
="Attenuation"      3
="Flow"    0
="Slice Rate"       .5

="Comment"

="Peak Data"
="Separation"       ="Top Height"       ="Start Time"       ="Top Time"         ="End Time"         ="Start Mol"
          ="Top Mol"          ="End Mol"          ="Start Height"     ="End Height"       ="Internal Standard"
          ="Mn"     ="Mw"     ="Mz"     ="Mz1"    ="Mv"     ="I.Visc"   ="Mw/Mn"              ="Mv/Mn" ="Mz/Mw"
 16       2268      19.933    23.750    29.800    308571    53374     3307      13          1         0
          28678     57259     94763     131166    0         0.00000   1.99660   0.00000   1.65498
 16       2847      33.933    35.666    37.933    494       223       79        -4          8         0         200
          220       242       265       0         0.00000   1.09964   0.00000   1.09963

="Slice Data"
="Peak Number"      ="Retention Time"   ="Volume"           ="Molecular Weight"             ="Height"
 1        20,00     0,000     299808    27680     << this is line 38 >>
 1        20,06     0,000     290759    33880
<< lines omitted >>
 1        29,73     0,000     3416      820
 1        29,80     0,000     3313      -260
 2        34,00     0,000     480       21800
 2        34,06     0,000     466       71840
<< lines omitted >>
 2        37,86     0,000     81        -28080
 2        37,93     0,000     79        -16100    << this is line 245 >>

="Total Average Molecular Weight"
="Mn"     ="Mw"     ="Mz"     ="Mz1"    ="Mv"     ="I.Visc"
614       38912     94591     131166    0         0.00000

="Sectional Average Molecular Weight"
="Area Number"      ="Mn"     ="Mw"     ="Mz"     ="Mz1"    ="Mv"     ="I.Visc"   ="Mw/Mn"            ="Mv/Mn"
          ="Mz/Mw"  ="Start Time"       ="End Time"         ="Start Mol"        ="End Mol"          ="Area"
          ="Area%"
```

The area of interest start just below “slice data”. There are data columns listing peak number, retention time, volume, molecular weight, and height.

Since we are interested in the molecular weight *distribution*, the height column data should be normalised, and then plotted versus the molecular weight column data.

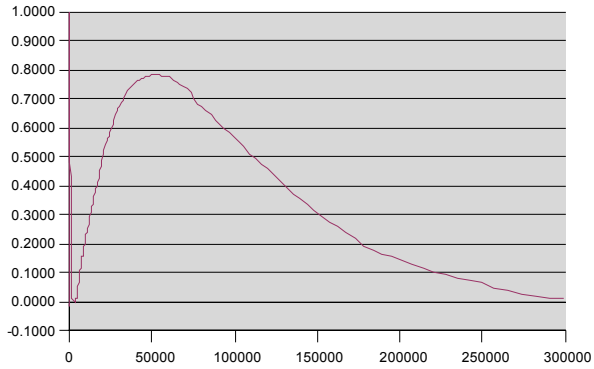
Analysis Using a Spreadsheet Program

The data readily read into Microsoft Excel™ or OpenOffice Calc, using the CSV text import[2].

At this point, it is important to remember that we are interested in normalised data. To this end, a normalised height data column is created. In cell F38, enter the formula **= E38 / MAX(E$38:E$245)**[3]. Do not forget to include the **$** signs to make the cell reference absolute. Now copy cell F38 to cells F39 .. F245.

In order to plot the data, a range of cells should now be selected; in this case D38-D245 and F38-F245. Next, use the spreadsheet's Chart command to start plotting the data. Select *XY Scatter plot* or *XY Chart* as the desired chart type, and *Lines only* or *Scatter with data points connected by smoothed Lines without markers* as the sub type. You may be asked if the X axis values may be sorted; agree to this.

The resulting chart looks something like this.



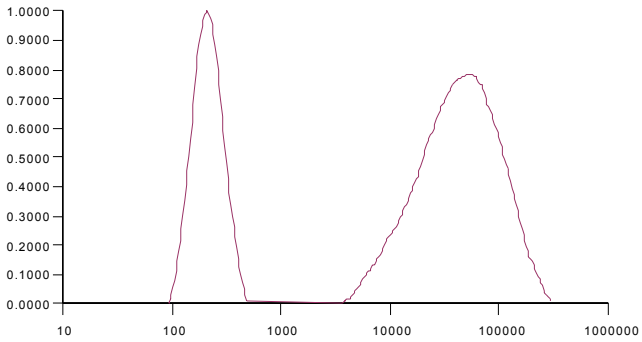
To refine:

- set the lower limit of the y scale to 0
- select a logarithmic scale for the X axis
- change the chart's background colour to *none*
- make the grid invisible

[2] The data are not actually comma separated, as CSV (comma separated values) would suggest. Separation with tabs is not an impediment for OpenOffice. Use of the CSV text import is necessary to force Open Office to open the data as a spreadsheet rather than a text document.

[3] Include **100*** to plot percentages instead of fractions.

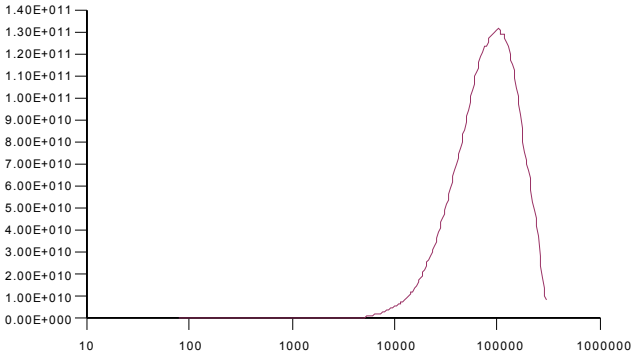
Now the plot looks like this.



Concentration versus Weight Plots

In molecular weight analyses, it is important to distinguish between two kinds of plots: concentration and weight plots.
In this example, the detector used was a refractometer. This typically measures concentrations, not weights. This means that the plot that was created above is a concentration plot.
In order to create a weight plot, the height values need to be multiplied by the molecular weight values. This is easily achieved as follows: enter the formula **= D38 * E38** in cell G38. Now copy the cell to cells G39..G245. Since the data have to be normalised, enter the formula **= G38 / MAX(G$38:G$245)** in cell H38, and copy it to H39..H245.
To change the plot, right click on it, and select the *modify data range* menu option. Change the Range to
$Sheet1.$D$38:$D$245;$Sheet1.$G$38:$G$245.

The plot changes to this:



Of course, you could also have created an additional chart[4], and changed the source data in it.

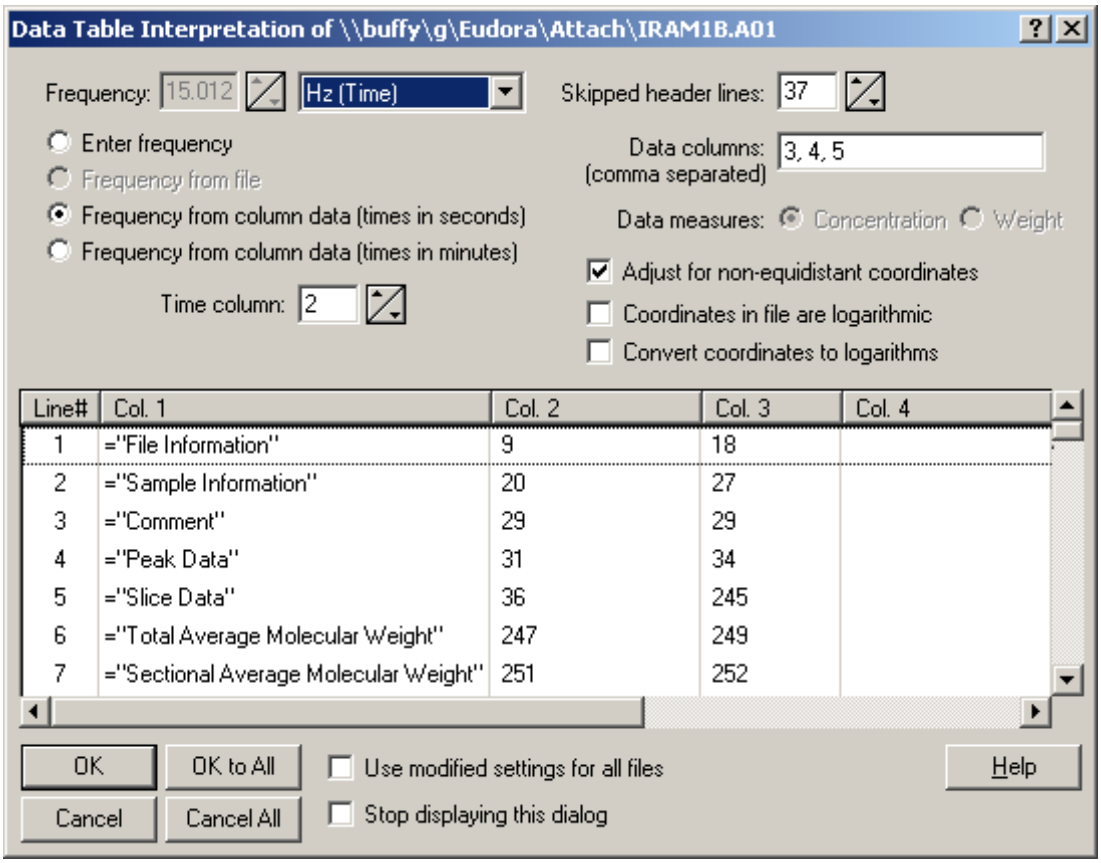
Eight data series

In the case study, eight data files for different silicone oil samples were measured. The procedure described above has to be repeated for each of these samples. It is possible to create a template to speed up this process (slightly hampered by the fact that the data files contain different numbers of data points). Finally, the data in all individual files can be combined to create a single overview plot.

[4] The existing chart can be copied by pressing the Ctrl key and dragging the chart to a new location.

Analysis using DAx Data Analysis software

In DAx, invoke the **File | Open** menu option. Under **List files of type**, select *Generic ASCII*. Select all eight data files, then click the **Open** button.
Next, the **Data Table Interpretation** dialog box is displayed.

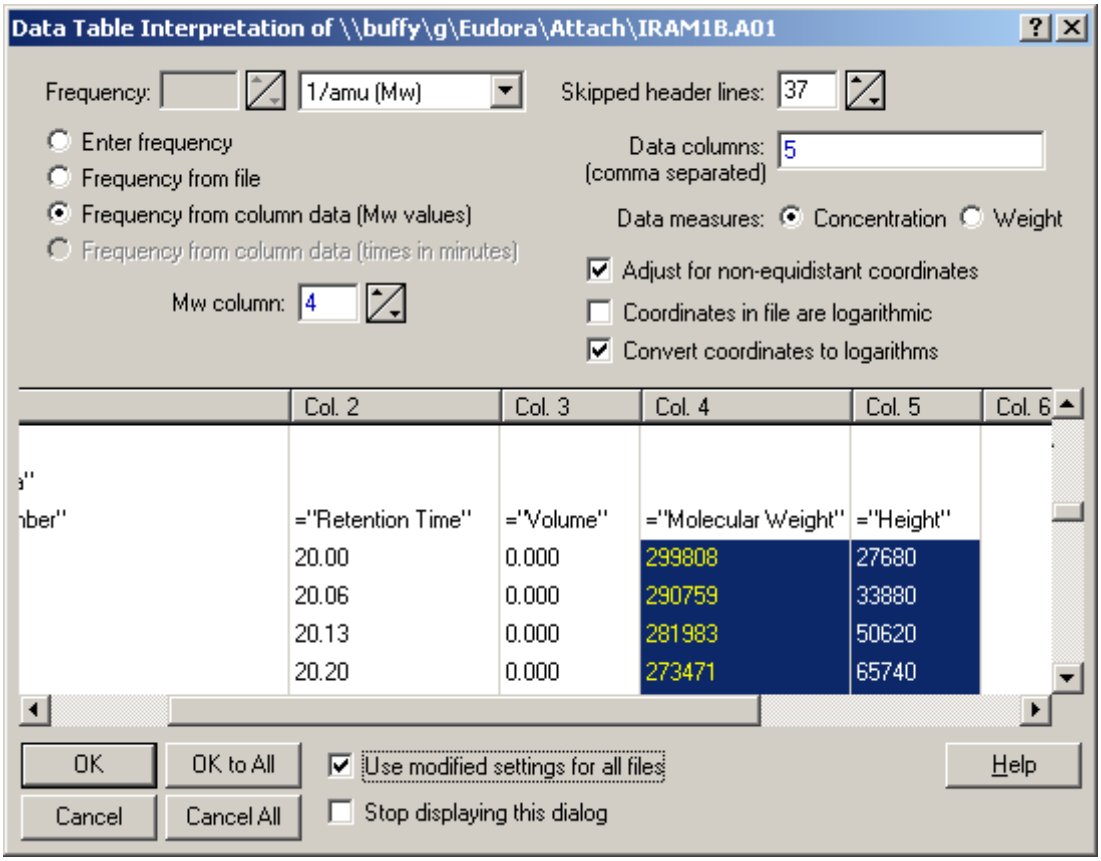


A number of items need to be adjusted:

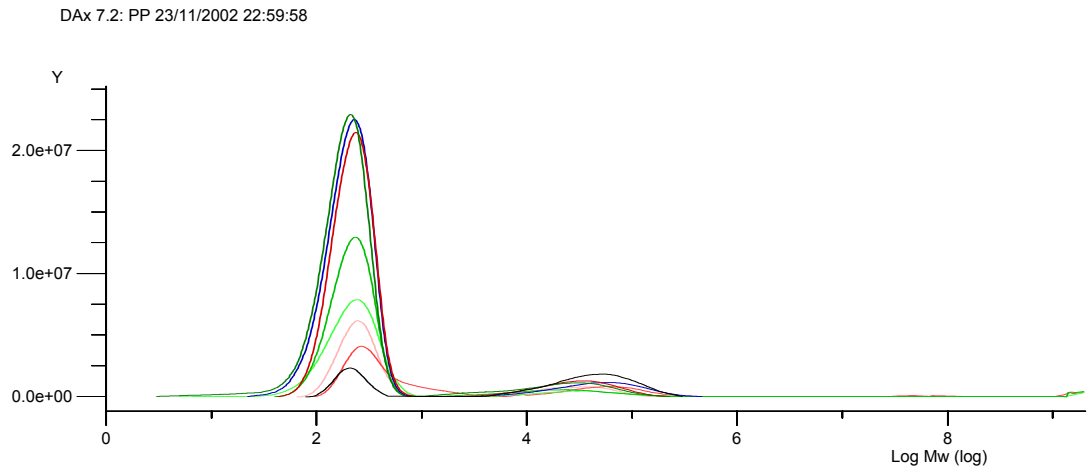
- change the *Frequency* type to **1/amu (Mw)**. This indicates to DAx that molecular weight data are going to be read.
- scroll down the data display to line 38. You can readily see that molecular weights are in column 4, and height data are in column 5. Change the *Mw column* item to **4**, and change the *Data columns* item to list only **5**.
- since these data were measured with a concentration detector, leave *Data measures* set to **Abundance**[5].
- check the *convert coordinates to logarithms* box. This is not a strictly necessary step, but it ensures a better distribution of data points.
- since all eight files are similar, check the *use modified settings for all files* box. This will make all *modified* items (frequency type, time column, data column) be used for all files. Other items, such as the number of skipped header lines, are still determined individually for each file.

The dialog box ends up looking like this.

[5] The *data measures* item can only be changed after the frequency type has been set to 1/amu (Mw)

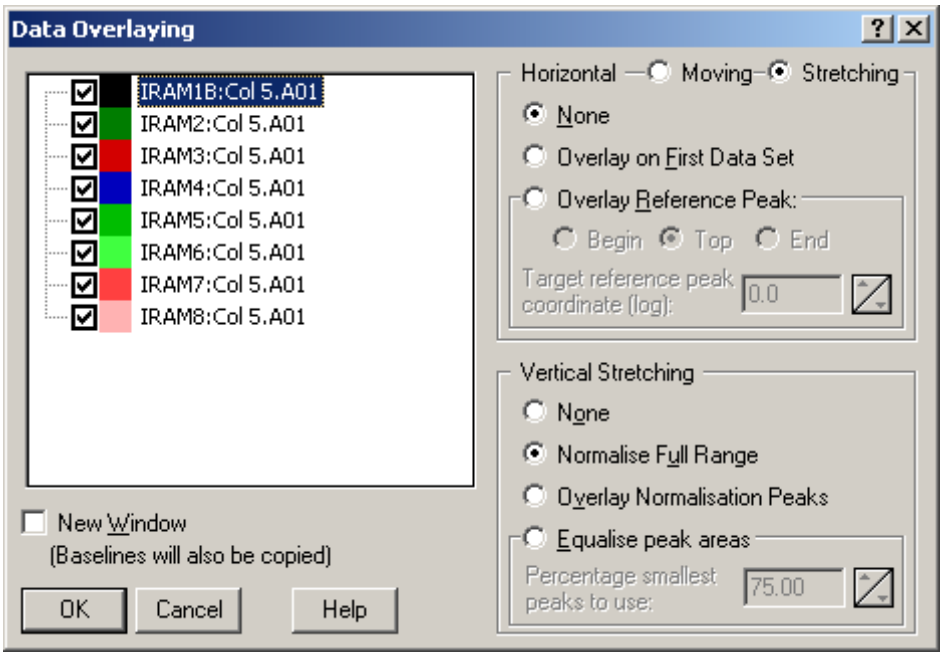


Now click the *OK to All* button. DAx now loads all 8 data files, creating this plot:



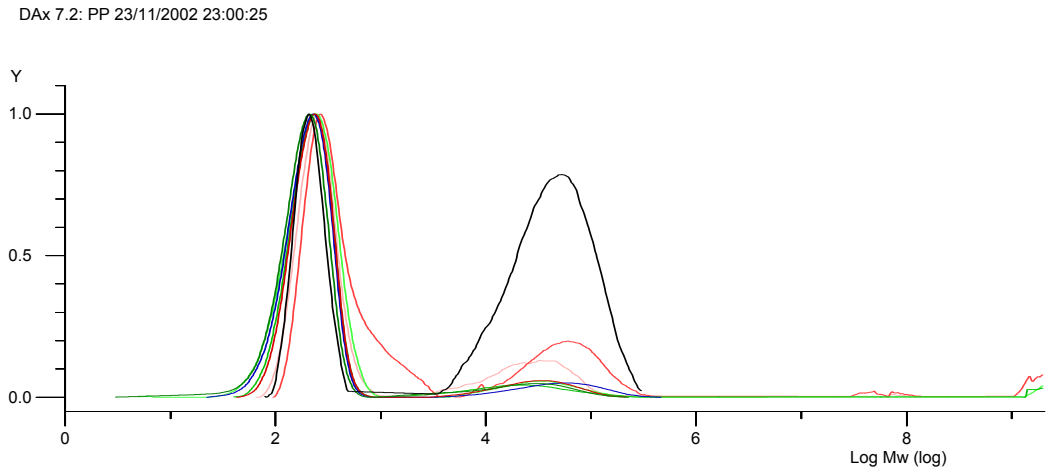
Normalising Data

In order to normalise the data, invoke the **Data | Overlay** menu option. Fill the dialog box in as follows:



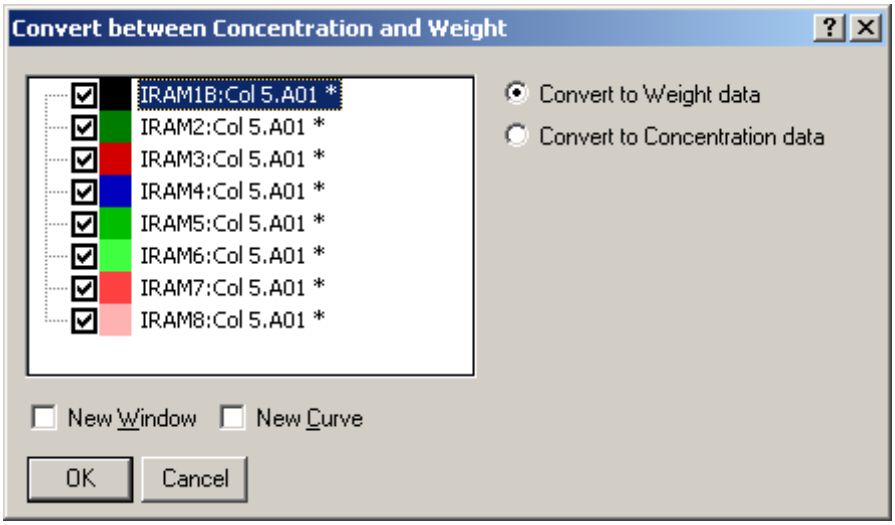
- *Horizontal stretching* is set to **None**
- *Vertical stretching* is set to **Normalise Full Range**

Click the OK button to get this plot:

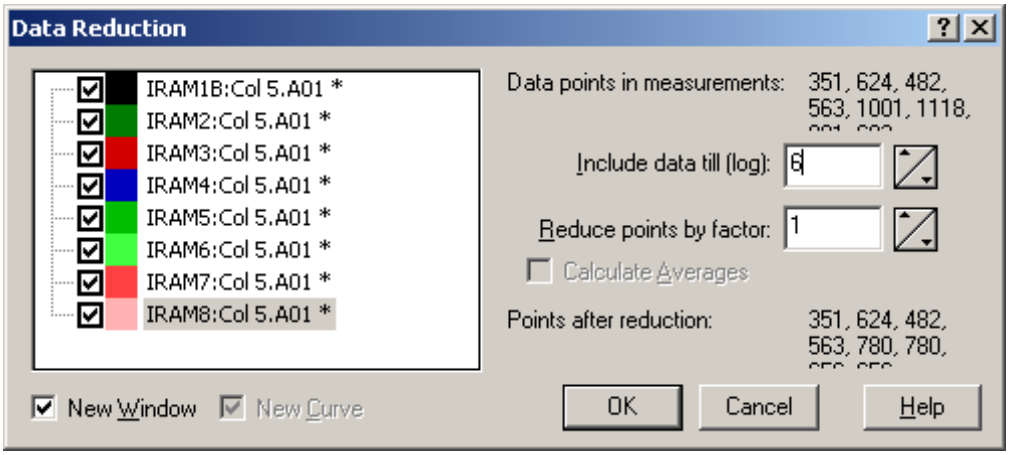


Converting to Weight Plots

Until now, the plots have displayed concentration (or abundance). In order to convert to weight plots, invoke the **GPC | Concentration / Weight Conversion** menu option. This displays a dialog box that looks like this:



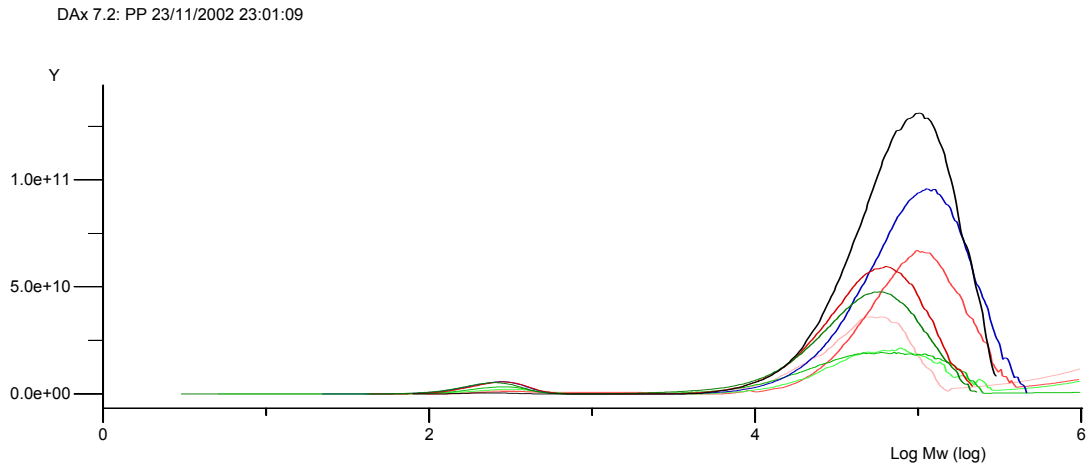
Check the *New window* item, then use the OK button to convert the data to weights. A second window will be created.
A problem becomes apparent when this is done: some of the data contain spurious values at high Mw's, typically above log Mw = 6[6].
To correct for this, invoke the **Data | Reduce** menu option. This displays the data reduction dialog box. In it, and set the *include data till (log)* item to 6. Check the *New window* item[7], and place checkmarks in front of each of the data sets. The dialog box now looks like this:



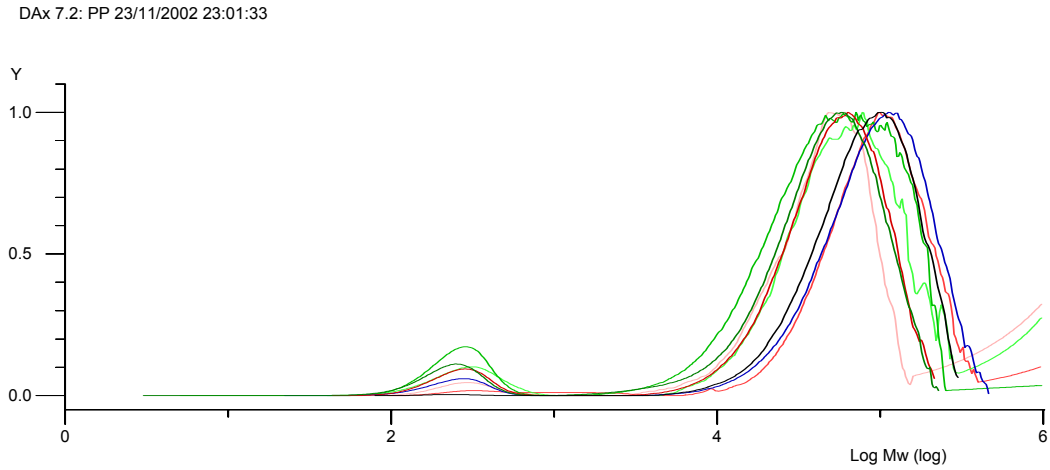
Click the OK button.

[6] In a spreadsheet, this leads to basically unusable plots.
[7] *New Window* must be checked, because the data are considered "raw data". In accordance with Good Laboratory Practise, raw data cannot be modified, but must be *copied*, in this case to a new window.

Now the data look like this:



Use the same overlay procedure that was used on the concentration data, to get this final plot:



Mw Axis plots

It is possible to display a secondary view of GPC data, using the **GPC | Mw Axis** menu option. To try this, activate the window containing the **concentration** data, and invoke the **GPC | Mw Axis** option. Under the **View** menu, make sure that:

- **Weight** is checked (and **Abundance** unchecked)

The following are the default settings:

- **Logarithmic** is checked (and **Linear** unchecked)
- **Differential** is checked
- **Subtract baseline** and **Normalise Peaks** are unchecked. **NB** The normalise peaks option here normalises the total peak area in the data. Since no peaks have been marked in the data, this option would not work.

As mentioned before, some of the data turn out to have spurious values at high Mw's, distorting the scale of the plot. After modifying the scale manually (using **View | Edit Scale**), the plot looks like this.

This option is quite useful for to create various types of plots quickly. However, it does not have the option of normalising the data on range.

Conclusion

It is certainly possible to view the silicone oil molecular weight data using a spreadsheet, even if some of the finer points will be awkward to visualise.

However, the process is far more laborious than using Dax, in part because in the spreadhsheet the entire visualisation process needs to be repeated for each data file[8]. Typically, the analysis of all eight data files would take up to two hours in a spreadsheet, compared to about ten minutes in DAx.

[8] Of course, it would be possible to create a spreadsheet template. However, since the eight data files in this study were subtly different, both in the start of the slice data, and the number of rows of slice data, this would not be straightforward.