RootProf

TUTORIAL 4

Covariance analysis

Motivation

Correlating measurements on the same samples by two different experimental techniques. This off-line analysis allows identifying the correlated features of the respective profiles.

The data set

Unidimensional profiles on polycrystalline mixtures have been acquired by using X-ray powder diffraction and FT-IR spectroscopy. Experimental samples have been produced by crystallization processes aiming at obtaining co-crystals formed by an active pharmaceutical ingredient (API) θa co-former. In our case polymorph III of carbamazepine constitutes the API, saccharine the co-former. The experimentally determined weight fractions of carbamazepine (CBZ III), saccharine (SAC) and co-crystal (CBZ-SAC) are reported in Table 1. The corresponding files are included as demo files. They are formed by two columns, the first containing the 2θ or λ^{-1} values, the second the corresponding values of intensity.

Table 1: Weight fractions of prepared mixtures. Samples 6-8 (shadowed) are composed by pure phases.

Sample n.	CBZ III	SAC	CBZ-SAC	XRPD	FT-IR
0	0	0.565	0.435	Rocco_S3_mac.txt	Sample 3a.asc.extract
1	0.500	0.500	0	Rocco_S5_mac.txt	Sample 5a.asc.extract
2	0.500	0	0.500	Rocco_S7_Como.txt	Sample S7a.asc.extract
3	0.347	0.334	0.319	Rocco_S11_mac.txt	Sample 11a.asc.extract
4	0.263	0.482	0.255	Rocco_S21_mac.txt	Sample 21a.asc.extract
5	0.238	0.364	0.399	Rocco_S22_mac.txt	Sample 22a.asc.extract
6	1	0	0	Rocco_CBZ_III_nomac.txt	CBZ III.asc.extract
7	0	1	0	Rocco_SAC_pura_nomac.txt	SAC.asc.extract
8	0	0	1	Rocco_CBZSAC_90511_n.txt	CBZ-SAC.asc.extract

The command file

The list of commands for covariance analysis of such dataset is the following.

```
  whichanalysis 5
  figpaper 1
!--first dataset: XRPD
  dataType 2
  range 10 50
  preprocess 0 2
  file  Rocco_S3_mac.txt
  file  Rocco_S5_mac.txt
  file  Rocco_S7_Como.txt
  file  Rocco_S11_mac.txt
  file  Rocco_S21_mac.txt
  file  Rocco_S22_mac.txt
  file  Rocco_CBZ_III_nomac.txt
  file  Rocco_SAC_pura_nomac.txt
  file  Rocco_CBZSAC_90511_n.txt
!--second dataset: FT-IR
  dataType2 1
  preprocess2 0 3
  file2  sample_3a.asc.extract
  file2  sample_5a.asc.extract
  file2  sample_S7a.asc.extract
  file2  sample_11a.asc.extract
  file2  sample_21a.asc.extract
  file2  sample_22a.asc.extract
  file2  CBZ_III.asc.extract
  file2  SAC.asc.extract
  file2  CBZ-SAC.asc.extract
  range2 3000 6000
  rowsel 3460 3500
  colsel 13.2 14.0
```

They have been included in the demo file named *fileInputCovariance*. See the user guide for an explanation of each command.

Running RootProf

Start ROOT by clicking on his icon, or by typing “root” on a terminal window. Then write the root command:

```
Root> .x RootProf_v15.C(“fileInputCovariance”)
```

or

```
Root> .> outputCovariance
.x RootProf_v15.C(“fileInputCovariance”)
.>
```

After some seconds, graphic windows will start appearing on your screen, while text output will appear on the terminal window, or redirected in the file named *outputCovariance*. When the run

ends, the root prompt will appear again on the ROOT terminal, and you will be able to edit each single graphic window and read the output file by your text editor.

The graphic output

The first graphic window produced contains the profiles of the two datasets (Figs. 1 to 4). In Figs. 1 and 3 they are shifted in the vertical axis, and ordered bottom to top; in Figs. 2 and 4 they are put together in a data matrix, colored according to the spectrum intensity. In our case data consists of XRPD patterns in the 2θ range [10°, 50°], where the highest peaks are included, and FT-IR spectra in the wavenumber range [3000 cm^{-1}, 6000 cm^{-1}], which is the region dominated by H-bonds.

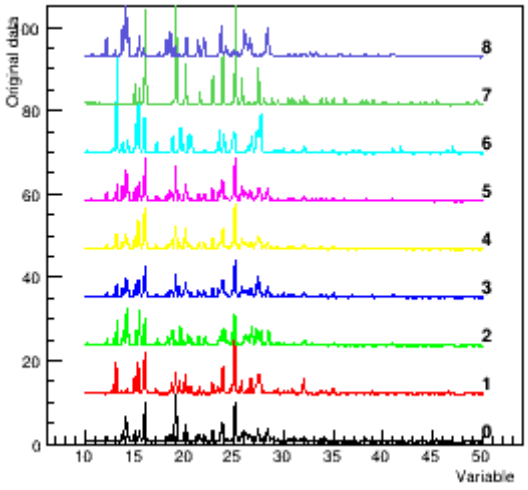

Fig.1 Original data shifted (First dataset)

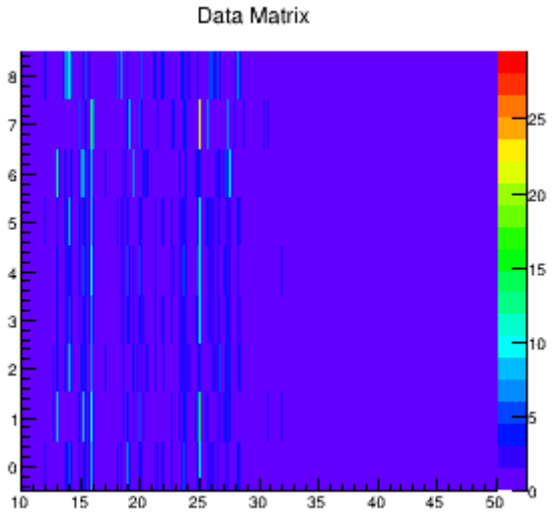

Fig.2 Data matrix (First dataset)

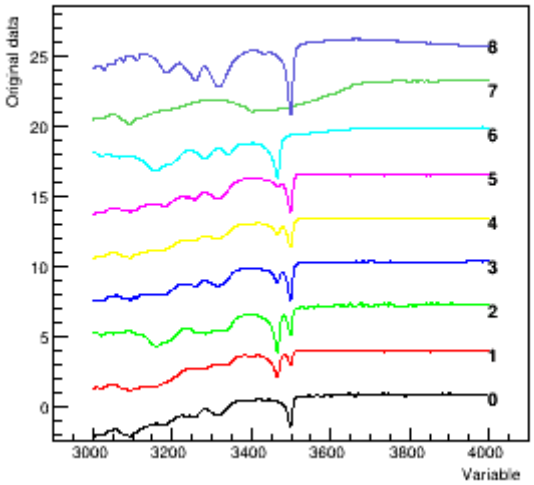

Fig.3 Original data shifted (Second dataset)

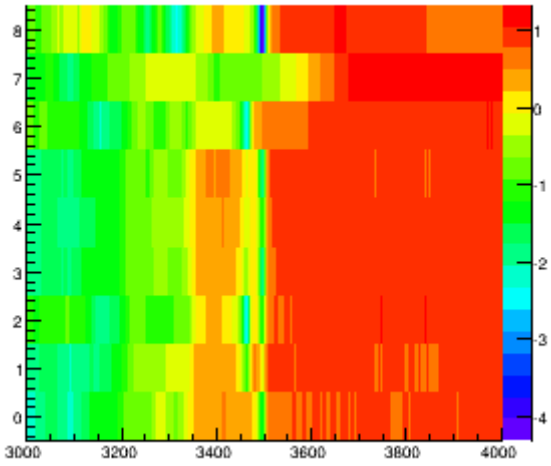

Fig.4 Data matrix (second dataset)

The covariance matrix is displayed as in Fig.5. The color scale is reported on the vertical bar on the right, so that red spots in the matrix indicate positive peaks, blue spots negative ones. Usually the number of bins is such high that the characteristic features of the matrix cannot be seen at first sight. A zoom in both X and Y scales made by the user on the graphic window can help in this case, so that each part of the matrix can be examined in details. In Fig.5a, for example, a region of the covariance matrix is highlighted, comprising 2θ angles between 12° and 20° in the X-axis and λ^{-1} values between 3440 and 3520 cm^{-1} in the Y-axis. By comparing these features with the XRPD and FT-IR spectra from pure phases (the top three spectra in Figs.1 and 3, respectively), it can be seen how the two peaks at 3460 cm^{-1} and 3500 cm^{-1} correlate with the XRPD signal. The spots in the covariance matrix at 13.2° and 15.5° have a negative peak at 3460 cm^{-1} and a positive one at 3500

cm^{-1}: these 2θ values coincide with those of the peaks of the CBZ III XRPD pattern which are not common with CBZ-SAC and SAC. An opposite trend is followed by the spot at 14°, which has a positive peak at 3460 cm^{-1} and a negative one at 3500 cm^{-1}: it is characteristic of the CBZ-SAC XRPD pattern. Finally, two positive spots in the covariance matrix at 3460 cm^{-1} and 3500 cm^{-1} occur at 16° and 19.5°, in correspondence of the main characteristic peaks of the SAC XRPD pattern. Thus, the analysis of the covariance matrix allows identifying the relationships between XRPD and FT-IR signals, or in general of the profiles from the two techniques considered. In this case, it was found that the relative height of two FT-IR peaks, at 3460 cm^{-1} and 3500 cm^{-1}, is very sensitive for the discrimination of CBZ III, SAC and CBZ-SAC.

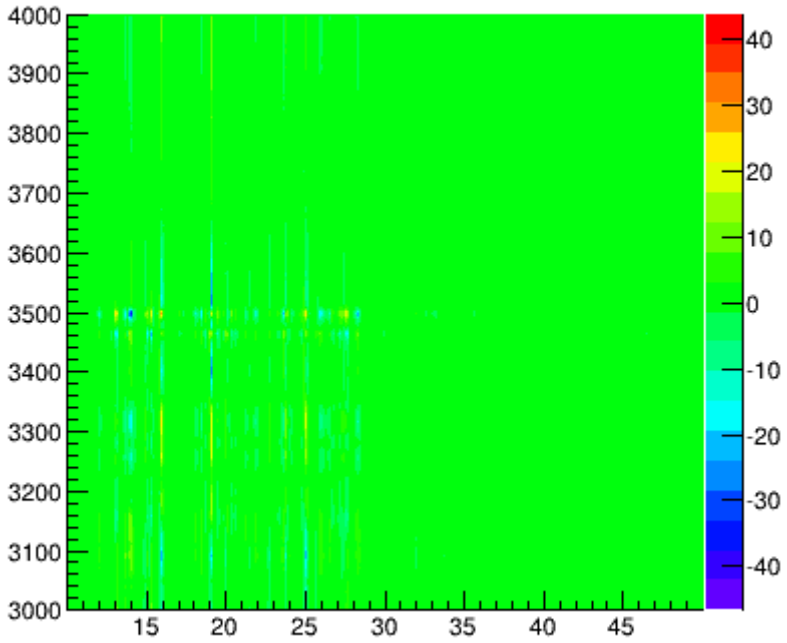

Fig.5 Covariance Matrix

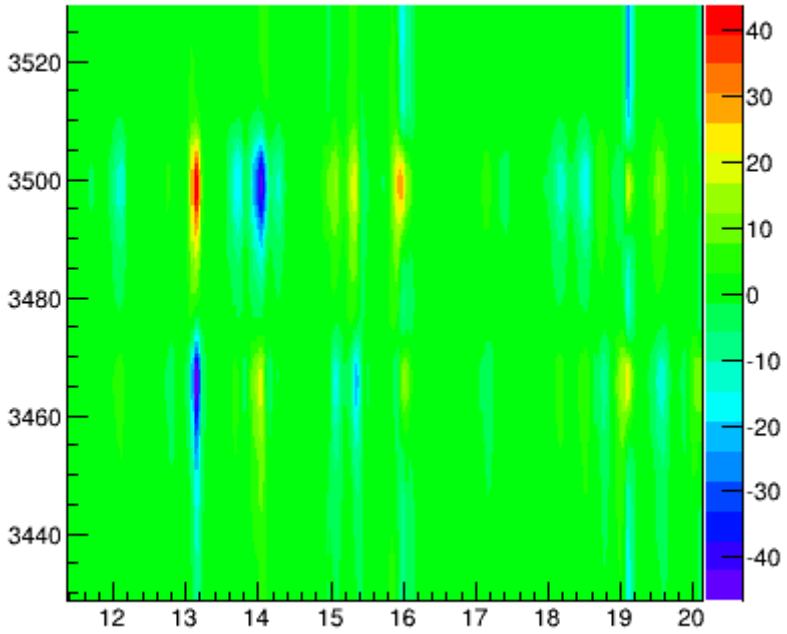

Fig.5a Covariance Matrix (zoomed)

Slices of the covariance matrix can be plotted row-wise and column-wise at wanted values, given in input by the commands *rowsel* and *colsel*, respectively. They are shown in Figs. 6 and 7, respectively. It can be seen that slices at rows 3460 and 3500 cm^{-1} (Fig. 6) are roughly one opposite the other, as expected from the previous findings. However, the peak at 20° is conserved in both slices. This is because both the CBZ-SAC and SAC XRPD patterns have a peak at such 2θ angle. Instead the slices at columns 13.2° and 14° (Fig. 7) show the features of the FT-IR spectra which are related to respectively CBZ III and CBZ-SAC pure phases.

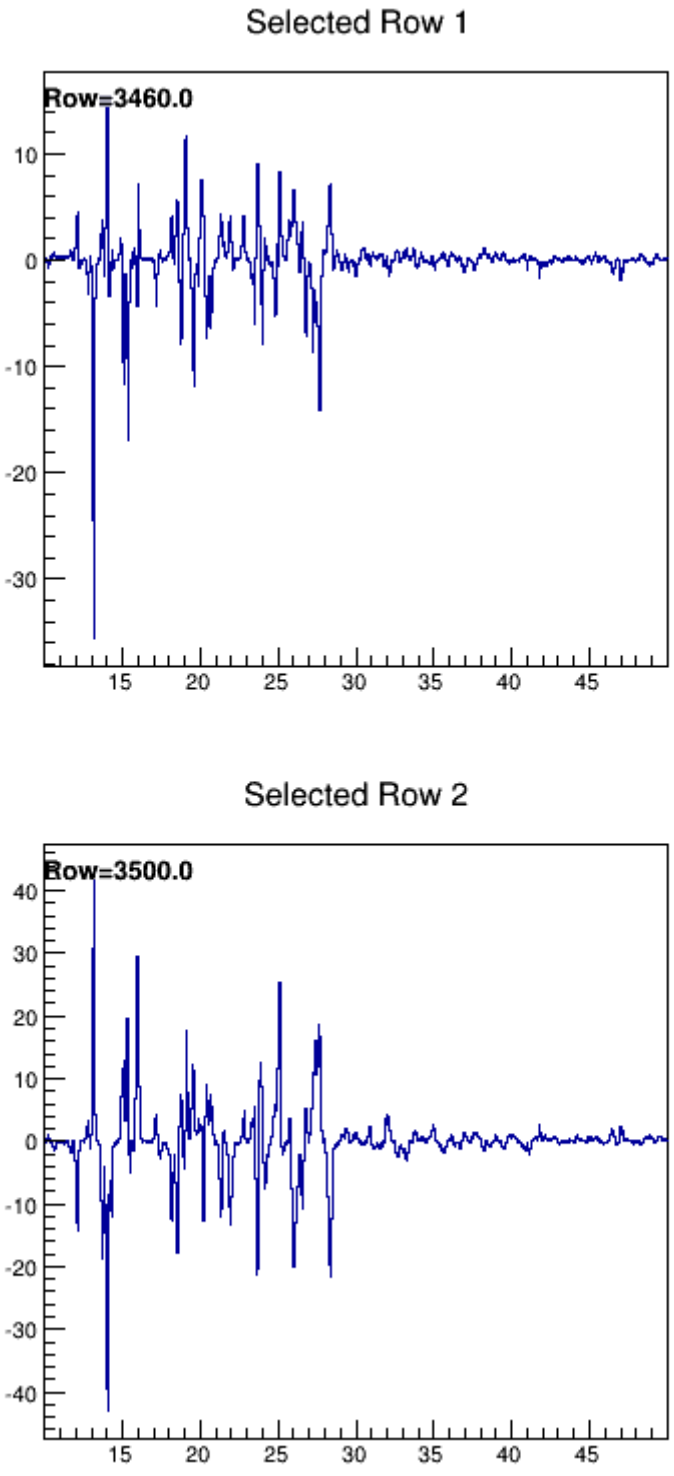

Fig.6 Selected rows

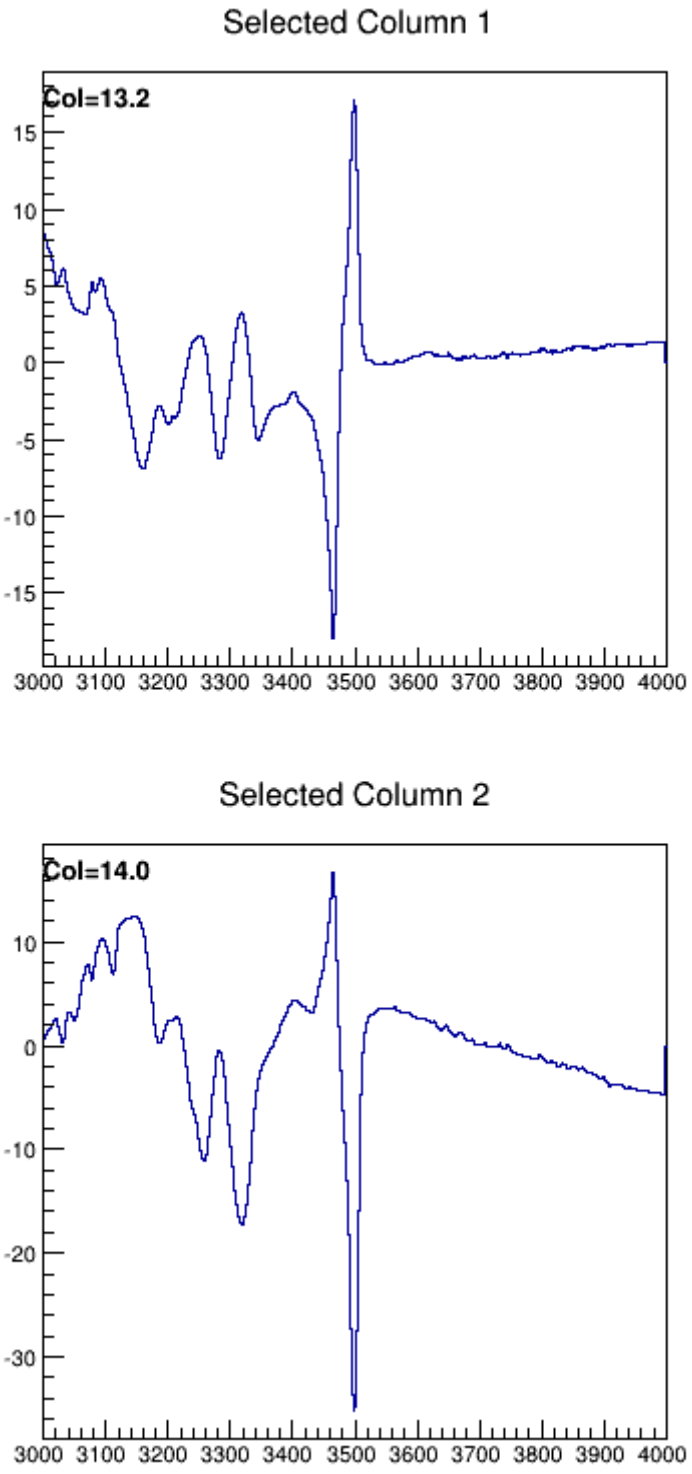

Fig.7 Selected columns

The Output file

The content of the output file named *outputCovariance* is reported below, with comments added.

```
 Input from file: fileInputCovariance
--------------------------------
  whichanalysis 5

  figpaper 1

  dataType 2

  range 10 50

  preprocess 0 2

  file  Rocco_S3_mac.txt

  file  Rocco_S5_mac.txt

  file  Rocco_S7_Como.txt

  file  Rocco_S11_mac.txt

  file  Rocco_S21_mac.txt

  file  Rocco_S22_mac.txt

  file  Rocco_CBZ_III_nomac.txt

  file  Rocco_SAC_pura_nomac.txt

  file  Rocco_CBZSAC_90511_n.txt

  dataType2 4

  preprocess2 0 3

  file2  sample_3a.asc.extract

  file2  sample_5a.asc.extract

  file2  sample_S7a.asc.extract

  file2  sample_11a.asc.extract

  file2  sample_21a.asc.extract

  file2  sample_22a.asc.extract

  file2  CBZ_III.asc.extract

  file2  SAC.asc.extract

  file2  CBZ-SAC.asc.extract

  range2 3000 6000

  rowsel 3460 3500

  colsel 13.2 14.0
```

The section above shows the commands read from the command file. It should be checked to ensure that they are interpreted correctly.

```
Starting Covariance analysis


First dataset: dataType=2

 Reading input files:
 ----------------------
Sample 0 -> file Rocco_S3_mac.txt
          Found 1999  points
Sample 1 -> file Rocco_S5_mac.txt
          Found 1999  points
Sample 2 -> file Rocco_S7_Como.txt
          Found 1999  points
Sample 3 -> file Rocco_S11_mac.txt
          Found 1999  points
Sample 4 -> file Rocco_S21_mac.txt
          Found 1999  points
Sample 5 -> file Rocco_S22_mac.txt
          Found 1999  points
Sample 6 -> file Rocco_CBZ_III_nomac.txt
          Found 1999  points
Sample 7 -> file Rocco_SAC_pura_nomac.txt
          Found 1999  points
Sample 8 -> file Rocco_CBZSAC_90511_n.txt
          Found 1999  points
```

The section above reports the number of data points read within each input file of the first dataset.

```
Second dataset: dataType=4

 Reading input files:
 ----------------------
Sample 0 -> file sample_3a.asc.extract
          Found 1001  points
Sample 1 -> file sample_5a.asc.extract
          Found 1001  points
Sample 2 -> file sample_S7a.asc.extract
          Found 1001  points
Sample 3 -> file sample_11a.asc.extract
          Found 1001  points
Sample 4 -> file sample_21a.asc.extract
          Found 1001  points
Sample 5 -> file sample_22a.asc.extract
          Found 1001  points
Sample 6 -> file CBZ_III.asc.extract
          Found 1001  points
Sample 7 -> file SAC.asc.extract
          Found 1001  points
Sample 8 -> file CBZ-SAC.asc.extract
          Found 1001  points
```

The section above reports the number of data points read within each input file of the second dataset.