Yappari tutorial : fitting multiple datasets

Version 22 08 2023, author ND

This is a short tutorial with a lot of images showing a simple way for fitting multiple datasets. Some variants are possible, but let's try to make it simple.

If you have a large number of datasets in a file, you can read them directly. In the /files directory you'll find a file with the name "**multiple_datasets_example.dat**" which has 3174 spectra that I simulated with Yappari and to which I added random noise up to 1% of Z.

This file (obtained directly from Yappari) has a format like this :

```
Freq /Hz, Zr , Zi ; Name: multiple_1237_
1.000000E+6  7.790801E-1  -7.921781E+0
8.317638E+5  1.135679E+0  -9.450379E+0
6.918310E+5  1.605364E+0  -1.131852E+1
5.754399E+5  2.322602E+0  -1.351846E+1
4.786301E+5  3.310995E+0  -1.585968E+1
…
Freq /Hz, Zr , Zi ; Name: multiple_1236_
1.000000E+6  7.823519E-1  -7.937434E+0
8.317638E+5  1.131766E+0  -9.414772E+0
6.918310E+5  1.618530E+0  -1.135973E+1
5.754399E+5  2.315114E+0  -1.352121E+1
4.786301E+5  3.304463E+0  -1.591682E+1
```

Each dataset is separated by `Freq /Hz, Zr , Zi ; Name: multiple_1236_`
The text after `Name:` is a name of a dataset
We can read this file as custom, we'll define the format by :

```
[header]=Freq /Hz, Zr , Zi ; Name:
[label_length]=0
#data_columns=1,2,3
```

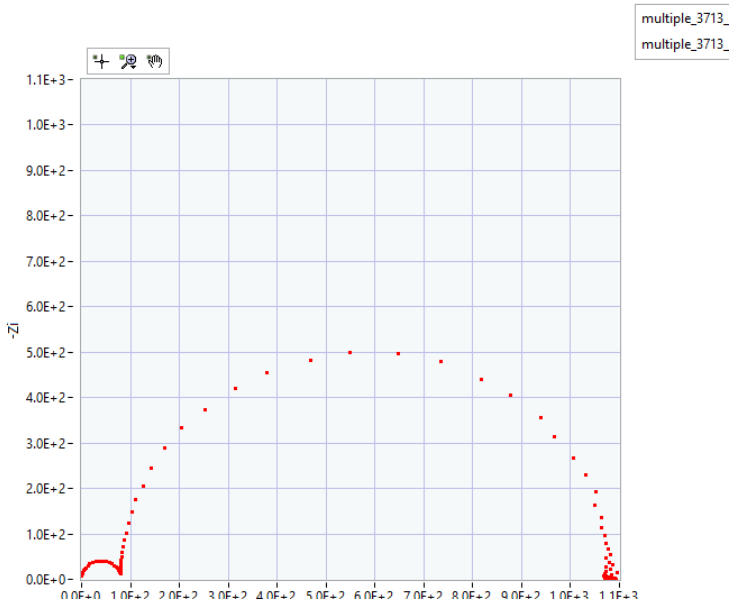
This definition is saved in the file **_definition_custom_multiple_datasets.ini** which is also included in the /files directory. Basically the instructions are to separate the data in pieces that are separated by the header `Freq /Hz, Zr , Zi ; Name:`

We don't care for the label if `label_length=0`, if you do put a number in the `[label_length]=` and the sets will be labelled with the text found after the header (the number of characters to hold as label).

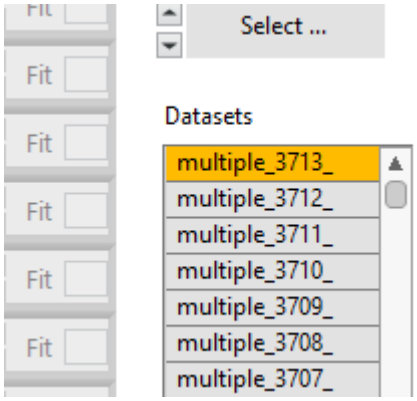
Go to *Read data* and select *Custom*, the program will ask for the name of the datafile, select **multiple_datasets_example.dat**

After that, the program will request a definition for the custom file, select **_definition_custom_multiple_datasets.ini**

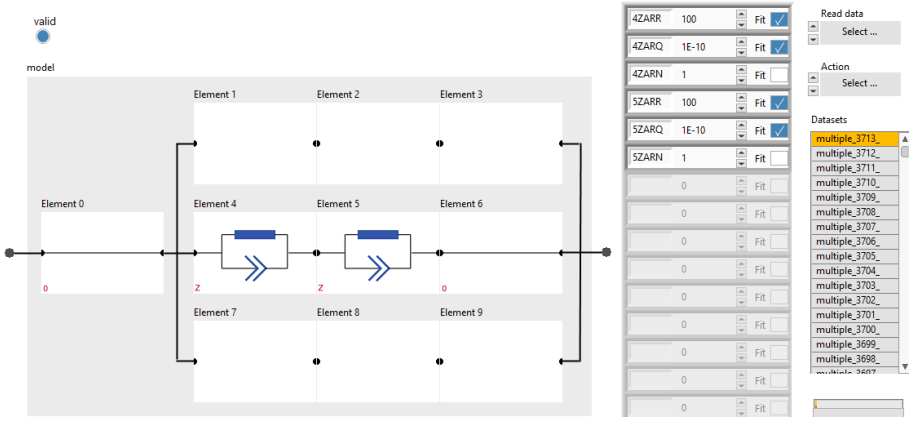
It will take a bit of time, 30 seconds on my computer, and you'll end up with this (if everything else is correct : separator set to TAB, in the Parameters panel in Yappari, the numbers on your PC have the decimal separator as . "dot", and not , "comma"…)



The program selects the first dataset, in this case number 3713, which is the last dataset found in the file (the first one is labelled as 0, is the last in the list). The Nyquist plot shows only the selected dataset. The selected dataset is shown colored in the right part of the program.

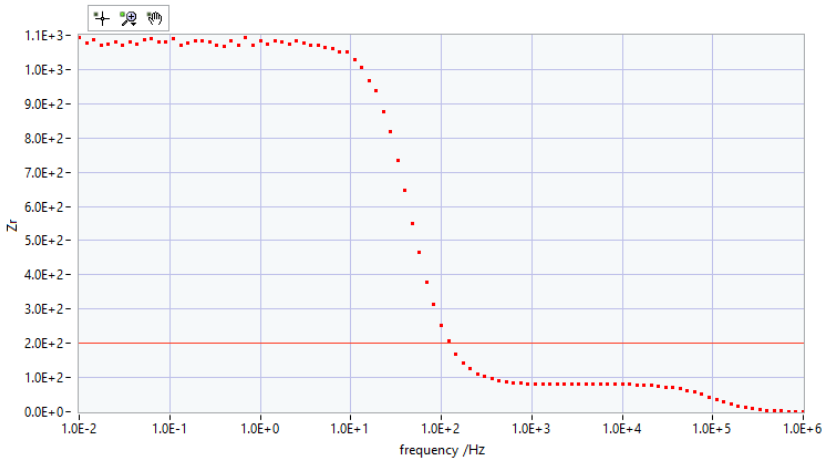


We can inspect the data in other plots, but for fitting the first thing to do is to create a circuit, an equivalent electrical model. To create a circuit you need to go to the **Model** panel

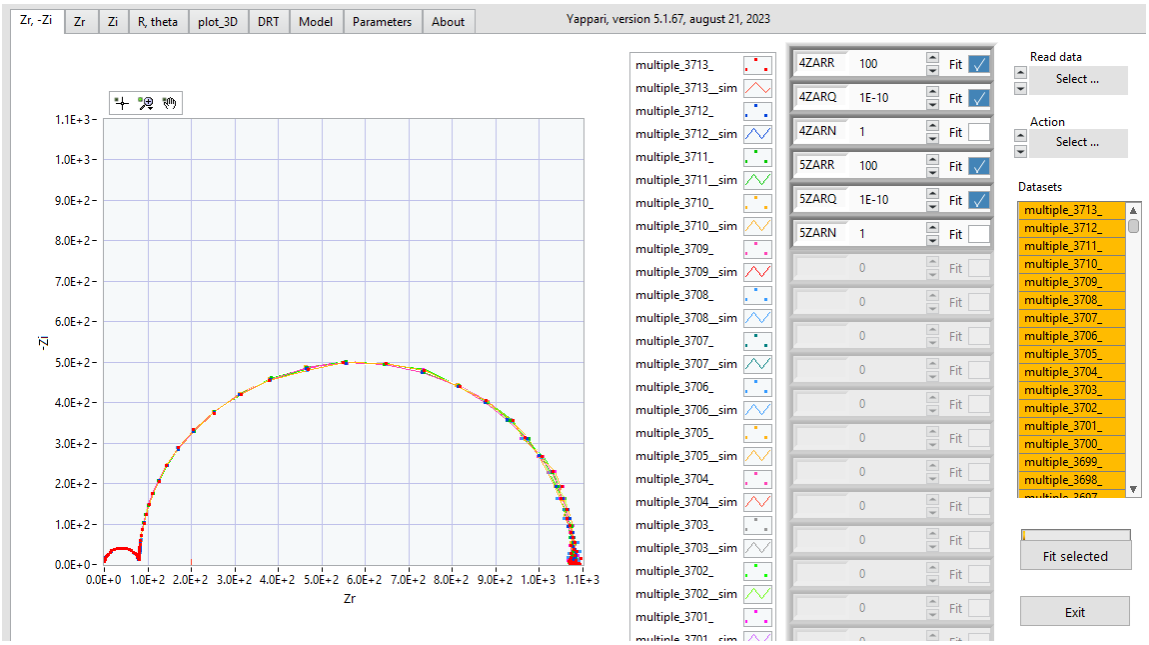


We create a circuit with two Zarcs (e.g. R and Q in parallel). The parameters are showing 4ZARR, 4ZARQ, 4ZARN and 5 ZARR, etc. The numbers 4 and 5 show where the objects are, in this case a ZARC in element 4 and another ZARC in element 5. The other letters are obvious here R, Q and N for a Zarc.

Once a circuit is created the LED valid is ON (see the top left in the previous image) and some dummy values for the elements are generated, a calculated spectrum is made for each active dataset. For this case we have the experimental, dots, and the simulated spectra.

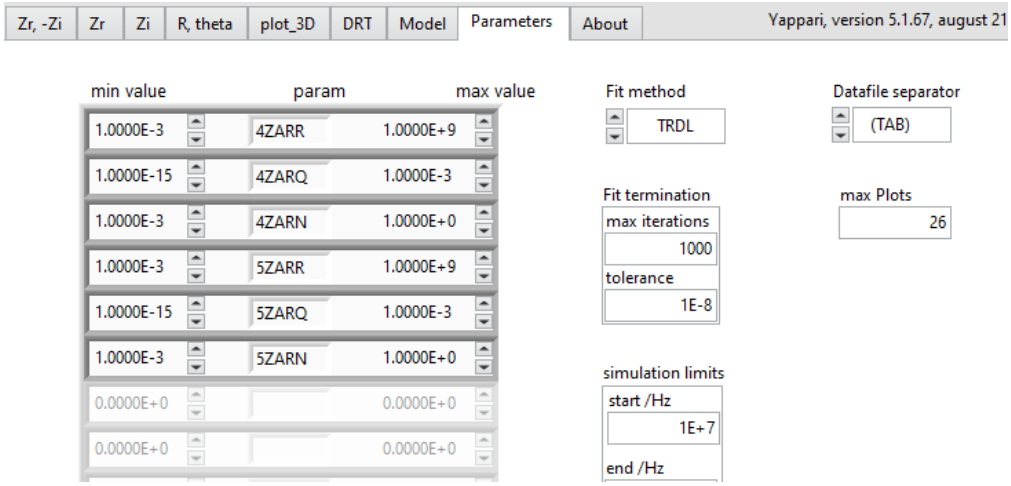


The colors and style of the graphs can be adjusted by the user in the Nyquist plot, they are propagated to the other plots. You can plot multiple datasets, just select the ones you want to see. You can use standard Windows commands in the *Datalists* listbox , Ctrl+A to select all, like in this image.

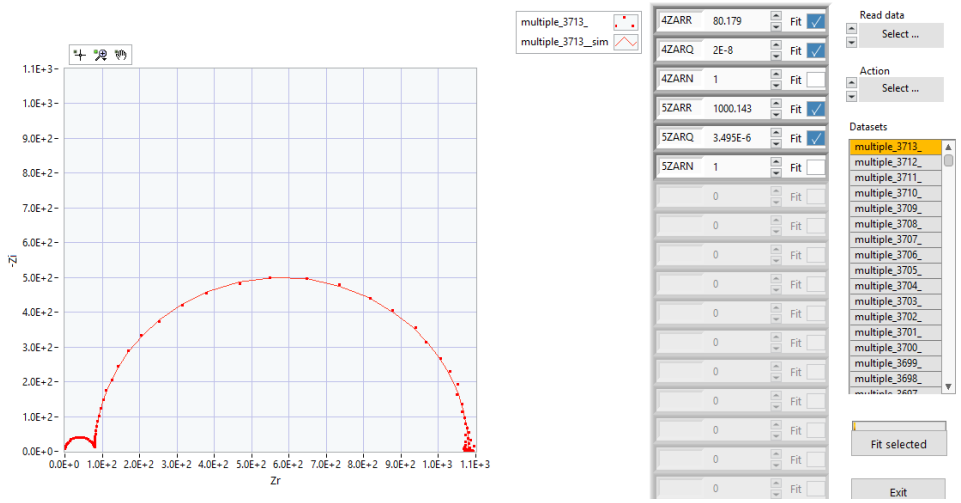


So, now we have 3174 datasets selected but not all are plotted, see *Max plots* in the *Parameters* panel.

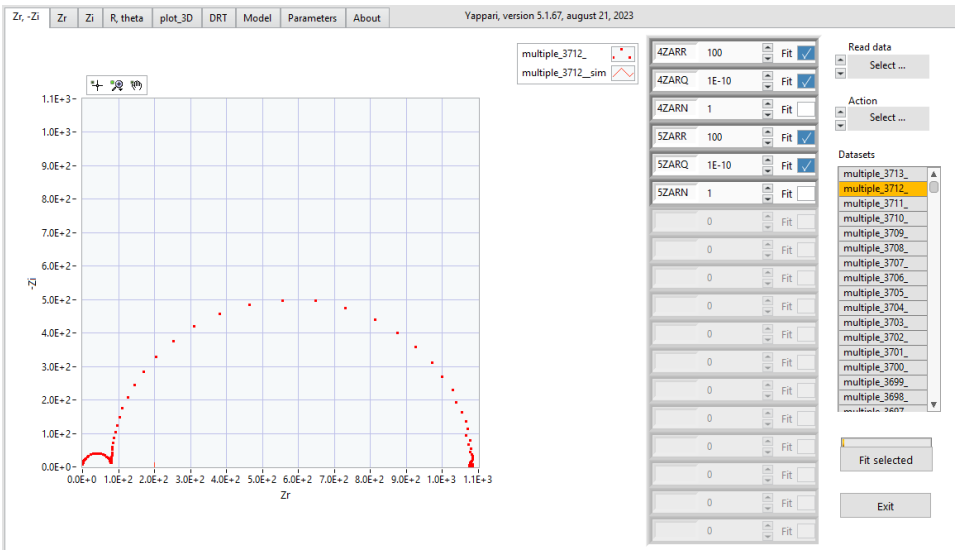
Before fitting you should go and setup the method, number of iterations,.. By default the method is TRDL, I suggest to keep it for now. You can adapt the parameter limits but in general there is no need.



The simplest way is to fit the first dataset : just select the first dataset and then press **Fit selected**, you get the results directly in the list : 80 Ohms for R1, etc.

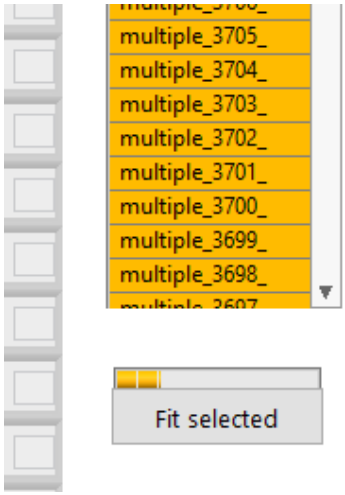


If we take a look at the second dataset, we see that the parameters are the default ones. It is normal, we did not fit it. All datasets have the same model and parameters but the values of these parameters are different for each dataset.

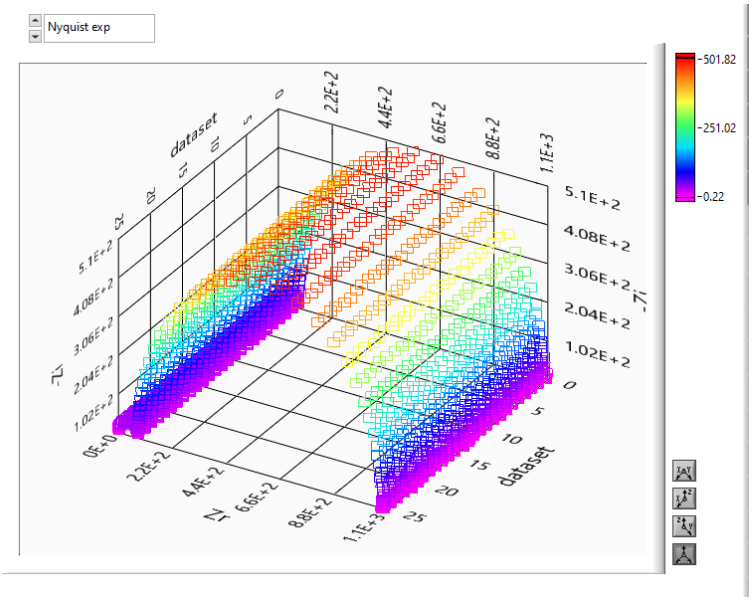


To fit all, one could select all datasets and fit all. It is possible but it's better to start from closer values, the fit will be faster and there is another consequence, I will explain below, there is a special situation for the case of symmetric circuits.

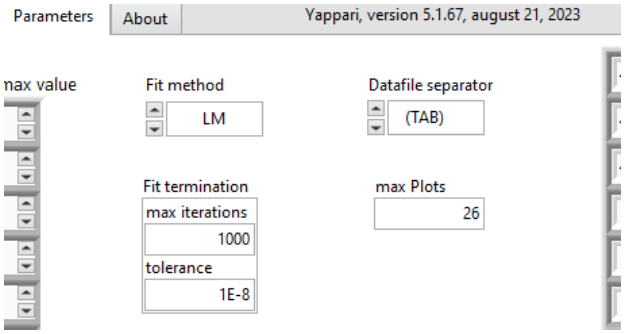
So, select the first dataset (with the fitted parameters) go to *Action*, then select *Clone the parameters to all*. Now the parameters of the first dataset are copied to all datasets providing a good starting point. It is better to change the method to LM, in order to obtain the esd's of the fitted parameters. Now click on the datasets and Ctrl+A, then *Fit selected* you should see something like this :



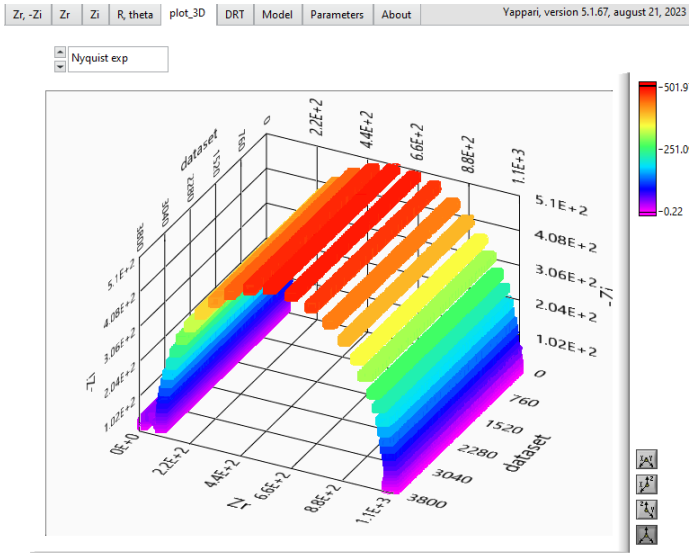
The yellow bar shows the advancement of the fit. For these 3700 datasets, the fit takes less than one minute on my PC. Let's look at the results in the graph. For instance the 3D graphics panel shows this image and we see only 26 datasets:



By default the graphs show 26 plots (and there is a max 26 legends, for reasons of space). If you want to plot more, go to Parameters panel and change **Max plots** value

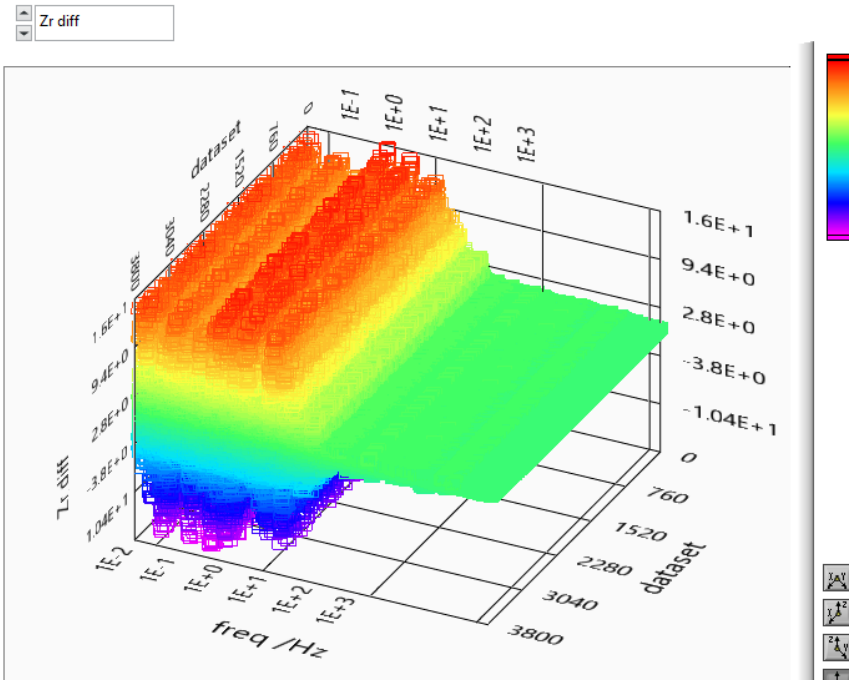


Put a big number if you want, instead of 26, I put 5000. Now you can see all 3714 plots in the graph, parameters, etc.



Beware that it will take more than a minute for this graph (I have to figure out why..) so don't "stay" on the 3D panel or the program will be very slow, it will appear frozen.

You can also plot the differences, again, it takes a lot of time for almost 4000 datasets :



Again : don't stay on this 3D view for other operations, since a draw is requested for each change, it will take a lot of time.

Now we have all the results, one can see in the 3D plots that the difference between calculated and experimental data is about 1%, which is likely due the noise I added to the data. You can save parameters in a datafile : *Action*, then *Save active parameters*. The data is saved in this format :

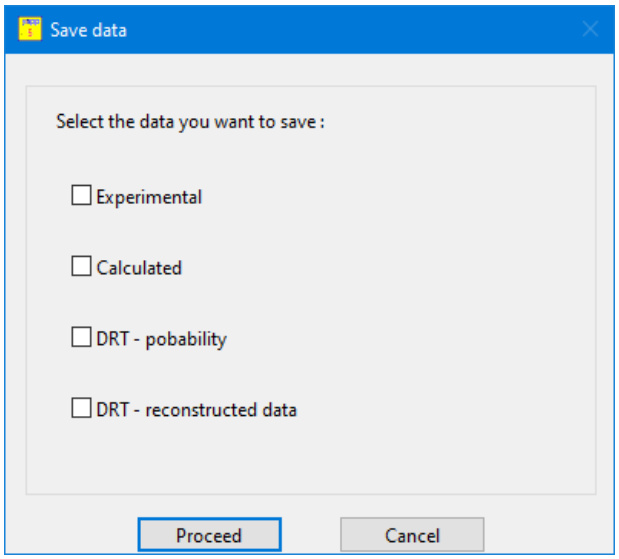
```
Yappari v5 - parameters saved : 21/08/2023 11:30:14

Dataset R2      chi2    4ZARR   sd_4ZARR        4ZARQ   sd_4ZARQ        4ZARN   sd_4ZARN        5ZARR
        sd_5ZARR        5ZARQ   sd_5ZARQ        5ZARN   sd_5ZARN

multiple_3713_ 9.999542E-1    1.224886E-2    8.017908E+1    5.380390E-1    1.999934E-8
        4.907372E-10    1.000000E+0    0.000000E+0    1.000143E+3    7.028813E-1    3.494608E-6
        7.821299E-9    1.000000E+0    0.000000E+0

multiple_3712_ 9.999891E-1    2.999330E-3    8.026979E+1    2.620355E-1    1.998851E-8
        2.386356E-10    1.000000E+0    0.000000E+0    9.995219E+2    3.423665E-1    3.506031E-6
        3.824405E-9    1.000000E+0    0.000000E+0

multiple_3711_ 9.999908E-1    2.503571E-3    7.992274E+1    2.414201E-1    2.000230E-8
        2.209878E-10    1.000000E+0    0.000000E+0    9.997856E+2    3.154165E-1    3.495010E-6
        3.511676E-9    1.000000E+0    0.000000E+0

multiple_3710_ 9.999925E-1    2.077267E-3    7.982877E+1    2.176960E-1    2.003200E-8
        1.998151E-10    1.000000E+0    0.000000E+0    9.995805E+2    2.844387E-1    3.499832E-6
        3.171765E-9    1.000000E+0    0.000000E+0
```

For each dataset there is a line containing the name of the dataset, R2, chi2 and all the parameters and their esds (if the fit was done with the LM method). A 0.000 for esd means that : the fit was not made with the LM method, or that parameter was not fitted.

Note that for symmetric circuits, like here where there are 2 Zarcs, you may have alternate results. For the Zarc1 the parameters are 80 Ohm, 2E-8F and n=1, while for the Zarc2 the parameters are 1Kohm and 3.6 microF, n=1. So, if you start fitting from far away positions you may get a result with Zarc2 as having 80 Ohm, 2E-8F and n=1, while for the Zarc1 the parameters can be 1Kohm and 3.6 microF, n=1. They are mathematically equivalent, so if you want to further plot these 3714 parameters these alternate results may appear strange. This is the reason to start fitting with parameters close to the solution and to avoid local minima.

You can also use *Report active* function which will create a file with all the graphs and all the results, but for large data sets I do not recommend it as it will be a huge file, for this case with 12000 images !

You can use *Save data* which opens a selection window in which you can select one or more sets of data:



If the data is not available, for instance there is no DRT calculation, in the file you will see NaNs.

More information is available in the Readme.md file on github repository.