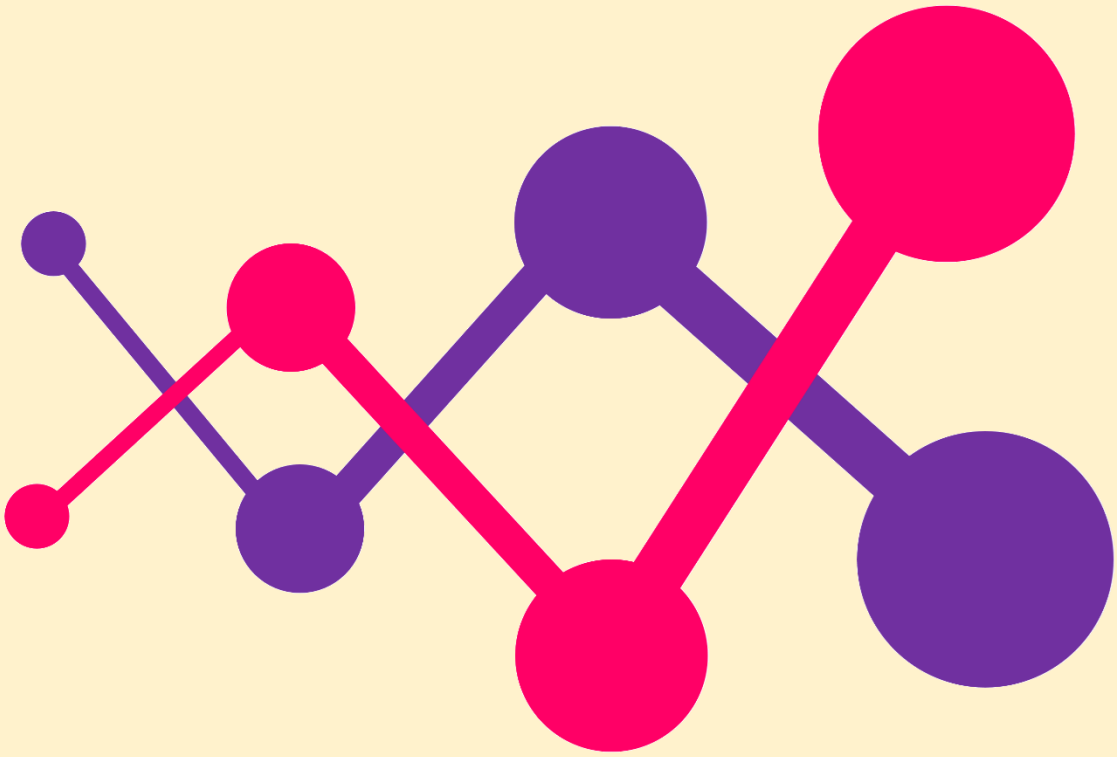
DataPro user manual



DP

DataPro©

Version: 1.0 beta.

Original idea and developed by: Robert Oliva Vidal

Table of contents

What is DataPro?

Description

DataPro is a software designed to perform simple operations on multiple numerical data files simultaneously. DataPro operates with data files containing columns of numbers, where the first column is interpreted as an abscissa (horizontal, x column) and the following columns are interpreted as ordinate (vertical, y column). The operations that can be performed on these columns include; *i*) arithmetic operations such as adding columns, or adding constant values to certain columns, *ii*) format operations such as replacing certain strings, or *iii*) numerical operations such as performing the Fourier transform. In total, the current version of DataPro can perform over 40 different operations.

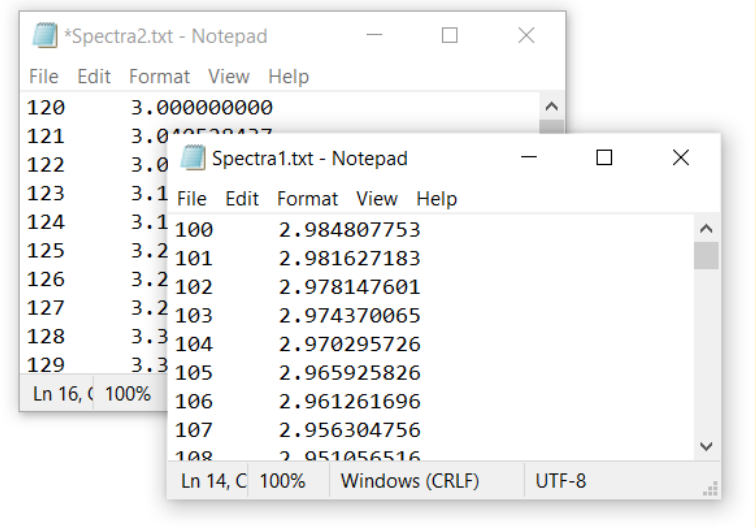


Fig. 1.

Example of two data files that are suitable to be operated with DataPro. Each data file contains two columns, Col. 1 and Col. 2, that will be plotted by an x VS y graph if opened by DataPro.

Contact

The current version of DataPro is at an early stage of development (beta). The developer will keep improving the present program if users find it useful. Please, feel free to ask for user-tailored improvements, report bugs and/or provide comments to the developer. You can contact him via facebook, his user name is "Robert Oliva Vidal".

First steps… File and graph options: opening, viewing and saving plots

First steps

Congratulations! You just downloaded DataPro and are willing to test its features. The present user manual will provide an explanation for each operation that the program can perform (these are named "Actions"). Moreover, the present guide presents examples to better illustrate the working mode of each Action. You can test these examples by following the present guide and operating with the example data files located in the folder of the program, named "0_Guide_tests". If you want to discover the capabilities of DataPro open the program by double clicking on its icon, no installation is required!

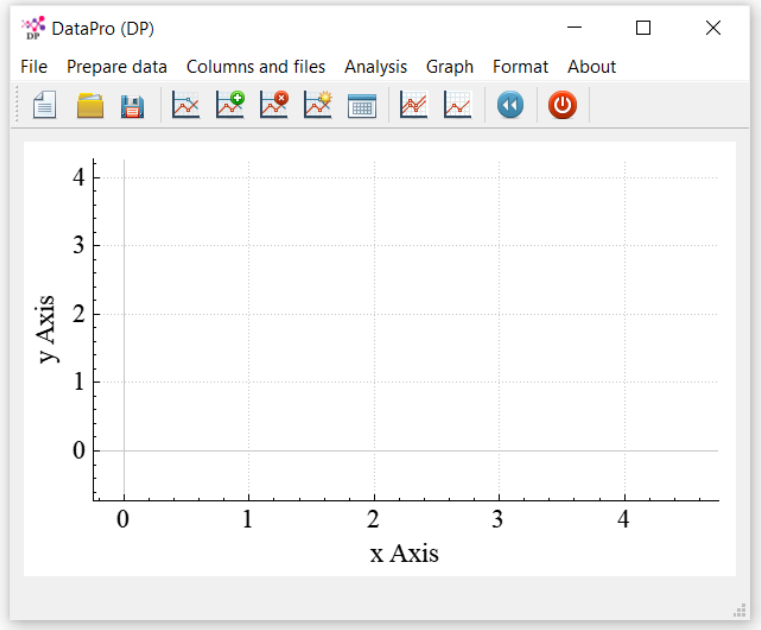


Fig. 2.

Once you open DataPro you will observe a "File Menu bar", a "Toolbar", a "plotting area" and a "Status bar" (bottom, in grey).

Opening and viewing data files

You can open multiple data files containing multiple columns each. DataPro will only accept data files whose numbers use dots for decimals, and have no other characters such as colons, semicolons or commas. The columns can be separated by one or many empty spaces or tabs. If your data files exhibit a different format, you can modify them in "Format" actions (see sections below). For the current manual, you can open all the example data files in the "0_Guide_tests/1_Open_XRD" folder. You can do so from the "File" menu and clicking Open or simply using the hotkey "Ctrl+O" (you can also drag and drop files with your mouse). By default, you will only observe the first opened file in your plotting area.

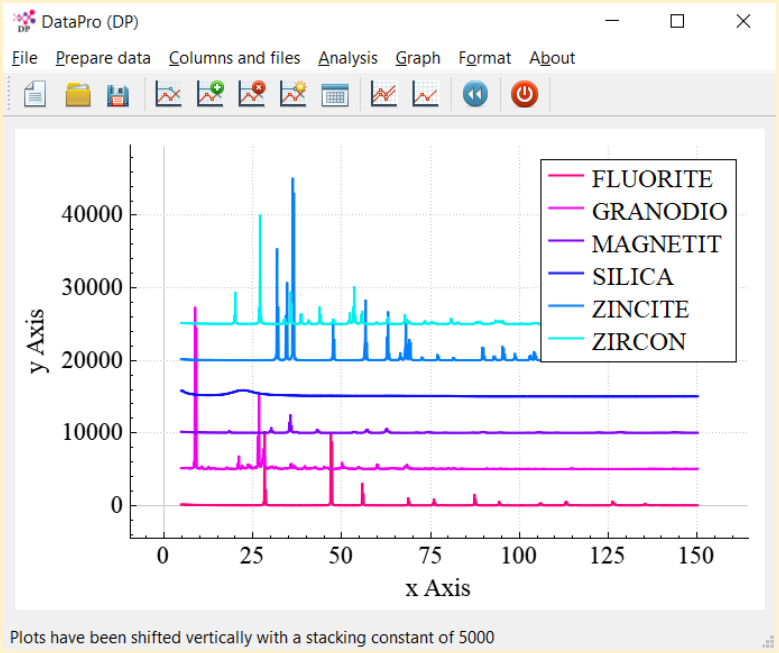


Fig. 3.

Vertically stacked diffraction patterns of different minerals.

Now is a good moment to make a backup by clicking "File->Make backup" or pressing "Ctrl+B". A folder "Guide_tests/1_Open_XRD/backup" will be created with a copy of all opened files. Now that we have made a backup, we can proceed by plotting all plots, one for each file by clicking "Graph->Plot…->All Data" (alternatively we can click the graph with a green symbol on the toolbar, or simply use the hotkey Ctrl+0). Next, you can stack them vertically for a better view selecting "Graph->Vertical stacking->Stacking activated" (also Ctrl+V) and then introducing the value "5000". Now you can observe diffraction patterns corresponding to different minerals vertically stacked. The name of the data files (i.e. name of the minerals) is shown in the "legend" at the top right corner in alphabetical order. The legend can be hidden or made visible by using the hotkeys "Ctrl+Down" or "Ctrl+Up", respectively. You should now see the patterns shown in Fig. 3. Note that after performing the stacking Action, a message will show up momentarily in the "status bar", as it can be seen in Fig. 3.

Sometimes performing operations in multiple data files can be computationally demanding, especially if we are visualizing all the plots simultaneously. For this reason, DataPro allows to perform operations to all files while only observing the changes in a few data files. The maximum number of plotted data files can be modified in "Graph->Plot…-> Set max…". Here we will simply select "Graph->Plot…-> Only first data file" option (Ctrl+1). Next, we will select "Graph->Vertical scale…-> Logarithmic". Now we can see the Fluorite diffraction pattern in a log scale. Note that you can use the mouse scroll wheel to zoom in and out in the plotting area. Moreover, by moving the wheel on each axis you can zoom in the axis individually. Finally, note that the plots can be dragged by holding and moving the mouse with the left-button on the plotting area. You can rescale the whole plot by selecting "Graph->Rescale graph" or pressing "Ctrl+R".

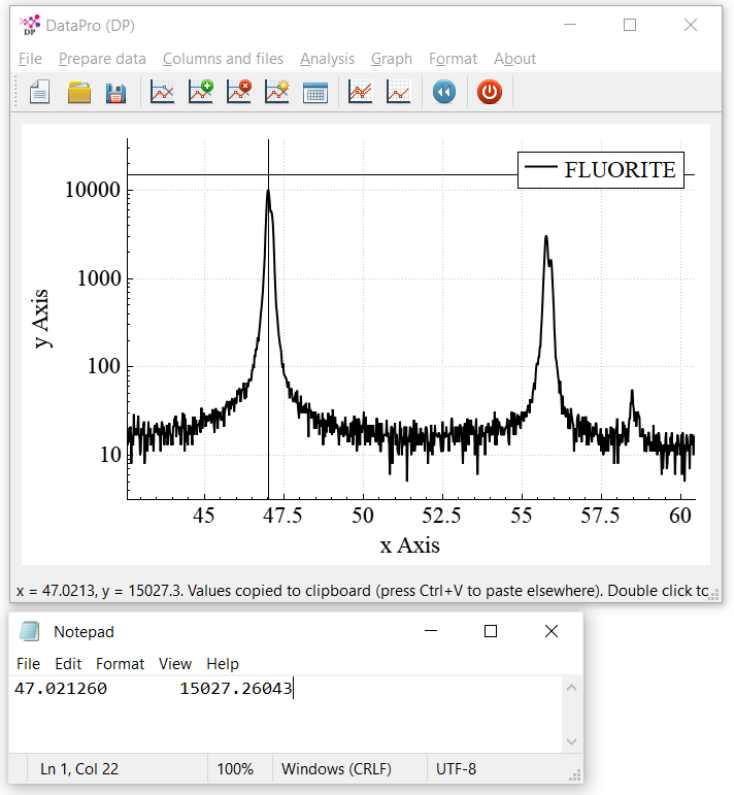


Fig. 4.

Zoomed-in diffraction pattern of Fluorite. The position of a peak has been pasted in a text editor without the need to manually introduce the values.

More interestingly, after left-clicking in any point of the plotting area, the X and Y coordinates will be saved in your clipboard, and you can paste them (Ctrl+V) anywhere else such as a text editor or spreadsheet like notepad or Excel. You can accurately pin point certain coordinates by left double clicking. This can be seen in Fig. 4.

Saving our changes

Following the previous example, we can set the linear scale, "Graph->Vertical scale…-> Linear", plot all data "Graph->Plot…-> All data" (or Ctrl+0), and then save all changes in a new folder. To do so, we click "File-> Save in…", a "Set directory" folder manager will prompt out, there we can open the folder "Guide_tests/1_Open_XRD" and create a new folder by right-clicking, we can name it "stacked". Once the folder manager is set to the directory "Guide_tests/1_Open_XRD/stacked" we click "Select folder" to safe all changes in the newly created folder. Now, if you open again the saved files, you will observe that they are already stacked, in contrast with the original files.

Let's assume you want to modify your data files once more. This time, we want to trim our data from 20 to 100 in the x-Axis. This is done by clicking "Prepare Data -> Trim" and

introducing those values. Once we trimmed our spectra, we can observe that the plots start at 20 and end at 100 units of the X-axis. We can undo this by clicking "Graph->Undo" or simply pressing "Ctrl+Z". Now we recovered the original plot range. Up to 5 Actions can be undone (this limit is set to prevent the RAM memory from being saturated when working with large data files). We can now Trim again from 20 to 100 and save "Ctrl+S". A pop-up message will warn us that we will overwrite our previously saved data. We click "Yes" to confirm. Finally, we can check the "Save As…" option (Ctrl+A), we navigate into the Guide_tests/1_Open_XRD/stacked folder and introduce the name "Diffraction_pattern". We will see that after saving, a copy of our data will be saved with file names stored in a numerical order ("diffraction_pattern_1.txt"). Hence, the "Save As…" option is useful when we want to numerically organize a bunch of data files with large file names. Note that DataPro can open all sort of text files such as .txt, .dat or .asc files. Finally, we can exit DataPro from "File->Exit" or pressing Ctrl+Q. You just finished the first Chapter!

Format options: correcting the format of the data files

Frequently scientists and technicians work with data files stored in different formats, which are difficult to read by different data-manipulation software such as Origin, Excel, Matlab… In this regard, DataPro allows you to easily modify the format of all your data files simultaneously. To do so, you only need to import the data, "File->Import" (Ctrl+I), modify it with the "Format" options explained in the present chapter, and save your changes, "File->Export" (Ctrl+E). Let's see how in more detail!

Rename filenames and extensions

You can import (Ctrl+I) some "damaged" files located in the "\Guide_tests\2_Format_tests" folder. Select and import all the files from the folder, whose names are similar to "PR_1_30s_160mW_g1200_szcz2200" (a temporary message will show up in the status bar). Throughout this chapter, we will improve their format so that these can be opened by DataPro or any other software operating with data files. Once imported, we will make a backup of the data files before manipulating their format. Keep in mind that before manipulating your data files with DataPro, it is always advised to make a backup of your original files. To do so, you just need to click "File->Make backup" or press Ctrl+B. A folder named "\Guide_tests\2_Format_tests\backup" will be created and all files are copied inside.

Now that we made a backup, we can proceed to simplify the names of the files, to do so we select "Format->Rename->File names". A message box will appear with different options, this time we will use the option "Replace a string". We will introduce "_30s_160mW_g1200_szcz2200" which will be replaced by nothing "" in order to delete that part of the file names. We repeat the operation but this time we introduce the string "_035s_g1200_szcz2200" to be replaced by nothing "" in order to delete it. We can see that now the names of the files are much simpler, such as "PR_1.xxx". Next we can proceed modifying the extension of the data files so that these can be visualized in a text editor such as notepad. Select "Format->Rename->Extensions", a question box will pop up, select "Yes", then introduce the three letters: *txt*. Now a temporary folder has been created in "\Guide_tests\2_Format_tests\TempPreview", open it in your folder explorer, and check your newly created data files with .txt extensions instead of .xxx. Don't open them since they are only temporary files. Now we can export the changes by pressing Ctrl+E. It is crucial that before exporting your changes, you make sure all files inside the \TempPreview are closed (not used by any other software such as notepad) since the temp folder and its contents will be deleted after exporting. Finally, our data files should look like in the bottom panel of Fig. 5.

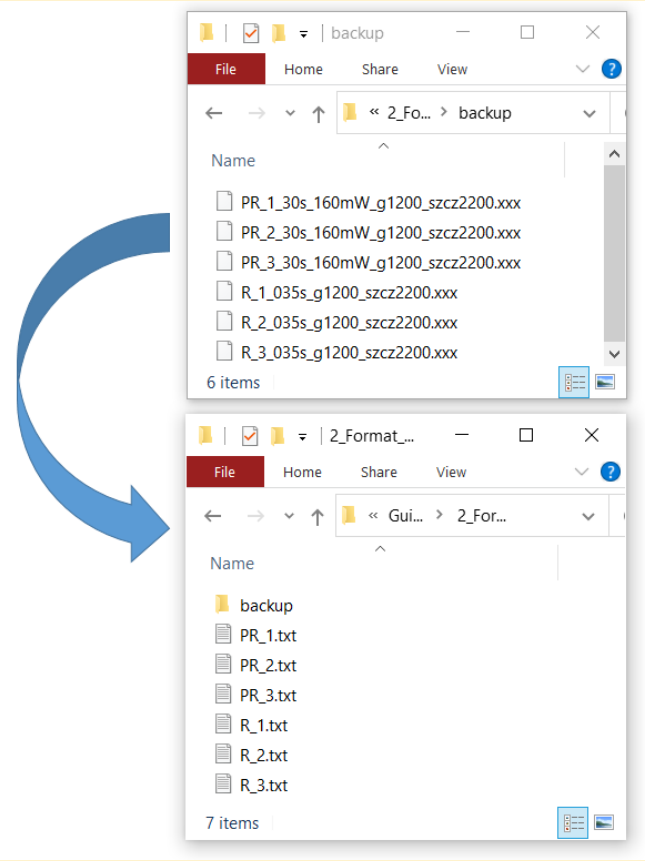


Fig. 5.

The file names and extensions can be modified with DataPro.

Formatting file contents: Replace strings, restore damaged data, and more!

Following the previous example, we import (Ctrl+I) all the data files located in "\Guide_tests\2_Format_tests" with names similar to "PR_1.txt". Before operating them, we can inspect them with a text editor such as notepad. Let's open the file "PR_1.txt" with a text editor. We will find many format errors inside. For instance, the file has some caption lines that we will want to delete. Also, the numbers have commas in their thousands. Another format error that we will correct are the presence of semicolons between columns, and some illegible numbers such as "ERROR", "Inf" or "0/0". Finally, the file contains some empty lines that need to be deleted. It can also be noted that the first column exhibits descending order instead of ascending. Next we will correct all these errors with DataPro.

First we will delete the semicolons. Before modifying the data files, it is very important to make sure that the data files are not being used by any other software such as the notepad. We select "Format->Format datafiles->Find and replace characters" or press Ctrl+F. We introduce a semicolon ";" and replace it by nothing in order to delete it "". Once finished, the status bar (below) will temporarily report the number of replacements made, in this case more than 10 000.

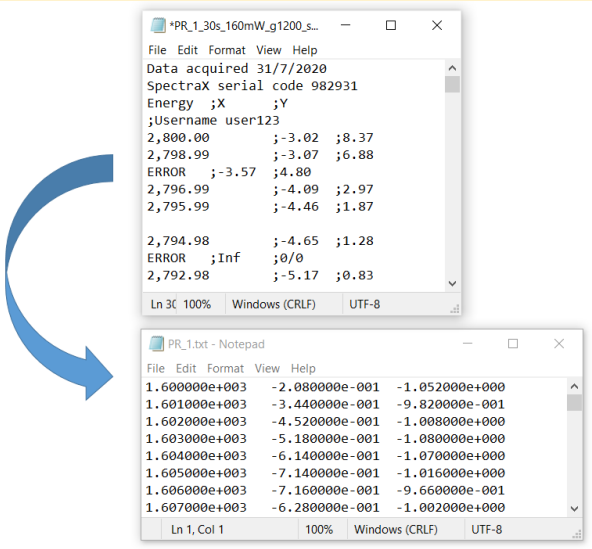


Fig. 6.

DataPro allows you to restore damaged data files, and set a proper multicolumn format.

The results can be previewed in the \TempPreview folder. We now close all files observed by the notepad and proceed deleting the thousand commas by repeating the operation, substituting commas "," by nothing "". Optionally we can substitute tabs by spaces by copy-pasting (from the data files) a tab " " which will be replaced by a single space " ". Next, we can delete all rows that contain captions or errors by selecting "Format->Format datafiles->Delete lines without data", we introduce "0" caption lines to be respected. We can now export our changes (Ctrl+E) and then check the changes from the notepad, now the data looks nicer, but we can notice that the step value (i.e. the value by which numbers increase or decrease) of the Column 1 (x) is not constant because some error lines have been deleted. We also notice that the step of Column X for files "PR_x.txt" are approximately 1, while for files "R_x.txt" are 2. If we want to divide both data files, we will need them to have the same values and step in their Column 1 (x). We can fix all these problems by importing all files (Ctrl+I) and clicking "Format->Modify datapoints->Set step of Col 1". A warning message will pop up, we click "Yes", we then introduce the step number "1" and click OK. In the \TempPreview folder we can see that all data is shown in scientific format, and the step of all columns is now constant and "1". The values of newly created rows are interpolated from adjacent values. We can export changes Ctrl+E.

Finally, we can flip the order of all columns in order to obtain an ascending order of the first column (X column), to do so we import all data files (Ctrl+I) and click "Format->Format datafiles->Data ordering…-> Ascending", we introduce "0" caption lines to be respected and export changes Ctrl+E.

Prepare data

Options "prepare data" in the Menu Bar are meant to be used to make a first preparation of our data files. To illustrate this, open (Ctrl+O) the spectroscopic files located in the folder "Guide_tests\3_Prepare_Data", then plot both spectra by clicking in "Graph->Plot…->All data". You will see that the plots exhibit some cosmic spikes, these are undesired signal that we will correct by selecting "Prepare data->clean spikes" (Ctrl+C) and then introducing the parameters 35 and 13. These are tolerance parameters that will help DataPro determine what is, and what is not a spike.

Now we can see that the plots correspond to absorption IR spectra of CO and CO_2, where the x Axis is in nanometer units. We will modify the units by clicking "Prepare data-> x Axis->Energy units", then select "nm" from the list, and then select "cm-1" as desired energy units. Finally, in order to assign the vertical scale in percentage (%) we will select "Analysis->y Axis->Multiply constant" and introduce "100". This will multiply the vertical scale by a factor of 100. You should now observe spectra like the one shown in Fig. 7. You can save your changes to end this chapter (Ctrl+S).

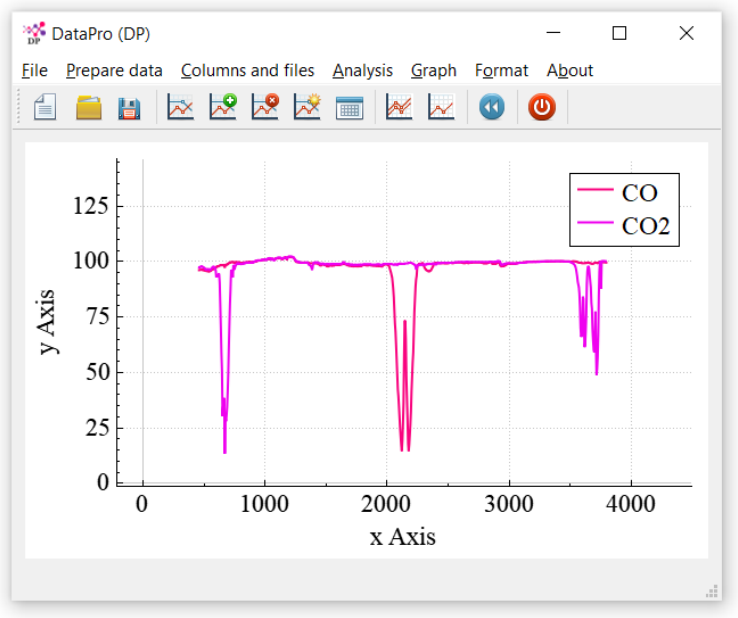


Fig. 7.

DataPro can be used to analyze all kind of spectroscopic data, from Raman to Infrared. This figure shows two absorbance spectra in the IR regime, units in cm^{-1}.

Operate between columns and files

Multiple columns operations

The spirit of DataPro is to facilitate performing simple operations on many data files simultaneously. If the user has all data collected in multiple columns, DataPro can also analyze them, even if there are also multiple files! Before we explore this option, it is worth noting that from "File -> File operations" it is also possible to merge multiple data files into a single multiple-column file, and vice versa.

Hence, DataPro allows to operate multiple files containing multiple columns each. To exemplify this, we can open the file "Multi_1.asc" located in the folder "\Guide_tests\4_MultiColumns". We can now observe a curve with a max around 15 named "Multi_1 (Col. 2)" in the legend. Since DataPro detected that this file exhibits multiple columns, the legend will indicate the particular column that is being plotted. To select a different column we click "Columns and files-> Set 'y' Column data" or simply press Ctrl+Enter. We can now introduce 9 and we will observe that the new curve exhibits a max around 10 and a similar shape as Col. 2. We can plot and operate all columns simultaneously by activating "All Columns mode", we can do so by introducing the number 0 after Ctrl+Enter. We should now observe the curves shown in Fig. 8.

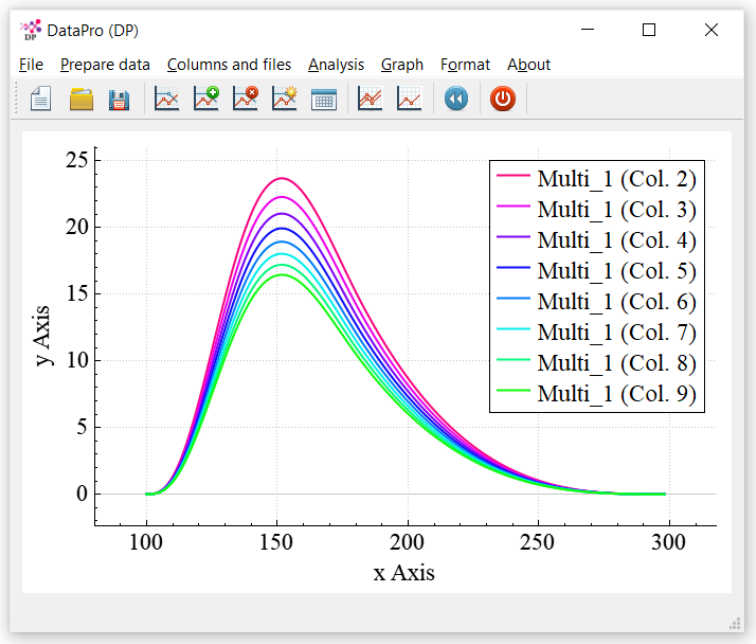


Fig. 8.

DataPro allows to plot and operate data from multiple columns simultaneously.

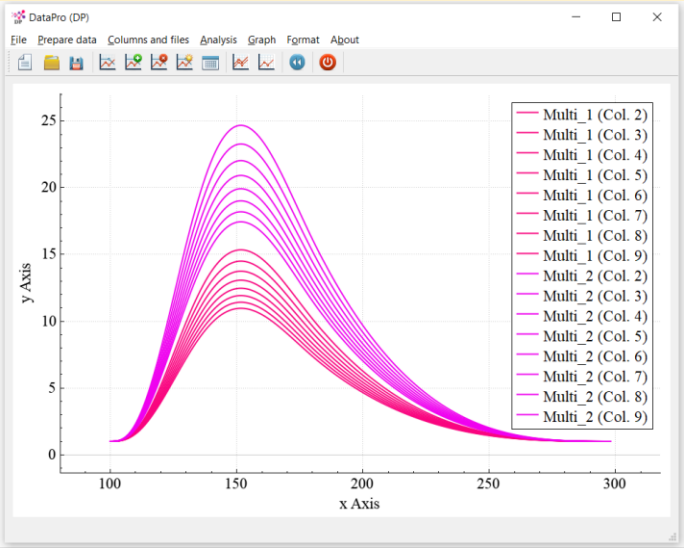


Fig. 9.

DataPro allows to plot and operate data from multiple columns (here columns from 2 to 9 are plotted) and multiple files simultaneously (here two data files are plotted, Multi_1 and Multi_2).

Note that multiple columns of multiple data files can also be plotted and operated simultaneously (see Fig. 9). To do so we will open the files "Multi_1.asc" and "Multi_2.asc", then click "Graph->Plot->All data" (now the plotting area will show the spectra corresponding the first column, Col. 2, of each data file, as indicated in the legend). Finally, to plot all columns we press "Ctrl+Enter" and introduce the value "0", and to plot All data we press "Ctrl+0". We can now see that DataPro is plotting 9 curves for each of the two data files. To illustrate that DataPro can manipulate all columns and files simultaneously, we now select "Analysis->Operations->Differentiate". We can now see that all plots changed the lineshape and become zero at a x Axis value of 150, which is the position at which all original spectra had a maximum and zero slope. If we now integrate the spectra by selecting "Analysis->Operations->Integrate", we will recover very similar curves as in the first case. This is so because integration and differentiation are somewhat opposite operations.

For the sake of simplicity, from now on we will only operate one data file, so we can open only "Multi_1.asc" without saving previous changes. Now we hit "Ctrl+Enter" and introduce 0 to plot all columns simultaneously. First, it is worth noting that we can delete or rearrange all columns in "Columns and files -> Columns" option. If we select "Columns and files -> Columns -> Average all columns", a new column number 10 (green color) will appear in the middle of all columns because it corresponds to the averaged values of all columns. We undo this action (Ctrl+Z) to operate only with the 9 columns. We are interested now in dividing all columns by the smaller column. To do so we select "Columns and files -> Operate with Ref. Column -> Set reference column" and introduce 9. Next we select "Columns and files -> Operate with Ref. Column -> Divide by reference". Since all plots are slightly larger than Col. 9, the resulting curves (and ratios) are slightly larger than 1 (zoom-in the Y axis between 1 to 1.4 to note that). Note that all curves resulting from

the quotient between Col. 9 and the rest columns have been included as new Columns, from 10 to 17.

It is worth noting that DataPro allows to export files with multiple (Y) columns into multiple files containing a Y column each. Following the previous example, we can save our file with 17 Y columns, then, we select "File->File operations…->Export column(s) to files", and introduce 0 to export a single file for each Y column. We can see that 16 new 2-column (X and Y) files have been created, from "Multi_1_Col2.asc" until "Multi_1_Col17.asc".

Operations between multiple files

DataPro not only allows to make operations between columns, but also between different files. Next we will see how. Start opening files "Multi_1.asc" and "Multi_2.asc" and plot all their columns (press "Ctrl+Enter", "0") for all files (click "Graph ->All Data" or press "Ctrl+0"). Then we select "Columns and files -> Operate with ref. file -> Set reference file(s)" and we introduce the file "Reference.asc", the status bar will show a message indicating that the reference has been imported. Finally, we select "Columns and files -> Operate with ref. file(s) -> Divide by ref. files", we then introduce column number 2, because the reference file has multiple columns. We now observe that the lineshape of all plots has been slightly modified, the resulting division has been overwritten in all the columns.

It is worth noting that the user can also operate the opened files with the same number of reference files (i.e. you can use multiple reference files instead of one). In this case files would be operated in a 1-to-1 fashion.

Analysis

DataPro allows to perform a number of operations simultaneously to all our data files and their columns. In the present section a few analysis options are presented. Let's start by opening (Ctrl+O) the file "PR.dat" located in the folder "\0_Guide_tests\5_Analysis". These are photoreflectance measurements of a direct band gap semiconductor (2 eV). This file contains two columns, 2 and 3, which correspond to measurements in the X and Y channels of a lock-in amplifier. Both columns can be plotted simultaneously by pressing "Ctrl+Enter" and introducing number "0". Unfortunately, the spectra present some interferences. The frequency of these interferences is around 71.5 (this can be calculated by taking the difference position of two adjacent interferences, here around 0.0140 eV, and calculating its inverse, 71.5). The frequency has been calculated from the position of the peaks, which can be easily obtained by double clicking with the mouse on the peak maxima, and pasting (Ctrl+V) the results in a spreadsheet like Excel. We can now correct these interferences from Fourier transform analysis. Select "Analysis->Special analysis->Clean interferences by FFT", then introduce the values 71, then 72, since these is the frequency range we want to block. Afterwards it can be seen that the interferences disappeared. The magnitude (R) and phase (θ) are related to the in-phase X and quadrature Y by $R = \sqrt{X^2 + Y^2}$ and $\theta = \arctan(Y/X)$. We can calculate the magnitude and phase by selecting "Analysis-> Special Analysis -> Modulation spectroscopy -> X Y -> R Theta". We introduce now the columns 2 and 3 for X and Y, respectively. It is a good moment to save the files (Ctrl+S). We observe that two new columns have been created, we select the 3rd "y" column, corresponding to R (Ctrl+Enter, option 4). We can now trim our spectra "Prepare data -> Trim" and select 1.8, then 2.79.

We can now subtract a baseline, "Prepare data… -> Baseline -> Linear baseline" which will bring the plot ends to zero. Now, we can calculate the modulus (from the Kramers–Kronig relations) of the spectra by selecting "Analysis->Special Analysis-> Modulation spectroscopy -> Calculate modulus from KKA". Finally, we can smooth our spectra by selecting "Analysis -> Special analysis -> Laplacian smoothing" and introducing the value 50 (larger values will yield larger smoothings). One last operation that we can perform is to apply a ball-like baseline. This is a baseline that is determined by an imaginary rolling ball under the spectra. To do so we introduce "Prepare data -> Baseline… -> Ball-like baseline", and then introduce the desired radii of the ball, here we set 0.2. Hence, bands larger than 0.4 eV will be subtracted while peaks smaller than 0.4 eV will be preserved. Now it can be seen that the spectra exhibit two dominant features, one around 2 eV and another around 2.75 eV. Before saving our changes, click "Columns and files -> Automate Actions -> Generate code" and copy the resulting code in a text file. You will need this code for the next section.

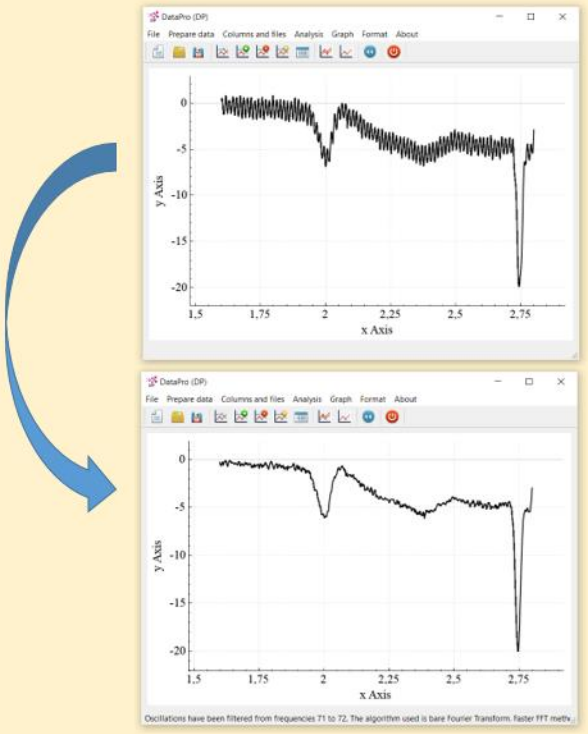


Fig. 10.

Interferences can be filtered by using Fourier transform analysis.

Now it is a good moment to test other features, such as "Prepare data -> Normalize -> Global maximum -> To one", this will set our plot from 0 to 1 in the vertical axis. We could export our spectra from "File -> File operations… -> Export column to single file" and select Col. 4, then introduce a name such as "Col_4_exported". Note that this analysis could have been performed simultaneously to many data files at the same time. To do so, we could have opened all those files at the same time and then set all the above operations while visualizing the first file, or all files simultaneously (the operation will be performed to all data files in both cases). You can discard current plots from "File->New" (Ctrl+N) in order to proceed to the… last Chapter!

Automate Actions

DataPro offers another feature to save the user's time when dealing with multiple data files. This feature, located in "Columns and files -> Automate Actions" allows to generate a text code that sequentially registered all the introduced Actions by the user since the data files were opened or last saved. The purpose of this code, is to allow the user to set a number of Actions in a file, and then be able to repeat them all in similar files by simply introducing the code. Hence, this feature can be useful if the user needs to make identical actions on data located in different folders, or during different days.

Let's exemplify this by opening all files from "PR_1.dat" to "PR_8.dat" located in "\0_Guide_tests\6_Automate actions". We can now plot all files (Ctrl+0) and all columns (Ctrl+Enter, then introduce "0"). We can now see that these files are similar to the ones in from the previous section. Columns 2 and 3 correspond to X and Y channels of photoreflectance spectra, and columns 4 and 5 correspond to R and the phase θ. We plot back only the column 4 (Ctrl+Enter, then introduce "4"), and then we introduce the code (Columns and files -> Automate Actions -> Introduce code) obtained from the previous section (in case you didn't save it, you can also find the code in the folder of the files, inside a file named "code.txt"). We can now plot all our data (Ctrl+0) and stack it "Graph -> Vertical stacking -> Stacking activated" introducing the value "0.4". You should now obtain a plot similar to the one shown in Fig. 11.

Hence, if the user regularly uses DataPro, it is advised that the user creates and saves their own codes to process its own data.

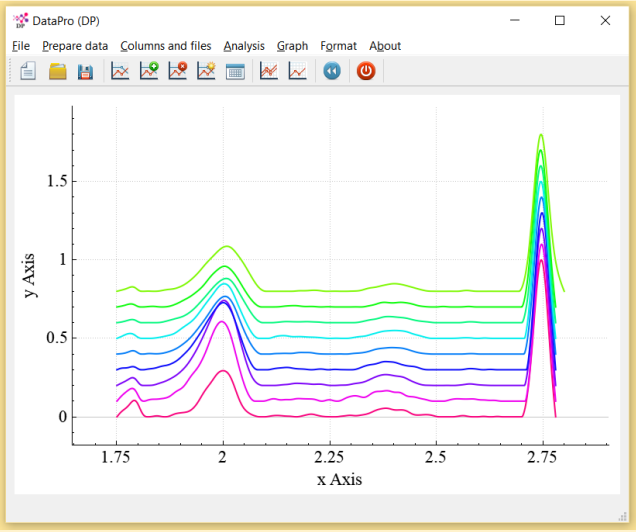


Fig. 11.

DataPro can sequentially repeat a series of operations (Actions) on multiple files.