RootProf: principles, workflow and new graphical user interface

Rocco Caliandro

CNR – Institute of Crystallography – Bari

E-mail: rocco.caliandro@ic.cnr.it
Multi-technique *in situ* experiments

- New generation X-ray sources
- Multi-probe experimental setups
- Sensitive and fast detectors

Monitor structural dynamics

Raman

XRPD
Multiple measurements

Same technique
Different times

Same time
Different techniques

Extract relevant information

Combine information

Qualitative and quantitative analysis

Statistical methods

Covariance analysis

Multivariate analysis

Simultaneous analysis of more than one variable

- Analysis of interdependence relationships among variables
  - Unsupervised (clustering, PCA)
- Analysis of dependence dependent variables are explained or predicted by other variables (PLS, Least squares)
New data analysis methods

Data Matrices
- All profiles processed at the same time
- Probe-independent processing
- Adapted variants of existing methods

RootProf

Prompt location of active atoms

Multivariate analysis

Fast extraction of structural kinetics

Combined information from different techniques

Raman

XRPD

RootProf

Combined information from different techniques
RootProf

Availability: www.ba.ic.cnr.it/softwareic/

Documentation: www.ba.ic.cnr.it/softwareic/rootprof

Forum: groups.google.com/forum/#!forum/RootProf

RootProf: software for multivariate analysis of unidimensional profiles

Rocco Caliandro* and Danilo Benny Belviso

Received 16 January 2014
Online help

www.ba.ic.cnr.it/softwareic/rootprof

Overview

User guide

Download/Install/Run

Updates/form

Tutorials

Qualitative analysis
Unsupervised quantitative analysis
Supervised quantitative analysis
Covariance analysis
Size analysis
Analysis of FT-IR spectra
Analysis of XAS spectra
Analysis of SAXS profiles
Analysis of single-crystal patterns
Developed at CERN for the data analysis of High-Energy Physics experiments

Freely available from https://root.cern.ch/releases

Version 6 (Linux, Mac) → RootProf v15
Version 5 (Windows) → RootProf v14

Besides the specific RootProf documentation, users can take advantage of the documentation and discussion of the wide community of ROOT users
RootProf

Root script, driven by the user through a GUI or a command file

Run ROOT

Run RootProf GUI

Prepare command file & Run RootProf from GUI

Inspect text output and editable graphic windows

RootProf can be modified to introduce new developments (C++ code)
Workflow

www.ba.ic.cnr.it/softwareic/rootprof
Step 1: Input
### CBZ-SAC mixtures

The CBZ-SAC mixtures are visualized with three components: CBZ I, SAC, and CBZ-SAC. The table below provides the composition of each sample along with the corresponding file names.

<table>
<thead>
<tr>
<th>Sample n.</th>
<th>CBZ III</th>
<th>SAC</th>
<th>CBZ-SAC</th>
<th>File name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.565</td>
<td>0.435</td>
<td>Rocco_S3_mac.txt</td>
</tr>
<tr>
<td>1</td>
<td>0.500</td>
<td>0.500</td>
<td>0</td>
<td>Rocco_S5_mac.txt</td>
</tr>
<tr>
<td>2</td>
<td>0.500</td>
<td>0</td>
<td>0.500</td>
<td>Rocco_S7_Como.txt</td>
</tr>
<tr>
<td>3</td>
<td>0.347</td>
<td>0.334</td>
<td>0.319</td>
<td>Rocco_S11_mac.txt</td>
</tr>
<tr>
<td>4</td>
<td>0.263</td>
<td>0.482</td>
<td>0.255</td>
<td>Rocco_S21_mac.txt</td>
</tr>
<tr>
<td>5</td>
<td>0.238</td>
<td>0.364</td>
<td>0.399</td>
<td>Rocco_S22_mac.txt</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Rocco_CBZ_III_nomac.txt</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Rocco_SAC_pura_nomac.txt</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Rocco_CBZSAC_90511_n.txt</td>
</tr>
</tbody>
</table>
Measurements

XRPD

FT-IR
Step 2: Pre-processing

- Enhances the performances of multivariate techniques
- Depends on the type of data
- Produces rescaled profiles from original ones: $y(i) \rightarrow \hat{y}(i)$
Background subtraction

Enhances:
- the signal in each profile
- differences among profiles

\[ \hat{y}'(i) = \hat{y}(i) - b(i) \]

Background

Can be estimated by:

Sensitive Nonlinear Iterative Peak (SNIP) clipping algorithm

\[ b(i) = \hat{y}_{N_{\text{clip}}}(i) \]

\[ \hat{y}_1(i) = \hat{y}(i) \]

\[ \hat{y}_2(i) = \text{Min} \left[ \hat{y}_1(i), \frac{\hat{y}_1(i+2) + \hat{y}_1(i-2)}{2} \right] \]

\[ \hat{y}_p(i) = \text{Min} \left[ \hat{y}_{p-1}(i), \frac{\hat{y}_{p-1}(i+p) + \hat{y}_{p-1}(i-p)}{2} \right] \]

\( N_{\text{clip}} \): width of clipping window & number of iterations
Pre-processing on XRPD

\[ \hat{y}'(i) = \hat{y}(i) - b(i) \]

\[ \hat{y}(i) = \frac{y(i)}{\sum_{j=1}^{N} y(j)} \]
Pre-processing on FTIR

FTIR, SNV

same spectrum with different amount of KBr pellet

\[
\hat{y}(i) = \frac{y(i) - \langle y(i) \rangle}{\sigma_y}
\]

a) Before  \rightarrow  b) After

FTIR, BS (Nclip=10)
Step 3: Task

Qualitative analysis
Qualitative analysis by PCA

XRPD

FTIR

Scores

Loadings
Step 3: Task

Quantitative analysis
The MultiFit approach

Fit model:

\[ f(i) = \sum_{j=1}^{M} p_j \hat{f}_j(i) \]

\[ \hat{f}_j(i) \] Pre-processed spectrum of j-th pure phase

M = number of pure phases in the sample

For IR

\[ w_j = p_j \]

or

\[ w_j = \frac{p_j}{\sum_{t=1}^{M} p_t} \]

For XRPD

\[ w_j = \frac{p_j}{\sum_{t=1}^{M} \frac{p_t}{\mu_i^*}} \]

Deviation from 1 indicates phases not included in the fit or amorphous content

Sum = \[ \sum_{t=1}^{M} p_t \]
MultiFit

XRPD

FTIR

Experimental (pre-processed) pattern

Calculated (best fitted) profile $f(x)$
MultiFit vs Rietveld

- Can be applied to profiles from any technique
- Can be applied if pure-phases profiles are measured
- A priori information of mixture composition not strictly necessary
  
  ![](Sum = \sum_{t=1}^{M} w_t)

- Specific for XRPD
- Can be applied if pure-phases crystal structures are known
- A priori information of mixture composition needed

free parameter

set to 1
PCA filtering

Profiles reconstructed by using 3 PCs ➔ Effect on quantitative analysis

- The dimensionality reduction minimize the experimental error
- Main features of the original spectra preserved

Experimental (pre-processed) pattern

Calculated (best fitted) profile f(x)
Quantitative analysis by unfolding

MultiFit

\[ \hat{y}(i) = \sum_{j=1}^{M} p_j \hat{f}_j(i) \]

Pre-processed profile of the j-th pure phase

M = number of pure phases in the sample

Unfolding

\[ p_j = \sum_{i=1}^{N} \hat{f}_j(i) \hat{y}(i) \]

Response matrix formed by the pure phase profiles

N = number of 2theta values

Less accurate, but faster and less dependent on pre-processing

Iterative LSQ

Iterative Gold deconvolution
Unfolding vs Multifit

AKLD = 0.121  
Time = 0.4 s

AKLD = 0.110  
Time = 54.4 s

Kullback-Leibler distance

\[ AKLD = \sum_{i=1}^{K} w'_i \ln \frac{w'_i}{w_i} \]
Supervised quantitative analysis

Subset of samples, with known weight fractions, used for calibration

Automatic choice of best pre-processing option

<table>
<thead>
<tr>
<th>Pre-Processing type</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Modifications</td>
<td>no-modification</td>
<td>Smoothing</td>
<td>Deconvolution</td>
<td>Log10</td>
<td>Powering by 0.8</td>
<td>Powering by 1.2</td>
</tr>
<tr>
<td>Level 2: Rescaling</td>
<td>No-rescaling</td>
<td>Mean centering</td>
<td>Normalization</td>
<td>Standard Normal Variate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3: Background subtraction</td>
<td>No-background subtraction</td>
<td>Clipping window=16</td>
<td>Clipping window=20</td>
<td>Clipping window=40</td>
<td>Clipping window=60</td>
<td>Clipping window=100</td>
</tr>
<tr>
<td>Level 4: Filtering</td>
<td>No-filtering</td>
<td>Multiplicative Scatter Correction</td>
<td>Multiplicative Scatter Correction, all profiles</td>
<td>Principal component filtering</td>
<td>Principal component filtering, all profiles</td>
<td></td>
</tr>
</tbody>
</table>

Best-fit determination of pure phase rescaling

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

<table>
<thead>
<tr>
<th>Samples in Calibration Set</th>
<th>Calibration Parameters</th>
<th>Overall AKLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>--</td>
<td>1.00, 1.00, 1.00</td>
<td>0.108</td>
</tr>
<tr>
<td>0,1,2</td>
<td>1.00, 0.88, 0.98</td>
<td>0.104</td>
</tr>
<tr>
<td>3,4,5</td>
<td>1.00, 1.02, 1.16</td>
<td>0.094</td>
</tr>
<tr>
<td>0,1,2,3,4,5</td>
<td>1.00, 0.93, 1.05</td>
<td>0.097</td>
</tr>
<tr>
<td>6,7,8</td>
<td>1.00, 1.00, 1.00</td>
<td>0.108</td>
</tr>
<tr>
<td>0,1,2,6,7,8</td>
<td>1.00, 0.87, 0.97</td>
<td>0.104</td>
</tr>
<tr>
<td>3,4,5,6,7,8</td>
<td>1.00, 1.02, 1.16</td>
<td>0.094</td>
</tr>
<tr>
<td>0,1,2,3,4,5,6,7,8</td>
<td>1.00, 0.93, 1.05</td>
<td>0.097</td>
</tr>
</tbody>
</table>
Step 3: Task

Covariance analysis
Covariance analysis

Averages calculated over the samples

\[
COV_{AB}(i, j) = \frac{\sum_{i=1}^{K}(\hat{y}_i(i) - \langle \hat{y}(i) \rangle_A)(\hat{y}_j(j) - \langle \hat{y}(j) \rangle_B)}{K - 1}
\]

\(\hat{y}_i\) and \(\hat{y}_j\) are the predicted values for samples \(i\) and \(j\) respectively, \(\langle \hat{y}(i) \rangle_A\) and \(\langle \hat{y}(j) \rangle_B\) are the average predicted values for \(i\) and \(j\) respectively.

\(K\) is the number of samples.

\(A=\text{XRPD}\) and \(B=\text{FTIR}\)

CBZ-SAC
SAC
CBZ III

3460 3500
Covariance matrix interpretation

$$\text{Cov}(x, y) = \sum (x_i - \bar{x})(y_j - \bar{y})$$

Crystal phase | XRPD signal (2θ) | FTIR signal (cm\(^{-1}\)) | Correlation |
--- | --- | --- | --- |
CBZ-SAC | 14.0 | 3460 | Positive |
| | | 3500 | Negative |
SAC | 16.0, 19.5 | 3460 | Positive |
| | | 3500 | Positive |
CBZ III | 13.2, 15.5 | 3460 | Negative |
| | | 3500 | Positive |

3460 and 3500 cm\(^{-1}\) FTIR peaks very sensitive for discrimination of CBZ III, SAC and CBZ-SAC
Step 4: Output & Step 5: Run
Round-robin mixtures

AKLD = 0.021

comparable results obtained by the Rietveld method
Polymorphism of glycine

γ-glycine

α-glycine
Classification of XRPD patterns

Mahalanobis distance

\[ dist = \sqrt{(u_j - u_i)^T COV (u_j - u_i)} \]

\( u_i = \text{sample mean of group i} \)

95% confidence ellipses and P-values

P-value = 5.2 \times 10^{-10}
Characterization of zucchini fruits

Biochar added to soil: 0% (B0), 2% (B2), 4% (B4), 8% (B8)

Mychorrhizial product addition: B0M, B2M

Crystalline Cellulose

I_α I_β

Amylose

15.5°
PCA analysis of fruits

Sensitive to biochar addiction (PC1)

Recognize mycorrhizial product addition (PC2)

PC1: relative abundance of crystalline cellulose

PC2: relative abundance of amylose (15.5°)
Structural characterization of halide perovskites

Temperature varied in situ 300 → 400 → 300 °C
Tetragonal-to-cubic phase transition monitored

R. Caliandro, D. Altamura, B.D. Belviso, A. Rizzo, S. Masi, C. Giannini
J. Appl. Cryst. 2019, 52

MAPbI$_3$ and PbI$_2$-MAI-DMSO peaks anticorrelated

Covariance matrix calculated by RootProf
Trends in time and space domains

**Scores**

**PCA on XRPD**

- PC1 scores
- PC2 scores
- PC3 scores

**PCA on PDF**

- PC1 scores
- PC2 scores
- PC3 scores

**Loadings**

- PC1 loadings
- PC2 loadings
- PC3 loadings

**Local disorder from misaligned octahedra**

**Anomalous contribution from nearest inter-octahedral distances**

**PC1**

**PC2**
Fitting of individual profiles

Automatic fitting of individual PDF profiles

 Weight fraction

 Thermal factor I2

 Dihedral angle I2-Pb-Pb-I2

 N quadrupole moment

 MAPbI$_3$ Pbl$_2$-MAI-DMSO

 MAPbI$_3$ Pbl$_2$-MAI-DMSO
Reversible generation, under illumination, of electron paramagnetic resonance signal from MAPbI$_3$ perovskite polycrystalline powder

Sample illuminated
Light on for 90 min (profiles 1-44)
Light off for 90 min (profiles 45-90)
Space-dependent PCA

PC1: structural variations induced by the X-ray illumination (lattice distortion, thermal atomic fluctuations…)

PC2: light-related trend, similar to that observed for the EPR signal

Higher light-induced effect for the shorter interatomic distances (R<3 Å), consistent with the range of Pb-O distances (2.25 Å)

Generation of paramagnetic Pb$^{3+}$ defects under illumination, induced by the presence of Pb-O defects that may trap photogenerated holes
References

- Palin et al. Rational design of the solid-state synthesis of materials based on poly-aromatic molecular complexes (2016) *CrystEngComm* 31
- Rizzuti et al. A combined approach for characterisation of fresh and brined vine leaves by X-ray powder diffraction, NMR spectroscopy and direct infusion high resolution mass spectrometry (2013) *Food Chemistry* 141