RootProf

TUTORIAL 8

Analysis of FT-IR spectra

Contents

Chapter 1

The data set

Unidimensional spectra from Fourier transformed infrared (FT-IR) spectroscopy measurements taken in transmission mode compose our dataset. All spectra have been acquire by the same spectrophotometer, by using KBr pellets. Experimental samples have been produced by crystallization processes aiming at obtaining co-crystals formed by an active pharmaceutical ingredient (API) and a 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 wavenumber values, the second the corresponding intensity.

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

Sample n.	CBZ III	SAC	CBZ-SAC	File name
0	0	0.565	0.435	Sample_3a.asc.extract
1	0.500	0.500	0	Sample_5a.asc.extract
2	0.500	0	0.500	Sample_S7a.asc.extract
3	0.347	0.334	0.319	Sample_11a.asc.extract
4	0.263	0.482	0.255	Sample_21a.asc.extract
5	0.238	0.364	0.399	Sample_22a.asc.extract
6	1	0	0	CBZ_III_nomac.asc.extract
7	0	1	0	SAC.asc.extract
8	0	0	1	CBZ-SAC.asc.extract

Chapter 2

Pre-processing of FT-IR spectra

Motivation

Obtaining a quick and joint view of all input spectra, comparing and inspecting their features, and testing the effect of pre-processing.

The command file

The list of commands is the following.

```
  whichanalysis 0
  figpaper 1
  dataType 4
  range 450 4000
   preprocess  0 3
!  preprocess  0 3 15
  file  sample_3a.asc.extract
  file  sample_5a.asc.extract
  file  sample_S7a.asc.extract
  file  sample_11a.asc.extract
  file  sample_21a.asc.extract
  file  sample_22a.asc.extract
  file  CBZ_III.asc.extract
  file  SAC.asc.extract
  file  CBZ-SAC.asc.extract
```

The commands have been included in the demo file named *fileInputIRFirstSight*. See the user guide for an explanation of their meaning.

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(“fileInputIRFirstSight”)

or

Root> .> outputIRFirstSight

.x RootProf_v15.C(“fileInputIRFirstSight”)

.>

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 *outputIRFirstSight*. 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 graphic output after pre-processing (Figs. 1 and 2) shows the FTIR spectra rescaled through SNV pre-processing (command *preprocess* 0 3). We verified that this rescaling allow eliminating the differences due to the amount of KBr put in the pellets. However it does not produce good classification of spectra, neither precise quantification of the weight fractions.

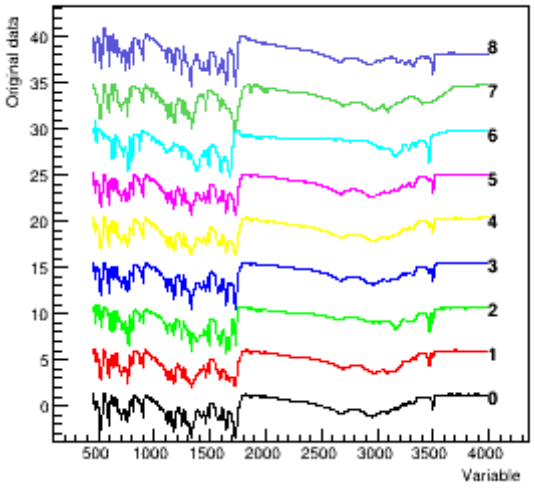


Fig.1 Original data shifted (after preprocessing)

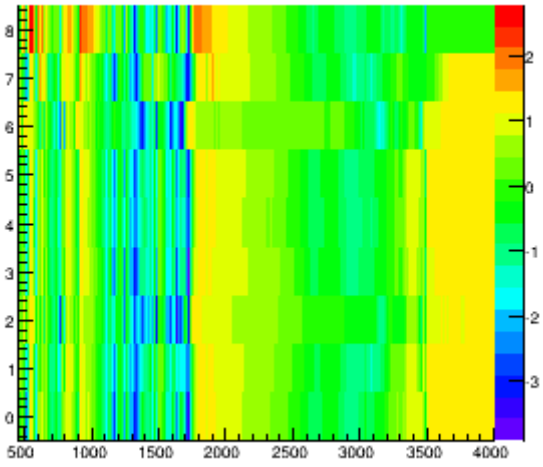


Fig.2 Data Matrix (after preprocessing)

If instead the command *preprocess* 0 3 15 is given in input, a background subtraction performed by the SNIP algorithm with a very small clipping window is performed, in addition to the SNV rescaling. The choice of such a small window cause the background to enter into the peaks, therefore rather that background subtraction, the operation could be called filtering by the SNIP algorithm. As such, the FT-IR spectra are completely modified after pre-processing (Fig.3), and they have now a common baseline, the features in the range 450-1750 cm^{-1} aligned to those in the range 3000-3500 cm^{-1}, and very sharp peaks, which form different patterns among the spectra. Even the data matrix (Fig.4) has now much more features than the previous one (Fig.2). Thus *preprocess* 0 3 15 was used for further qualitative and quantitative analysis.

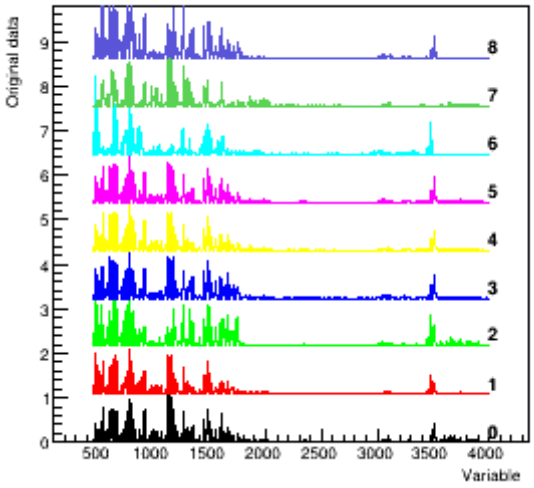


Fig.3 Original data shifted (after preprocessing)

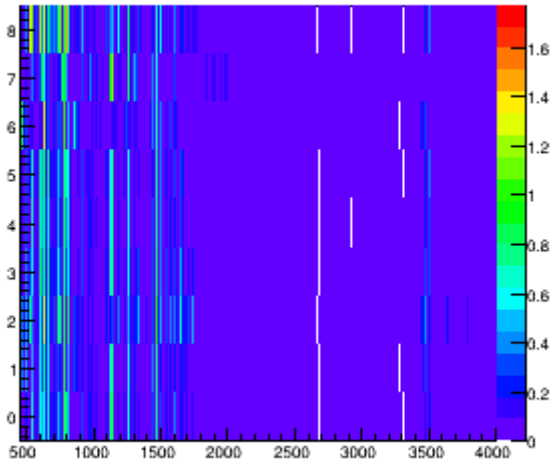


Fig.4 Data Matrix (after preprocessing)

The output file

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

```
 Input from file: fileInputIRFirstSight
------------------------------
  whichanalysis 0

  figpaper 1

  dataType 4

  range 450 4000

   preprocess  0 3

  file  sample_3a.asc.extract

  file  sample_5a.asc.extract

  file  sample_S7a.asc.extract

  file  sample_11a.asc.extract

  file  sample_21a.asc.extract

  file  sample_22a.asc.extract

  file  CBZ_III.asc.extract

  file  SAC.asc.extract

  file  CBZ-SAC.asc.extract
```

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

```
Reading input files:
 ----------------------
Sample 0 -> file sample_3a.asc.extract
         Found 3551  points
Sample 1 -> file sample_5a.asc.extract
         Found 3551  points
Sample 2 -> file sample_S7a.asc.extract
         Found 3551  points
Sample 3 -> file sample_11a.asc.extract
         Found 3551  points
Sample 4 -> file sample_21a.asc.extract
         Found 3551  points
Sample 5 -> file sample_22a.asc.extract
         Found 3551  points
Sample 6 -> file CBZ_III.asc.extract
         Found 3551  points
Sample 7 -> file SAC.asc.extract
         Found 3551  points
Sample 8 -> file CBZ-SAC.asc.extract
         Found 3551  points
```

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

Chapter 3

Qualitative analysis of FT-IR spectra

Motivation

Apply PCA for classification of FT-IR spectra pre-processed by SNV and SNIP filtering.

The command file

The list of commands is the following.

```
whichanalysis 1
figpaper 1
dataType 4
range 450 4000
skipdata 5
preprocess  0 3 15
file  sample_3a.asc.extract
file  sample_5a.asc.extract
file  sample_S7a.asc.extract
file  sample_11a.asc.extract
file  sample_21a.asc.extract
file  sample_22a.asc.extract
file  CBZ_III.asc.extract
file  SAC.asc.extract
file  CBZ-SAC.asc.extract
clusterswitch 0
```

The commands have been included in the demo file named *fileInputIRQualitative*. See the user guide for an explanation of their meaning.

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.C(“fileInputIRQualitative”)
```

or

```
Root> .> outputIRQualitative
.x RootProf.C(“fileInputIRQualitative”)
.>
```

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 *outputIRQualitative*. 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

PCA analysis has been applied to the spectra and data matrix shown respectively in Figs.1 and 2. They differs from those shown in Figs 1 and 2 of chapter 2 because of the command *skipdata* 5, which was included to speed up the PCA analysis. Since a data point every 5 is taken from input file, this command causes a slight broadening of the spectra.

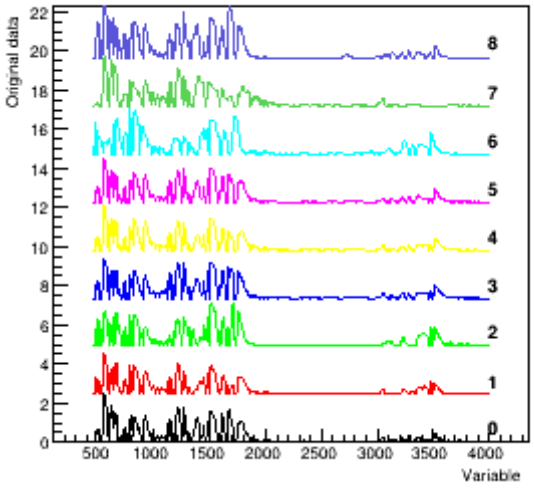


Fig.1 Original data shifted

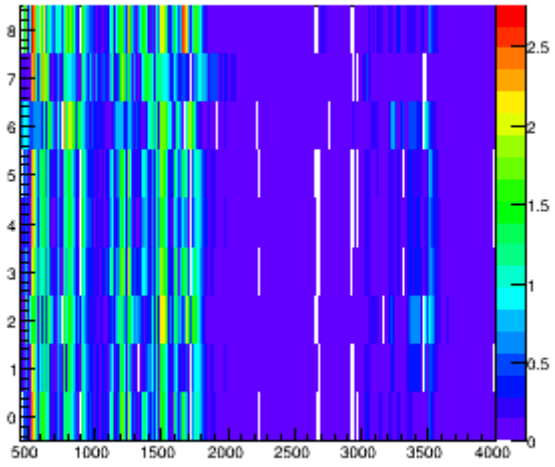


Fig.2 Data Matrix

In Fig.3 is reported the Scree plot, and in Figs. 4, 5 respectively the loadings and scores for the first two principal components. PC1 is able to distinguish CBZ III from SAC and CBZ-SAC phases, while PC2 separates SAC from CBZ-SAC. Both the 450-1750 cm^{-1} and 3000-3500 cm^{-1} ranges heavily contribute to PC1 and PC2.

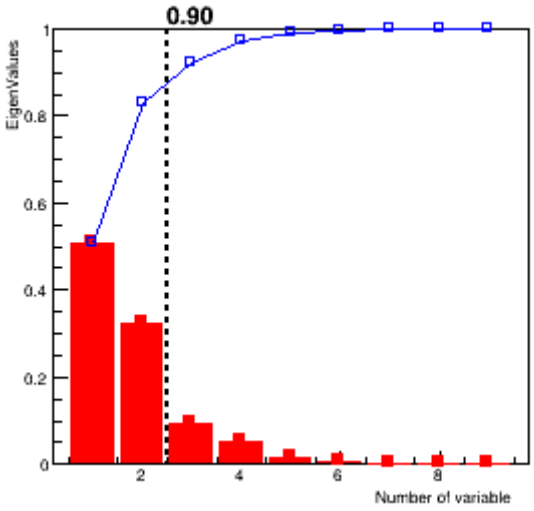


Fig.3 Scree plot

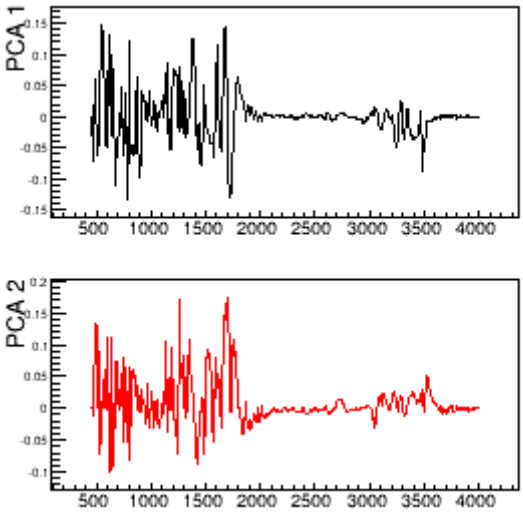


Fig.4 Loadings

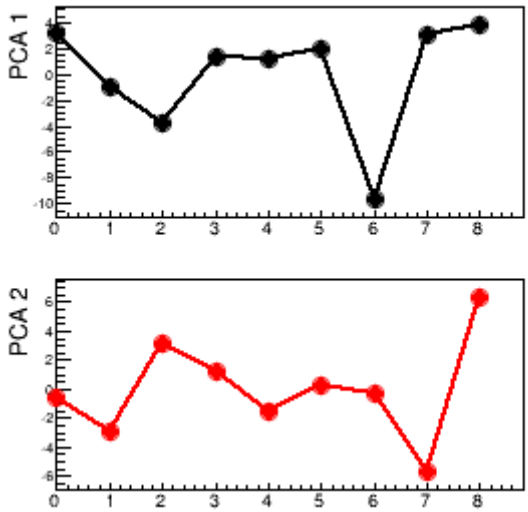
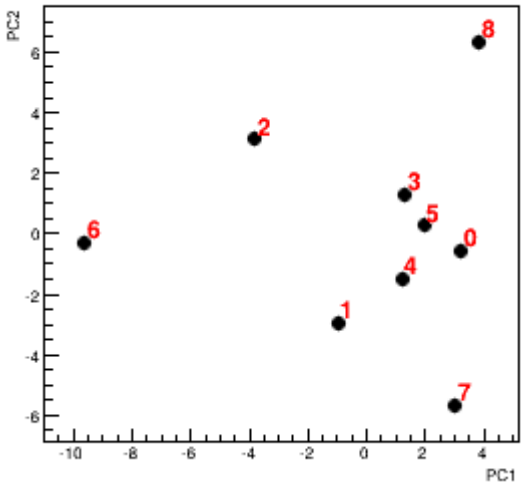


Fig.5 Scores

The score plot in Fig.6 shows that pure phase spectra (samples 6,7,8) are well separated, binary samples (0,1,2) are roughly placed in between two pure phase points, and intermediate samples (3,4,5) are located within the triangle formed by the pure phase spectra. Such approximated ternary diagram is much more distorted if no background subtraction is applied.



/

Fig.6 Score plot PC1-PC2

The output file

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

```
 Input from file: fileInputIRQualitative
------------------------------
  whichanalysis 1
```

```
  figpaper 1

  dataType 4

  range 450 4000

  skipdata 5

  preprocess  0 3 15

  file  ./datiIR/sample_3a.asc.extract

  file  ./datiIR/sample_5a.asc.extract

  file  ./datiIR/sample_S7a.asc.extract

  file  ./datiIR/sample_11a.asc.extract

  file  ./datiIR/sample_21a.asc.extract

  file  ./datiIR/sample_22a.asc.extract

  file  ./datiIR/CBZ_III.asc.extract

  file  ./datiIR/SAC.asc.extract

  file  ./datiIR/CBZ-SAC.asc.extract

  clusterswitch 0
```

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

```
Reading input files:
 ----------------------
Sample 0 -> file ./datiIR/sample_3a.asc.extract
         Found 711  points
Sample 1 -> file ./datiIR/sample_5a.asc.extract
         Found 711  points
Sample 2 -> file ./datiIR/sample_S7a.asc.extract
         Found 711  points
Sample 3 -> file ./datiIR/sample_11a.asc.extract
         Found 711  points
Sample 4 -> file ./datiIR/sample_21a.asc.extract
         Found 711  points
Sample 5 -> file ./datiIR/sample_22a.asc.extract
         Found 711  points
Sample 6 -> file ./datiIR/CBZ_III.asc.extract
         Found 711  points
Sample 7 -> file ./datiIR/SAC.asc.extract
         Found 711  points
Sample 8 -> file ./datiIR/CBZ-SAC.asc.extract
         Found 711  points
```

The section above reports the number of data points read within each input file, as determined by the command *range*.

```
Starting Qualitative analysis
```

```
 n. points 711
Eigenvalues: 1 --> 50.56% (50.6%)
Eigenvalues: 2 --> 32.27% (82.8%)
Eigenvalues: 3 --> 9.48% (92.3%)
Eigenvalues: 4 --> 4.97% (97.3%)
Eigenvalues: 5 --> 1.74% (99.0%)
Eigenvalues: 6 --> 0.59% (99.6%)
Eigenvalues: 7 --> 0.32% (99.9%)
Eigenvalues: 8 --> 0.07% (100.0%)
Eigenvalues: 9 --> 0.00% (100.0%)

 Chosen value of k=2: ratio=0.92 error=0.038
```

The section above shows the results of the PCA analysis. The first eigenvalues are listed as a function of their value, and the number of eigenvalues selected for PCA analysis is reported (k), together with the values of the threshold on the cumulative eigenvalue distribution (ratio), and an estimate of the corresponding error between original and reconstructed data (error). The threshold value is chosen on the basis of the command *threshold*. No clustering is performed on spectra, because of the command *clusterswitch* 0

Chapter 4

Unsupervised quantitative analysis of FT-IR spectra

Motivation

Apply unsupervised quantitative analysis by the MultiFit approach to FT-IR spectra filtered by the SNIP algorithm.

The command file

The list of commands is the following.

```
whichanalysis 3
figpaper 1
dataType 4
range 450 4000
preprocess  0 3 15
file  sample_3a.asc.extract
referw 0 0.565 0.435
file  sample_5a.asc.extract
referw 0.5 0.5 0
file  sample_S7a.asc.extract
referw 0.5 0 0.5
file  sample_11a.asc.extract
referw 0.347 0.334 0.319
file  sample_21a.asc.extract
referw 0.263 0.482 0.255
file  sample_22a.asc.extract
referw 0.238 0.364 0.399
file  CBZ_III.asc.extract
referw 1 0 0
purephase
file  SAC.asc.extract
referw 0 1 0
purephase
file  CBZ-SAC.asc.extract
referw 0 0 1
purephase
```

The commands have been included in the demo file named *fileInputIRQuantitative*. See the user guide for an explanation of their meaning.

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(“fileInputIRQuantitative”)
```

or

```
Root> .> outputIRQuantitative
.x RootProf_v15.C("fileInputIRQuantitative")
.>
```

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 *outputIRQuantitative*. 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 graphic window in Fig.1 show the result of the MultiFit procedure applied to the SNIP filtered FT-IR spectrum of sample 0. The pre-processed spectra (brack) are superimposed to the best fit model (red).

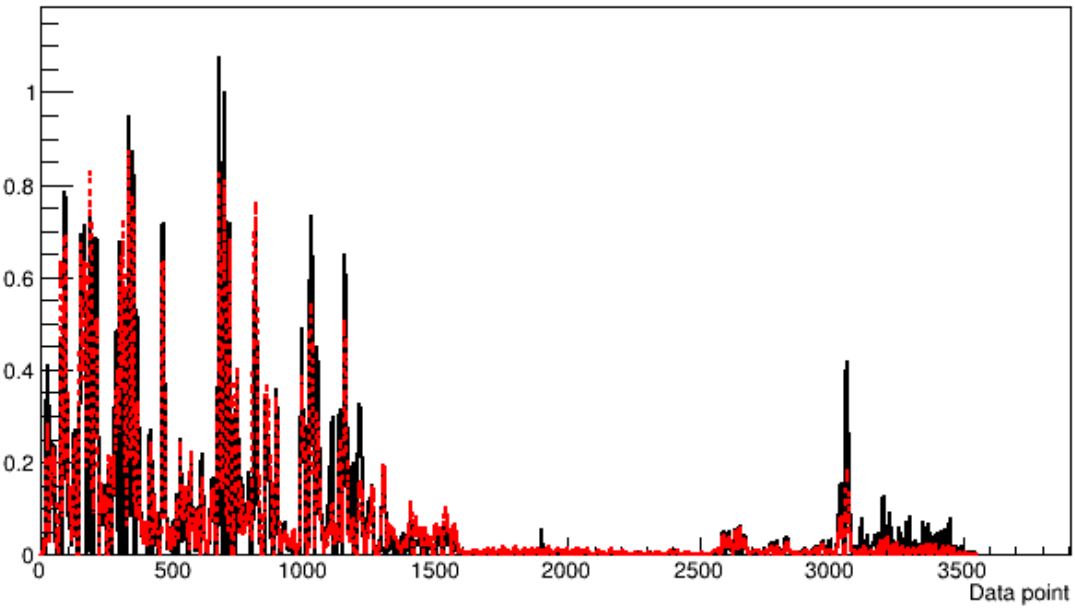


Fig.1 MultiFit on Sample 0

Weight fractions estimated by the fitting procedure are plotted against the sample number in Fig.2. The sum of the two weight fractions (green line) deviates from unity, since no such constraint was used during fitting. It can be noted that the qualitative results obtained by PCA are confirmed: Samples 6-8 are pure phases, samples 0-2 are almost binary mixtures, and samples 3-5 are almost equipopulated mixtures. Estimated weight fractions reseble the reference ones (Fig.3), and are not

very different from those obtained from analysis of X-ray powder diffraction patterns taken on the same samples (see Quantitative tutorial, chapter 3).

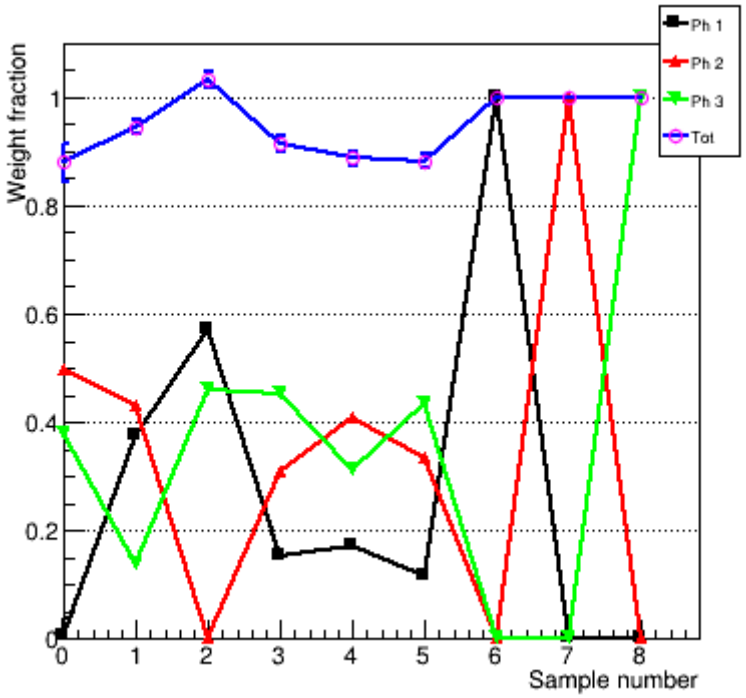


Fig.2 Quantitative Fit graph

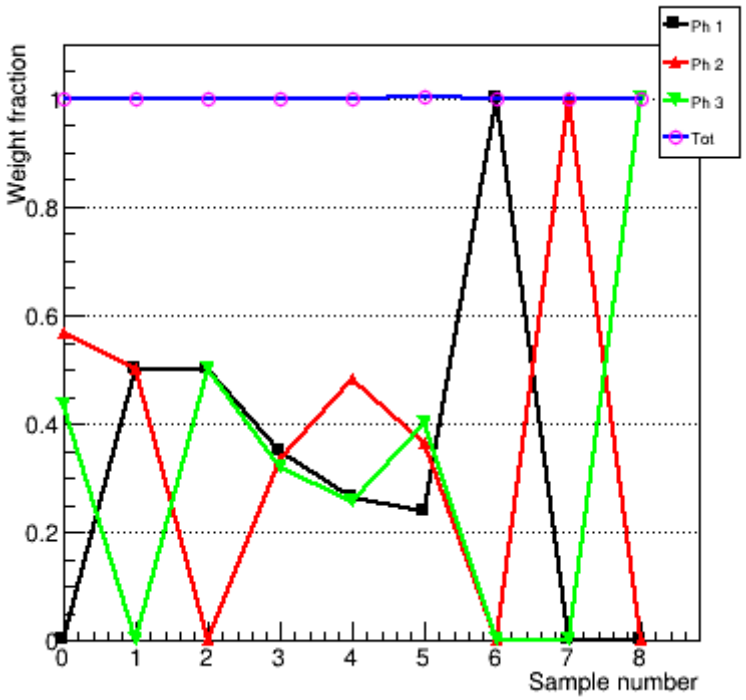


Fig.3 Quantitative Fit graph (reference weight fractions)

Figs.4-6 report the calibration plots for the three pure phases. The best fit line (red) is in all cases almost superimposed to the diagonal line (back, dashed), indicating good quantitative estimates.

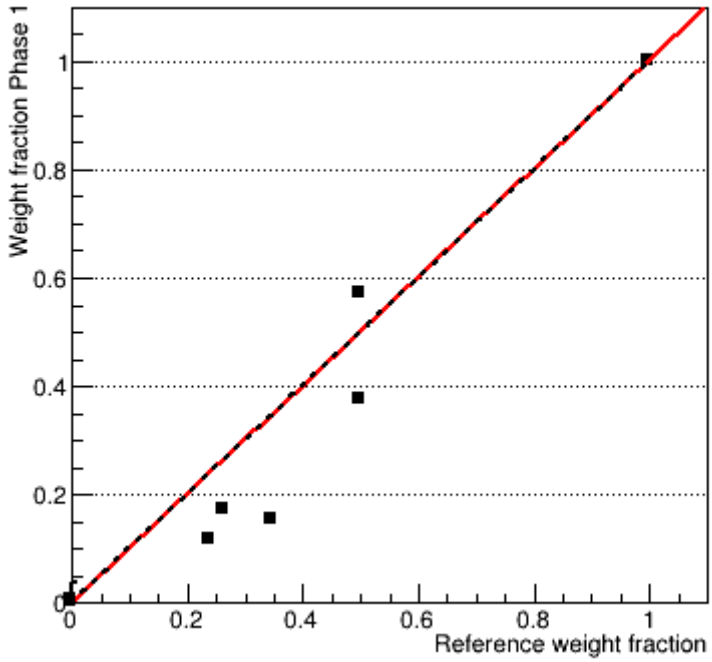


Fig.4 Calibration plot Phase 1

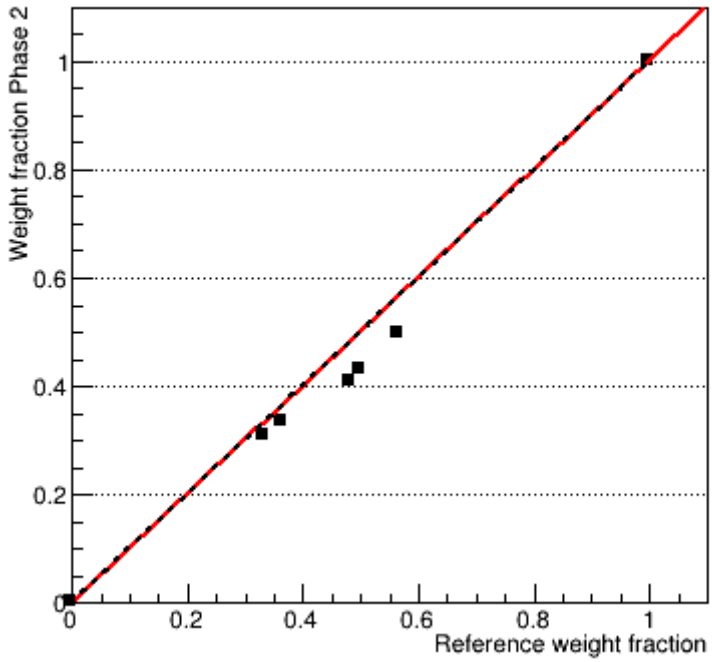


Fig.5 Calibration plot Phase 2

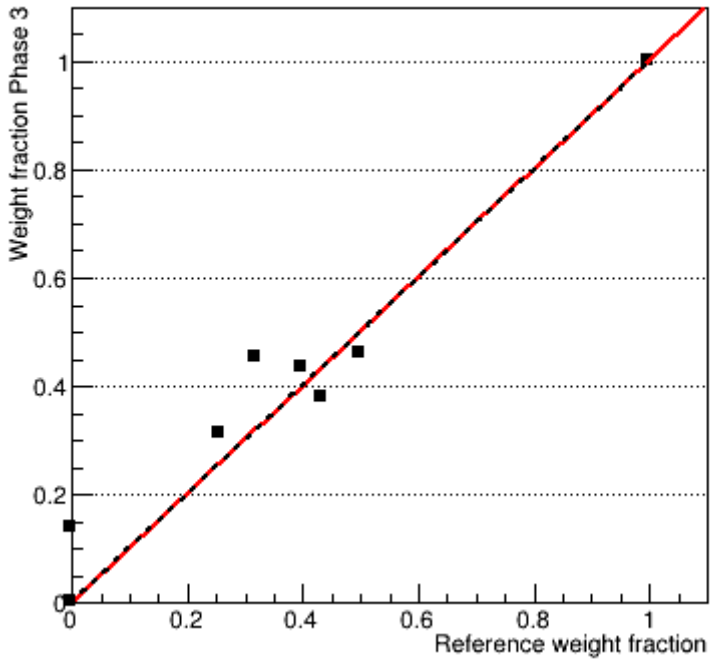


Fig.6 Calibration plot Phase 3

The output file

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

```
 Input from file: fileInputIRQuantitative
------------------------------
  whichanalysis 3

  figpaper 1

  dataType 4

  range 450 4000

  preprocess  0 3 15

  file  sample_3a.asc.extract

  referw 0 0.565 0.435

  file  sample_5a.asc.extract

  referw 0.5 0.5 0

  file  sample_S7a.asc.extract

  referw 0.5 0 0.5

  file  sample_11a.asc.extract
```

```
  referw 0.347 0.334 0.319

  file  sample_21a.asc.extract

  referw 0.263 0.482 0.255

  file  sample_22a.asc.extract

  referw 0.238 0.364 0.399

  file  CBZ_III.asc.extract

  referw 1 0 0

  purephase

  file  SAC.asc.extract

  referw 0 1 0

  purephase

  file  CBZ-SAC.asc.extract

  referw 0 0 1

  purephase
```

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

```
 Reading input files:
 ----------------------
Sample 0 -> file sample_3a.asc.extract
         Found 3551  points
Sample 1 -> file sample_5a.asc.extract
         Found 3551  points
Sample 2 -> file sample_S7a.asc.extract
         Found 3551  points
Sample 3 -> file sample_11a.asc.extract
         Found 3551  points
Sample 4 -> file sample_21a.asc.extract
         Found 3551  points
Sample 5 -> file sample_22a.asc.extract
         Found 3551  points
Sample 6 -> file CBZ_III.asc.extract
         Found 3551  points
Sample 7 -> file SAC.asc.extract
         Found 3551  points
Sample 8 -> file CBZ-SAC.asc.extract
         Found 3551  points
```

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

```
 Starting Quantitative analysis


 FIT RESULTS:
 ------------------------
```

```
Spectrum 0: sample_3a.asc.extract
Chi-Square=1.09e+01, Reduced Chi-Square=3.08e-03, NDF=3547
Weight fraction Phase 1  0.006 +- 0.034
Weight fraction Phase 2  0.496 +- 0.007
Weight fraction Phase 3  0.378 +- 0.005
Total weight fraction 0.879 +- 0.035
------------------------
Spectrum 1: sample_5a.asc.extract
Chi-Square=1.47e+01, Reduced Chi-Square=4.13e-03, NDF=3547
Weight fraction Phase 1  0.376 +- 0.007
Weight fraction Phase 2  0.430 +- 0.009
Weight fraction Phase 3  0.139 +- 0.006
Total weight fraction 0.945 +- 0.013
------------------------
Spectrum 2: sample_S7a.asc.extract
Chi-Square=2.25e+01, Reduced Chi-Square=6.34e-03, NDF=3547
Weight fraction Phase 1  0.572 +- 0.009
Weight fraction Phase 2  0.000 +- 0.010
Weight fraction Phase 3  0.461 +- 0.007
Total weight fraction 1.033 +- 0.015
------------------------
Spectrum 3: sample_11a.asc.extract
Chi-Square=1.74e+01, Reduced Chi-Square=4.92e-03, NDF=3547
Weight fraction Phase 1  0.154 +- 0.008
Weight fraction Phase 2  0.310 +- 0.009
Weight fraction Phase 3  0.451 +- 0.007
Total weight fraction 0.915 +- 0.014
------------------------
Spectrum 4: sample_21a.asc.extract
Chi-Square=1.53e+01, Reduced Chi-Square=4.30e-03, NDF=3547
Weight fraction Phase 1  0.170 +- 0.008
Weight fraction Phase 2  0.406 +- 0.009
Weight fraction Phase 3  0.311 +- 0.006
Total weight fraction 0.888 +- 0.013
------------------------
Spectrum 5: sample_22a.asc.extract
Chi-Square=1.24e+01, Reduced Chi-Square=3.49e-03, NDF=3547
Weight fraction Phase 1  0.115 +- 0.007
Weight fraction Phase 2  0.332 +- 0.008
Weight fraction Phase 3  0.435 +- 0.006
Total weight fraction 0.882 +- 0.012
------------------------
Spectrum 6: CBZ_III.asc.extract
Chi-Square=7.33e-07, Reduced Chi-Square=2.07e-10, NDF=3547
Weight fraction Phase 1  1.000 +- 0.000
Weight fraction Phase 2  0.000 +- 0.000
Weight fraction Phase 3  0.000 +- 0.000
Total weight fraction 1.000 +- 0.000
------------------------
Spectrum 7: SAC.asc.extract
Chi-Square=4.84e-07, Reduced Chi-Square=1.36e-10, NDF=3547
Weight fraction Phase 1  0.000 +- 0.000
Weight fraction Phase 2  1.000 +- 0.000
Weight fraction Phase 3  0.000 +- 0.000
Total weight fraction 1.000 +- 0.000
------------------------
Spectrum 8: CBZ-SAC.asc.extract
Chi-Square=1.08e-06, Reduced Chi-Square=3.04e-10, NDF=3547
Weight fraction Phase 1  0.000 +- 0.000
Weight fraction Phase 2  0.000 +- 0.000
```

```
 Weight fraction Phase 3  1.000 +- 0.000
 Total weight fraction 1.000 +- 0.000
```

The above section reports the results of the fitting procedure applied to each input spectrum separately. Fit results include goodness-of-fit estimates (Chi Square and Reduced Chi Square), number of degrees of freedom (NDF), and the best fit estimates of the pure phases weight fractions. Finally, the rescaled weights are reported.

```
   -----------------------------------
 Phase 1: RMSE=0.095 RMSE'=0.117 R2=0.935 R2(NPP)=0.778 AKLD=0.130
      Calibration fit: Chi-Square=1.37e+03 intercept=-0.000+-0.000
slope=1.000+-0.000
      Limits estimates: LOD=0.03  LOQ=0.09
 NPP: Calibration fit: Chi-Square=4.55e+02 intercept=-0.177+-0.011
slope=1.215+-0.029
 NPP: Limits estimates: LOD=2.95  LOQ=8.95
  -----------------------------------
 Phase 2: RMSE=0.043 RMSE'=0.053 R2=0.991 R2(NPP)=0.995 AKLD=0.048
      Calibration fit: Chi-Square=2.54e+02 intercept=0.000+-0.000
slope=1.000+-0.000
      Limits estimates: LOD=0.04  LOQ=0.11
 NPP: Calibration fit: Chi-Square=9.35e+00 intercept=0.009+-0.009
slope=0.856+-0.021
 NPP: Limits estimates: LOD=3.50  LOQ=10.60
  -----------------------------------
 Phase 3: RMSE=0.072 RMSE'=0.067 R2=0.955 R2(NPP)=0.834 AKLD=0.050
      Calibration fit: Chi-Square=1.18e+03 intercept=0.000+-0.000
slope=1.000+-0.000
      Limits estimates: LOD=0.03  LOQ=0.09
 NPP: Calibration fit: Chi-Square=3.32e+02 intercept=0.160+-0.006
slope=0.623+-0.015
 NPP: Limits estimates: LOD=2.93  LOQ=8.89
  -----------------------------------
 Average total weight fraction: 0.924 +- 0.008
 Overall agreement: FOM=0.210 AKLD=0.228
```

The section above includes the comparison between the phase abundances estimated by RootProf with the true ones. The following quantities are reported to express the agreement between estimated and reference weight fractions: the Root Mean Square Error (RMSE), the same quantity calculated by including only samples for which the reference weight fraction of the given pure phase is not zero (RMSE'), the Squared Pearson's correlation coefficient (R2), and the same quantity calculated without considering pure phases [R2(NPP)]. Optimal RMSE and RMSE' values should be close to 0, optimal R2 values should be close to 1. The Chi-Square, intercept and slope of the best fit line determined by the calibrations plots are also reported. The slope parameter is used to obtain estimates of the Limit of Quantification (LOQ) and Limit of Detection (LOD). The same calculations are repeated by excluding pure phases from the fit of the calibration plots (NPP). A measure of the overall agreement between estimated and true weight fractions is given through the parameter FOM, which is the sum of the RMSE calculated over all the three pure phases, and the parameter Absolute value of the Kullback-Leibner distance (AKLD), also calculated. As a sum over the three pure phases. Note that AKLD has finite values only if reference weight fraction are different from zero.

Chapter 5

Supervised quantitative analysis of FT-IR spectra

Motivation

Apply supervised quantitative analysis by the MultiFit approach to FT-IR spectra filtered by the SNIP algorithm.

.

The command file

The list of commands is the following.

```
whichanalysis 4
calib 2
figpaper 1
dataType 4
range 450 4000
preprocess  0 3 15
file  sample_3a.asc.extract
referw 0 0.565 0.435
test
file  sample_5a.asc.extract
referw 0.5 0.5 0
test
file  sample_S7a.asc.extract
referw 0.5 0 0.5
test
file  sample_11a.asc.extract
referw 0.347 0.334 0.319
file  sample_21a.asc.extract
referw 0.263 0.482 0.255
file  sample_22a.asc.extract
referw 0.238 0.364 0.399
file  CBZ_III.asc.extract
referw 1 0 0
purephase
file  SAC.asc.extract
referw 0 1 0
purephase
file  CBZ-SAC.asc.extract
referw 0 0 1
purephase
```

The commands have been included in the demo file named *fileInputIRSupervised*. See the user guide for an explanation of their meaning. Samples (0,1,2) have been chosen as calibration set. They are constituted by binary samples, which are positioned in the expected positions in the PC1-PC2 score plot (see chapter 2).

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(“fileInputIRSupervised”)
```

or

```
Root> .> outputIRSupervised
.x RootProf_v15.C(“fileInputIRSupervised”)
.>
```

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 *outputIR*Supervised. 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 graphic window in Fig.1 shows the curve which is used to obtain the estimates of the calibration parameters. The known weight fractions of spectra belonging to the calibration set (black line) are fitted by the weight fractions calculated by quantitative analysis, performed by MultiFit or Unfolding techniques, according to the value of the *unfold* command (red, dashed line). In the fitting model the pure phases are multiplied by constants, which represent the free parameters of the fitting. The best fit is thus performed on NxP points, where N is the number of spectra in the calibration set, and P is the number of pure phases. In this case N=4, P=3, and there are 12 points.

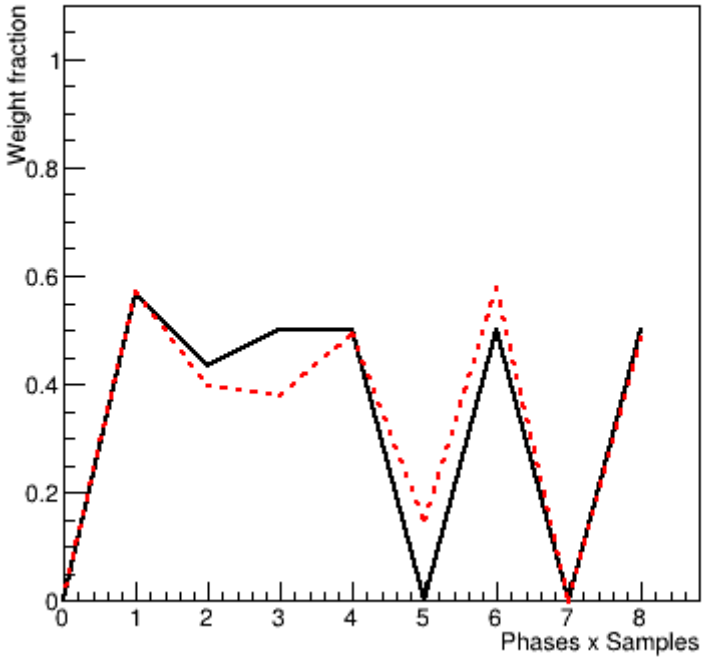


Fig.1 Calibration by LSQ

The output file

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

```
 Input from file: fileInputIRSupervised
--------------------------------
  whichanalysis 4

  calib 2

  figpaper 1

  dataType 4

  range 450 4000

  preprocess  0 3 15

  file  sample_3a.asc.extract

  referw 0 0.565 0.435

  test

  file  sample_5a.asc.extract

  referw 0.5 0.5 0

  test

  file  sample_S7a.asc.extract
```

```
  referw 0.5 0 0.5

  test

  file  sample_11a.asc.extract

  referw 0.347 0.334 0.319

  file  sample_21a.asc.extract

  referw 0.263 0.482 0.255

  file  sample_22a.asc.extract

  referw 0.238 0.364 0.399

  file  CBZ_III.asc.extract

  referw 1 0 0

  purephase

  file  SAC.asc.extract

  referw 0 1 0

  purephase

  file  CBZ-SAC.asc.extract

  referw 0 0 1

  purephase
```

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

```
Starting pure phase scale calibration

Reading input files:
 ----------------------
Sample 0 -> file sample_3a.asc.extract
         Found 3551  points
Sample 1 -> file sample_5a.asc.extract
         Found 3551  points
Sample 2 -> file sample_S7a.asc.extract
         Found 3551  points
Sample 3 -> file sample_11a.asc.extract
         Found 3551  points
Sample 4 -> file sample_21a.asc.extract
         Found 3551  points
Sample 5 -> file sample_22a.asc.extract
         Found 3551  points
Sample 6 -> file CBZ_III.asc.extract
         Found 3551  points
Sample 7 -> file SAC.asc.extract
         Found 3551  points
Sample 8 -> file CBZ-SAC.asc.extract
```

```
         Found 3551  points
```

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

```
FIT RESULTS:
 ------------------------
 Spectrum 0: sample_3a.asc.extract
 Chi-Square=1.09e+01, Reduced Chi-Square=3.08e-03, NDF=3547
 Weight fraction Phase 1  0.006 +- 0.034
 Weight fraction Phase 2  0.496 +- 0.007
 Weight fraction Phase 3  0.378 +- 0.005
 Total weight fraction 0.879 +- 0.035
 ------------------------
 Spectrum 1: sample_5a.asc.extract
 Chi-Square=1.47e+01, Reduced Chi-Square=4.13e-03, NDF=3547
 Weight fraction Phase 1  0.376 +- 0.007
 Weight fraction Phase 2  0.430 +- 0.009
 Weight fraction Phase 3  0.139 +- 0.006
 Total weight fraction 0.945 +- 0.013
 ------------------------
 Spectrum 2: sample_S7a.asc.extract
 Chi-Square=2.25e+01, Reduced Chi-Square=6.34e-03, NDF=3547
 Weight fraction Phase 1  0.572 +- 0.009
 Weight fraction Phase 2  0.000 +- 0.010
 Weight fraction Phase 3  0.461 +- 0.007
 Total weight fraction 1.033 +- 0.015
```

The section above reports the results of a first run of quantitative analysis, performed without special rescaling pure phase spectra on the spectra belonging to the calibration set. The obtained weight fractions represent the starting point for the calibration procedure.

```
Calibration fit results:
 ------------------------------
 Chi-Square=0.044  Reduced Chi-Square=0.0073 NDF=6
 Phase 0: Best fit coefficient= 1.00 +-  0.17
 Phase 1: Best fit coefficient= 0.88 +-  0.15
 Phase 2: Best fit coefficient= 0.96 +-  0.17
```

The section above report the result of the best fit applied on the true weight fractions of the spectra belonging to the calibration set (see Fig.1). The coefficient so obtained are the scaling factors to be applied to pure phase spectra.

```
FIT RESULTS:
 ------------------------
 Spectrum 0: sample_3a.asc.extract
 Chi-Square=1.09e+01, Reduced Chi-Square=3.08e-03, NDF=3547
 Weight fraction Phase 1  0.006 +- 0.033
 Weight fraction Phase 2  0.563 +- 0.008
 Weight fraction Phase 3  0.394 +- 0.006
 Total weight fraction 0.962 +- 0.035
 ------------------------
 Spectrum 1: sample_5a.asc.extract
 Chi-Square=1.47e+01, Reduced Chi-Square=4.13e-03, NDF=3547
 Weight fraction Phase 1  0.376 +- 0.007
 Weight fraction Phase 2  0.489 +- 0.010
```

```
 Weight fraction Phase 3  0.145 +- 0.006
 Total weight fraction 1.009 +- 0.014
 ------------------------
 Spectrum 2: sample_S7a.asc.extract
 Chi-Square=2.25e+01, Reduced Chi-Square=6.34e-03, NDF=3547
 Weight fraction Phase 1  0.572 +- 0.009
 Weight fraction Phase 2  0.000 +- 0.012
 Weight fraction Phase 3  0.481 +- 0.007
 Total weight fraction 1.053 +- 0.016
 ------------------------
 Spectrum 3: sample_11a.asc.extract
 Chi-Square=1.74e+01, Reduced Chi-Square=4.92e-03, NDF=3547
 Weight fraction Phase 1  0.154 +- 0.008
 Weight fraction Phase 2  0.352 +- 0.011
 Weight fraction Phase 3  0.470 +- 0.007
 Total weight fraction 0.976 +- 0.015
 ------------------------
 Spectrum 4: sample_21a.asc.extract
 Chi-Square=1.53e+01, Reduced Chi-Square=4.30e-03, NDF=3547
 Weight fraction Phase 1  0.170 +- 0.008
 Weight fraction Phase 2  0.462 +- 0.010
 Weight fraction Phase 3  0.324 +- 0.007
 Total weight fraction 0.956 +- 0.014
 ------------------------
 Spectrum 5: sample_22a.asc.extract
 Chi-Square=1.24e+01, Reduced Chi-Square=3.49e-03, NDF=3547
 Weight fraction Phase 1  0.115 +- 0.007
 Weight fraction Phase 2  0.377 +- 0.009
 Weight fraction Phase 3  0.454 +- 0.006
 Total weight fraction 0.946 +- 0.013
 ------------------------
 Spectrum 6: CBZ_III.asc.extract
 Chi-Square=1.74e-07, Reduced Chi-Square=4.91e-11, NDF=3547
 Weight fraction Phase 1  1.000 +- 0.000
 Weight fraction Phase 2  0.000 +- 0.000
 Weight fraction Phase 3  0.000 +- 0.000
 Total weight fraction 1.000 +- 0.000
 ------------------------
 Spectrum 7: SAC.asc.extract
 Chi-Square=2.20e-06, Reduced Chi-Square=6.21e-10, NDF=3547
 Weight fraction Phase 1  0.000 +- 0.000
 Weight fraction Phase 2  1.000 +- 0.000
 Weight fraction Phase 3  0.000 +- 0.000
 Total weight fraction 1.000 +- 0.000
 ------------------------
 Spectrum 8: CBZ-SAC.asc.extract
 Chi-Square=5.62e-07, Reduced Chi-Square=1.58e-10, NDF=3547
 Weight fraction Phase 1  0.000 +- 0.000
 Weight fraction Phase 2  0.000 +- 0.000
 Weight fraction Phase 3  1.000 +- 0.000
 Total weight fraction 1.000 +- 0.000
```

The section above reports the results of a second run of quantitative analysis, performed by rescaling pure phase spectra with the previously determined coefficient.

```
  ------------------------------------
 Phase 1: RMSE=0.095 RMSE'=0.117 R2=0.935 R2(NPP)=0.778 AKLD=0.130
      Calibration fit: Chi-Square=1.37e+03 intercept=-0.000+-0.000
slope=1.000+-0.000
```

```
      Limits estimates: LOD=0.03  LOQ=0.08
 NPP: Calibration fit: Chi-Square=4.55e+02 intercept=-0.177+-0.011
slope=1.214+-0.029
 NPP: Limits estimates: LOD=2.95  LOQ=8.95
  ----------------------------------
 Phase 2: RMSE=0.011 RMSE'=0.013 R2=0.999 R2(NPP)=0.995 AKLD=0.011
      Calibration fit: Chi-Square=1.12e+01 intercept=0.000+-0.000
slope=1.000+-0.000
      Limits estimates: LOD=0.01  LOQ=0.04
 NPP: Calibration fit: Chi-Square=9.31e+00 intercept=0.011+-0.010
slope=0.972+-0.024
 NPP: Limits estimates: LOD=3.50  LOQ=10.60
  ----------------------------------
 Phase 3: RMSE=0.077 RMSE'=0.074 R2=0.953 R2(NPP)=0.834 AKLD=0.050
      Calibration fit: Chi-Square=1.25e+03 intercept=0.000+-0.000
slope=1.000+-0.000
      Limits estimates: LOD=0.00  LOQ=0.01
 NPP: Calibration fit: Chi-Square=3.32e+02 intercept=0.167+-0.006
slope=0.649+-0.016
 NPP: Limits estimates: LOD=2.93  LOQ=8.89
  ----------------------------------
 Average total weight fraction: 0.984 +- 0.008
 Overall agreement: FOM=0.183 AKLD=0.191
```

The above section includes the comparison between the phase abundances estimated by the calibrated RootProf and the reference ones. The quantities reported are the same as those described in Chapter 2 of the Quantitative tutorial. Note that the overall AKLD (0.191) is lower than that obtained by the unsupervised MultiFit (0.228), hence the performances of the supervised quantitative analysis are better those of the unsupervised one.

Notes

- By considering results reported on Table 1, it results that mixtures with non-zero phase abundances (3,4,5) are not particularly suitable for calibration purposes. Binary mixtures (0,1,2) have instead higher performances. This fact reflects what seen in qualitative analysis, where data points (3,4,5) were shifted from expected positions in the ternary diagram in the PC1-PC2 score plot.
- Pure phase spectra (6,7,8) give very poor calibration performances. A calibration set constituted by pure phases only should not be used. It does not add any new information with respect to the unsupervised case.
- The calibration procedure applies to MultiFit or Unfolding techniques, depending on the settings of the *unfold* command.
- The calibration parameters can be view as a multiplicative rescaling of the pure phase spectra.
- If the pre-processing calibration is run on FT-IR spectra (command *calib* 2 commented in *fileInputIRSupervised*), pre-processing 0 3 80 0 is chosen, and L1=0, L2= 3 L3=80 L4=0 are selected as best pre-processing types (by looking at projection plots). However the analysis does not produce a better results that that shown here, probably because the test set (mixtures 0,1,2) is not reliable for such an analysis.

Table 1: Results of the calibration procedure by using different compositions of the calibration set.

Samples in Calibration Set	**Calibration Parameters**	**Overall AKLD**
--	1.00, 1.00, 1.00	0.228
0,1,2	1.00, 0.88, 0.96	0.191
3,4,5	1.00, 1.69, 2.35	0.630
0,1,2,3,4,5	1.00, 0.98, 1.23	0.236
6,7,8	1.00, 1.00, 1.00	0.228
0,1,2,6,7,8	1.00, 0.96, 0.99	0.213
3,4,5,6,7,8	1.00, 1.02, 1.13	0.239
0,1,2,3,4,5,6,7,8	1.00, 0.98, 1.08	0.200