RootProf: principles, workflow and new graphical user interface

IC CI
CNR Istituto di Cristallografia

Rocco Caliandro

CNR – Institute of Crystallography – Bari

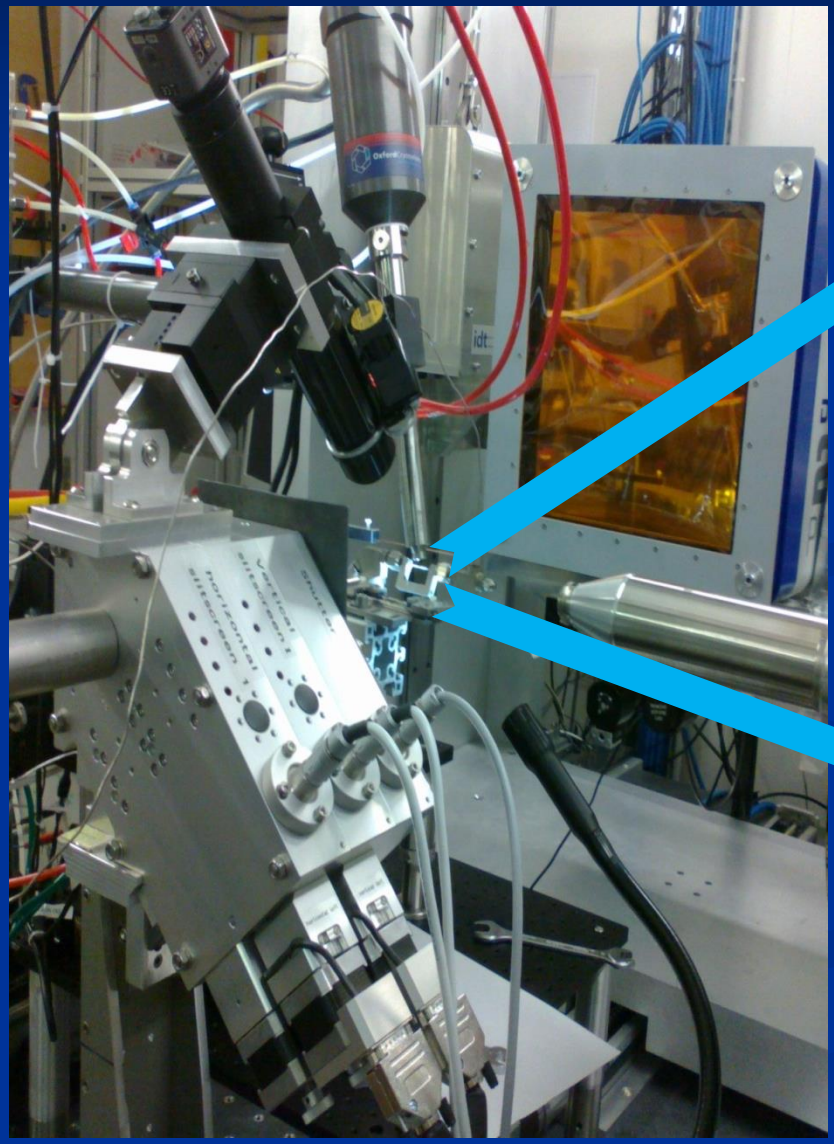
E-mail: rocco.caliandro@ic.cnr.it

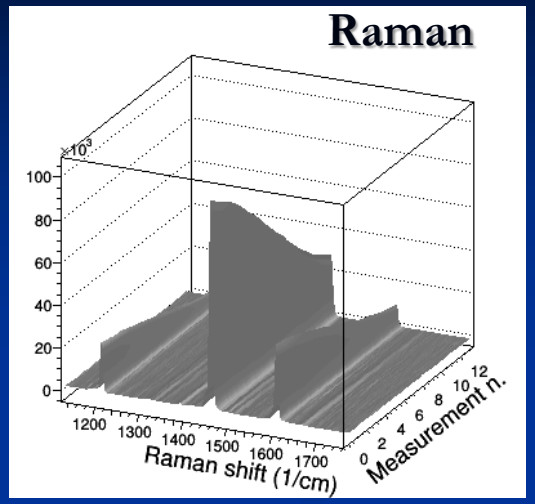
Multi-technique *in situ* experiments

New generation
X-ray sources

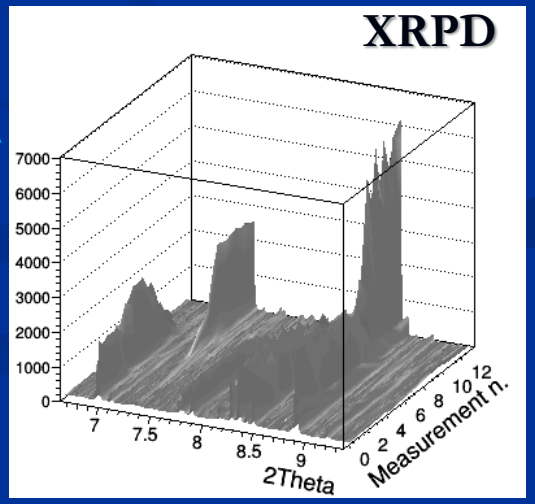
Multi-probe
experimental setups

Sensitive and fast
detectors





Monitor
structural dynamics



Multiple measurements

Same technique
Different times

Same time
Different techniques

Statistical methods

Extract relevant information

Combine information

Qualitative and quantitative analysis

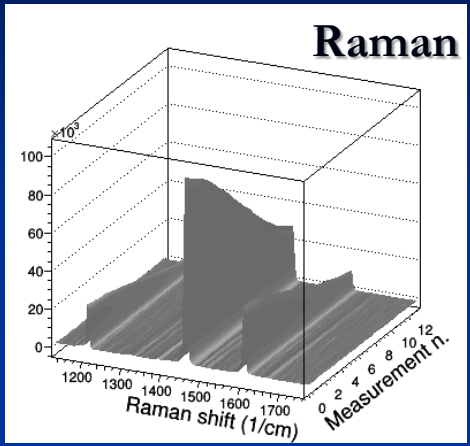
Covariance analysis

Multivariate analysis

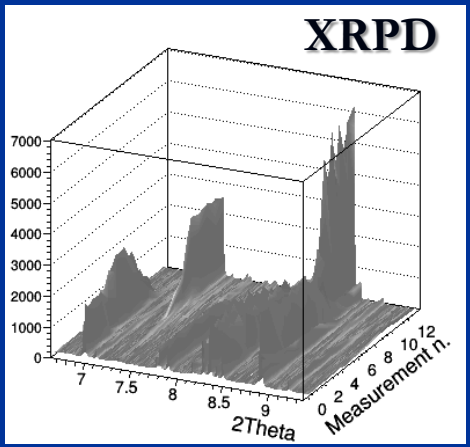
Simultaneous analysis of more than one variable

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

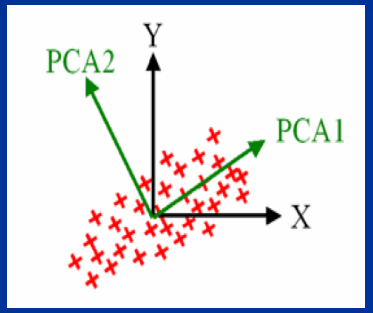
New data analysis methods



Data Matrices

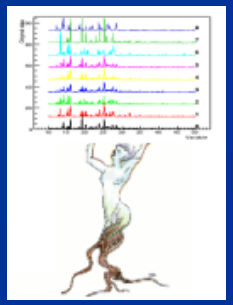


Multivariate analysis

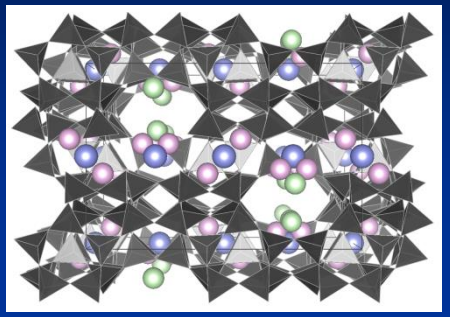


- ✓ All profiles processed at the same time
- ✓ Probe-independent processing
- ✓ Adapted variants of existing methods

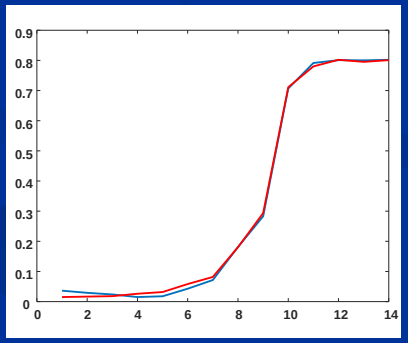
RootProf



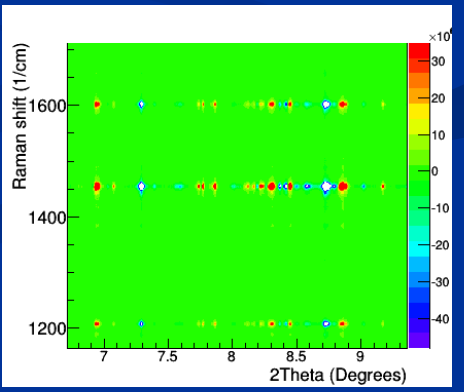
Prompt location of active atoms



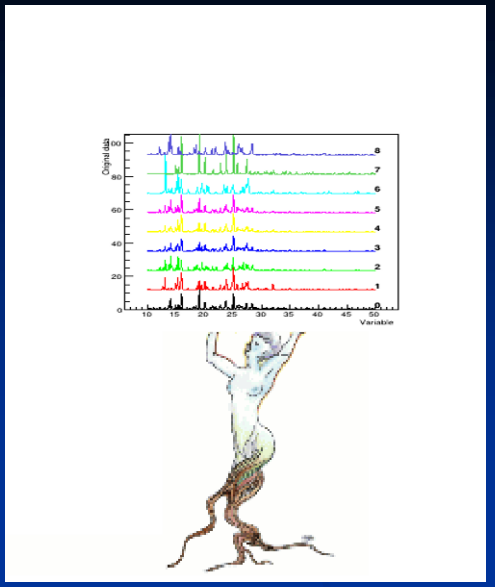
Fast extraction of structural kinetics



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

Journal of
Applied
Crystallography
ISSN 1600-5767

Received 16 January 2014

RootProf: software for multivariate analysis of unidimensional profiles

Rocco Caliandro* and Danilo Benny Belviso

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

ROOT

Developed at CERN for the data analysis of High-Energy Physics experiments

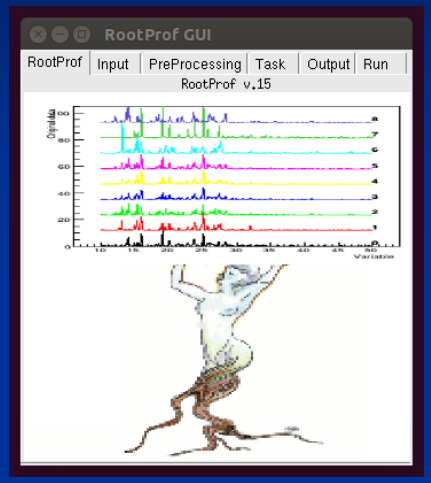
Freely available from https://root.cern.ch/releases

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

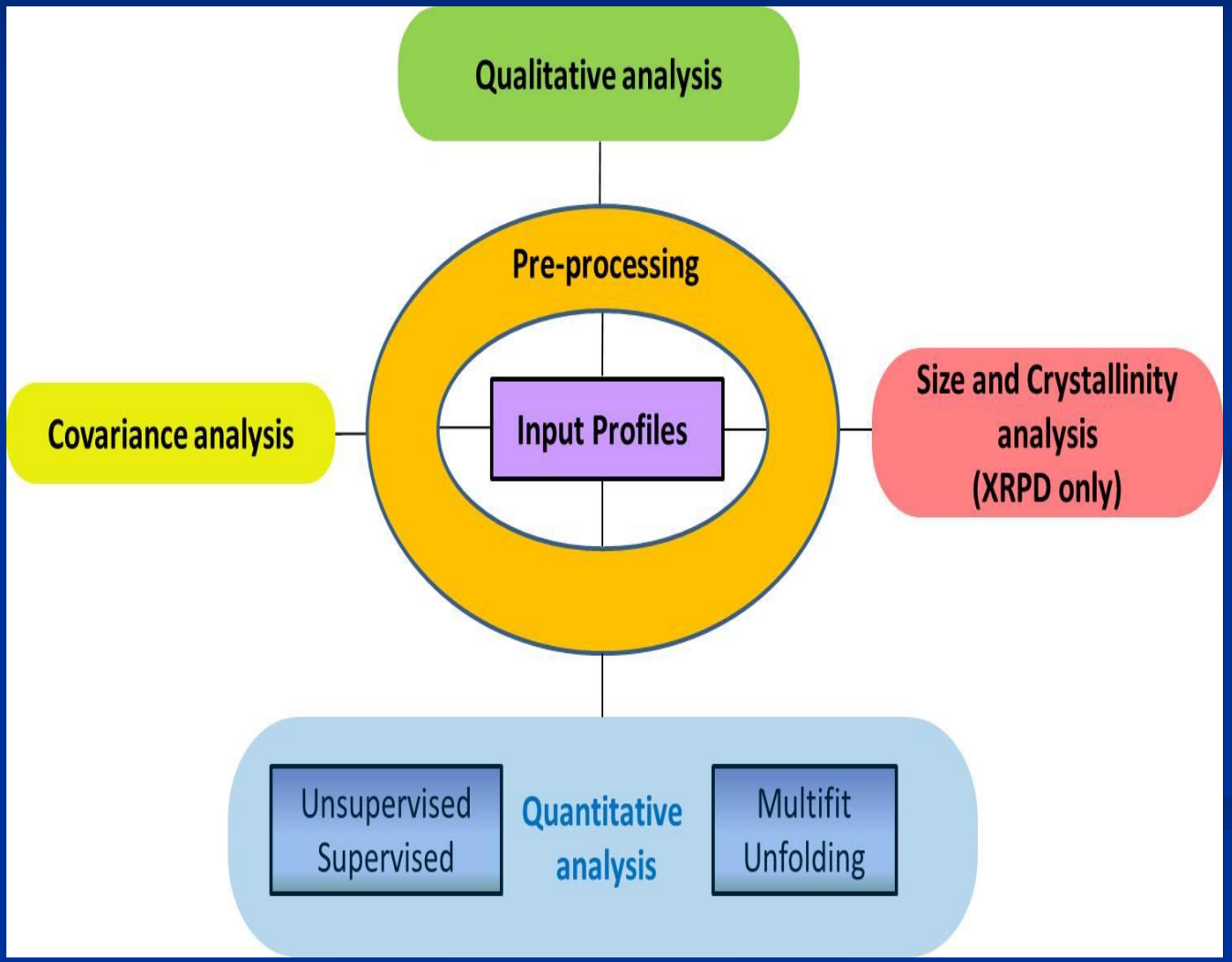
Prepare command file & Run RootProf from GUI

Inspect text output and editable graphic windows

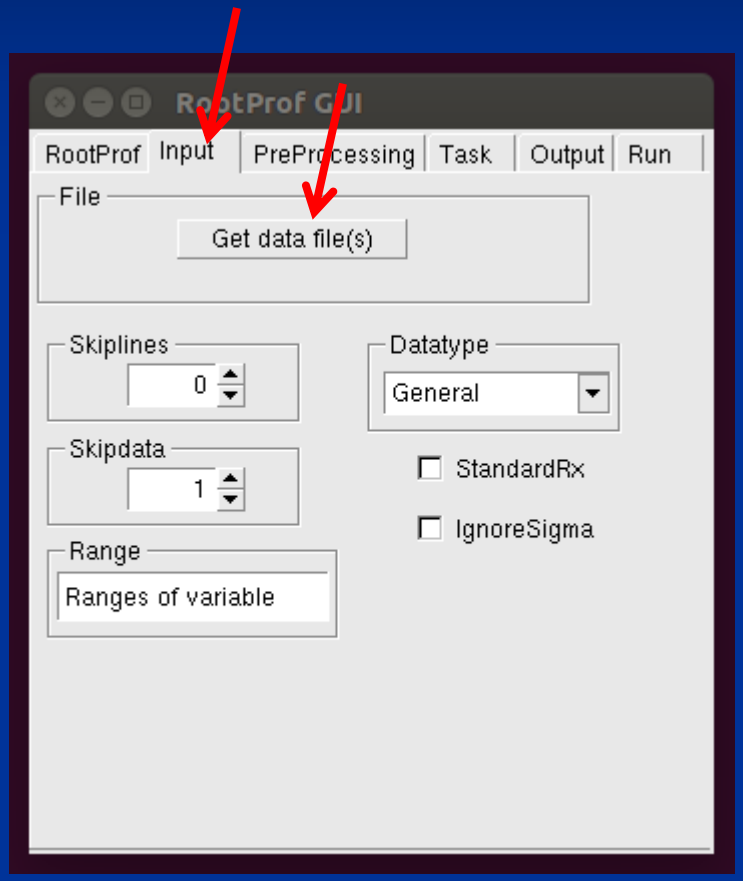
RootProf can be modifyed to introduce new developments (C++ code)

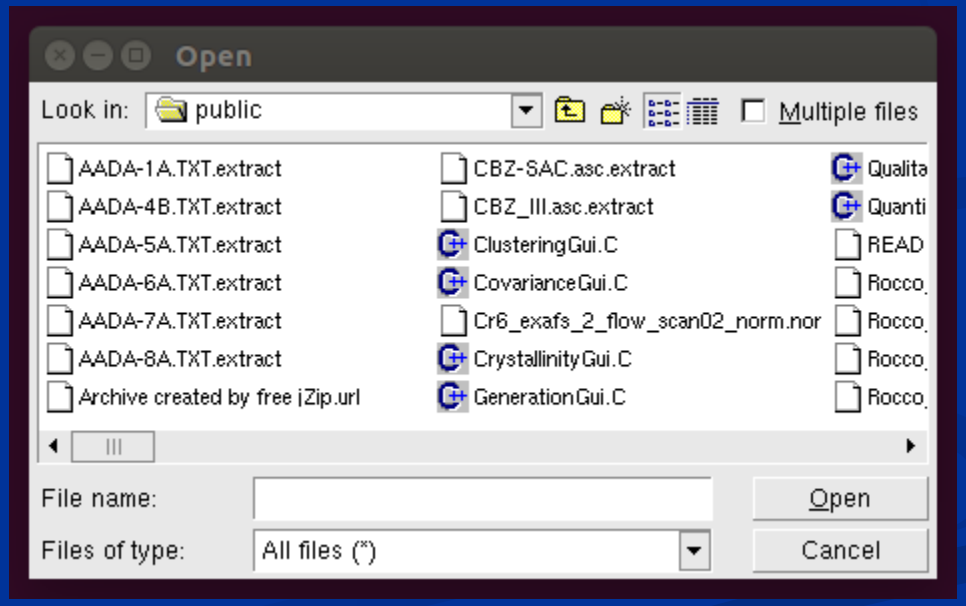
Workflow

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



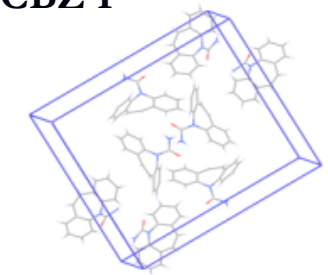
Step 1: Input


RootProf GUI
RootProf
Input
PreProcessing
Task
Output
Run
File
Get data file(s)
Skiplines
0
Skipdata
1
Range
Ranges of variable
Datatype
General
StandardRx
IgnoreSigma

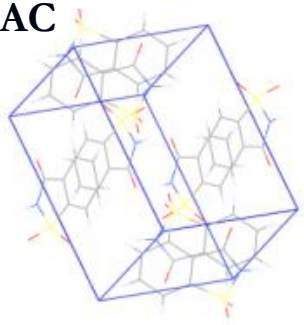

Open
Look in:
public
Multiple files
AADA-1A.TXT.extract
AADA-4B.TXT.extract
AADA-5A.TXT.extract
AADA-6A.TXT.extract
AADA-7A.TXT.extract
AADA-8A.TXT.extract
Archive created by free jZip.url
CBZ-SAC.asc.extract
CBZ_III.asc.extract
ClusteringGui.C
CovarianceGui.C
Cr6_exafs_2_flow_scan02_norm.nor
CrystallinityGui.C
GenerationGui.C
Qualita
Quanti
READ
Rocco
Rocco
Rocco
Rocco
File name:
Open
Files of type:
All files (*)
Cancel

CBZ-SAC mixtures

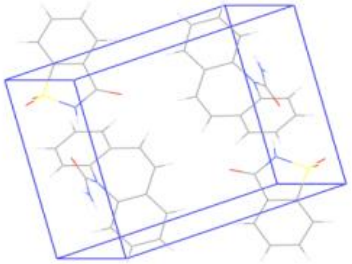
CBZ I



SAC



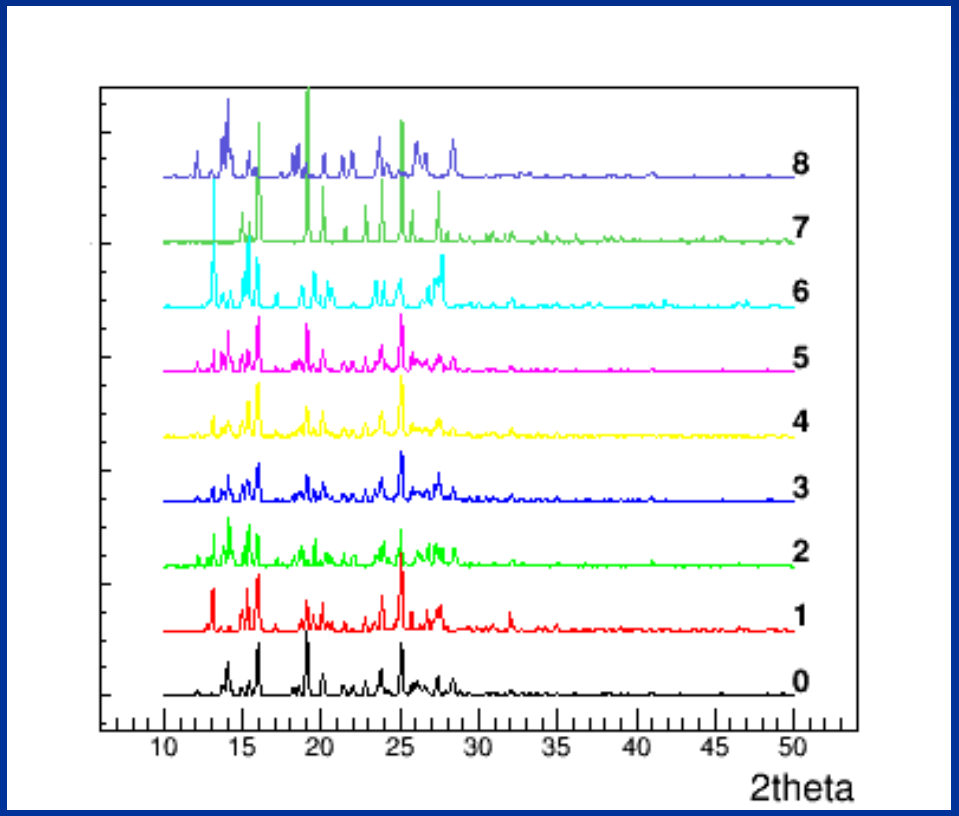
CBZ-SAC



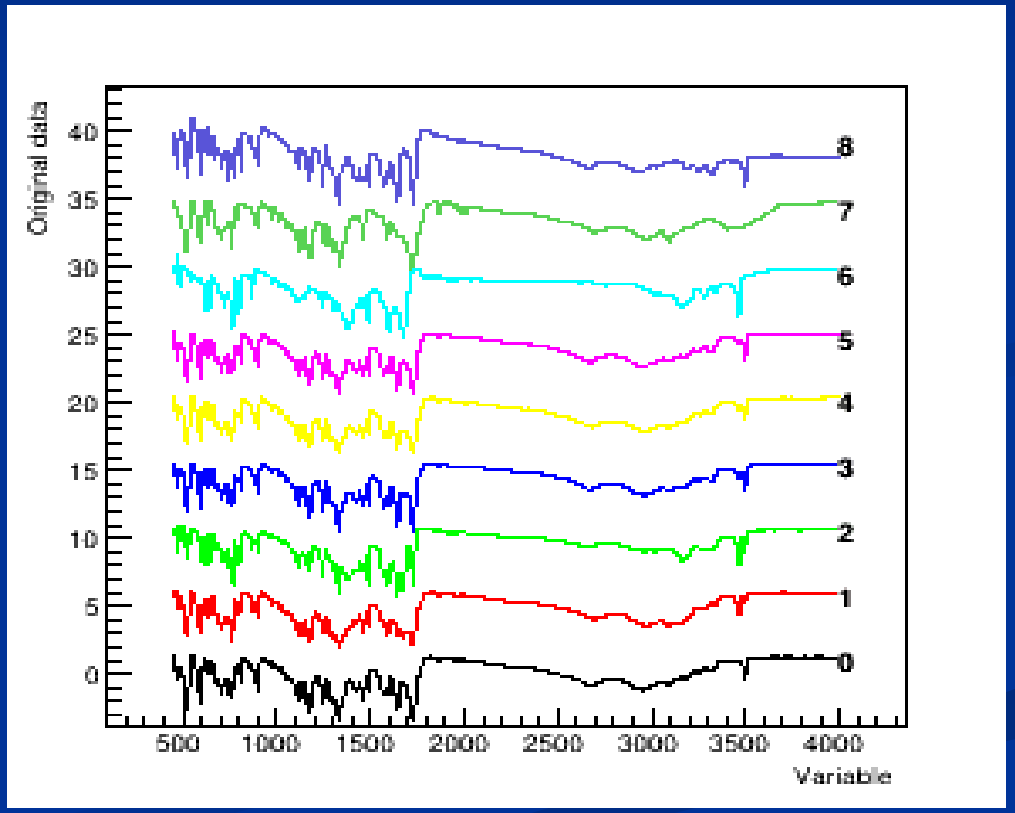
Sample n.	CBZ III	SAC	CBZ-SAC	File name
0	0	0.565	0.435	Rocco_S3_mac.txt
1	0.500	0.500	0	Rocco_S5_mac.txt
2	0.500	0	0.500	Rocco_S7_Como.txt
3	0.347	0.334	0.319	Rocco_S11_mac.txt
4	0.263	0.482	0.255	Rocco_S21_mac.txt
5	0.238	0.364	0.399	Rocco_S22_mac.txt
6	1	0	0	Rocco_CBZ_III_nomac.txt
7	0	1	0	Rocco_SAC_pura_nomac.txt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

Measurements

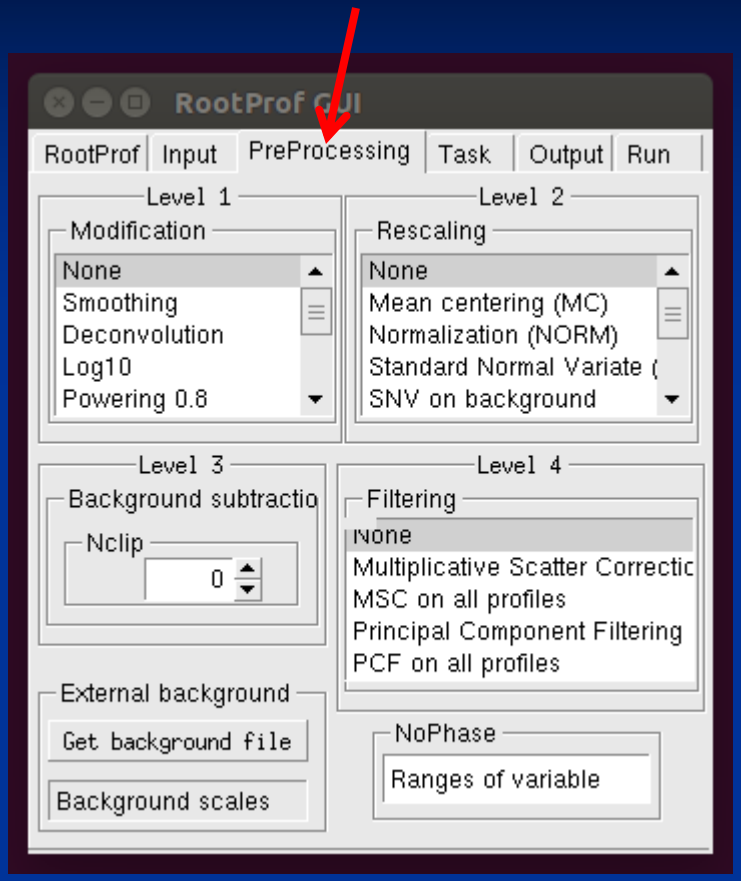
XRPD



FT-IR



Step 2: Pre-processing



Level 1
Modification

Level 2
Rescaling

Level 3
Background
subtraction

Level 4
Filtering

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

Backgroud subtraction

Enhances: - the signal in each profile

- differences among profiles

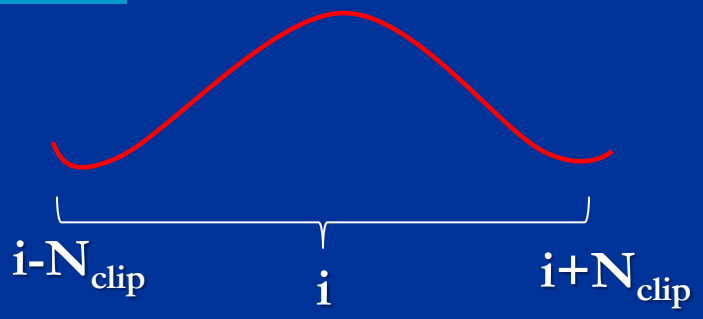
$$\hat{y}'(i) = \hat{y}(i) - b(i)$$

→ Background

Can be estimated by:

Sensitive Nonlinear Iterative Peak (SNIP) clipping algorithm

$$b(i) = \hat{y}_{Nclip}(i)$$



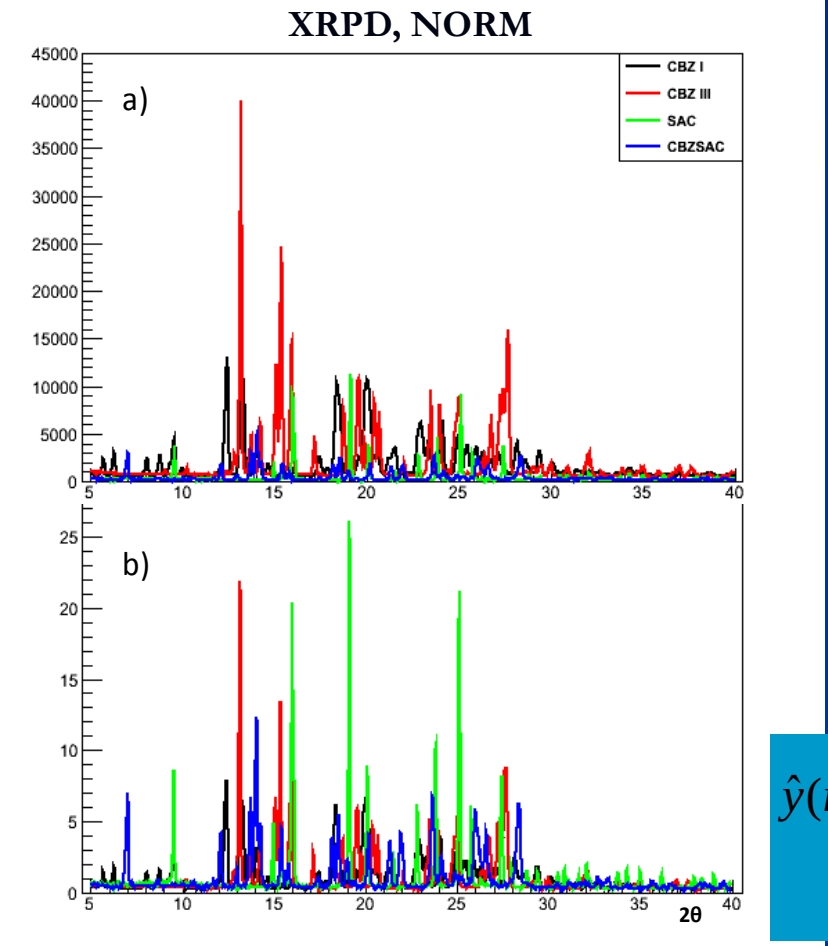
$$\hat{y}_1(i) = \hat{y}(i)$$

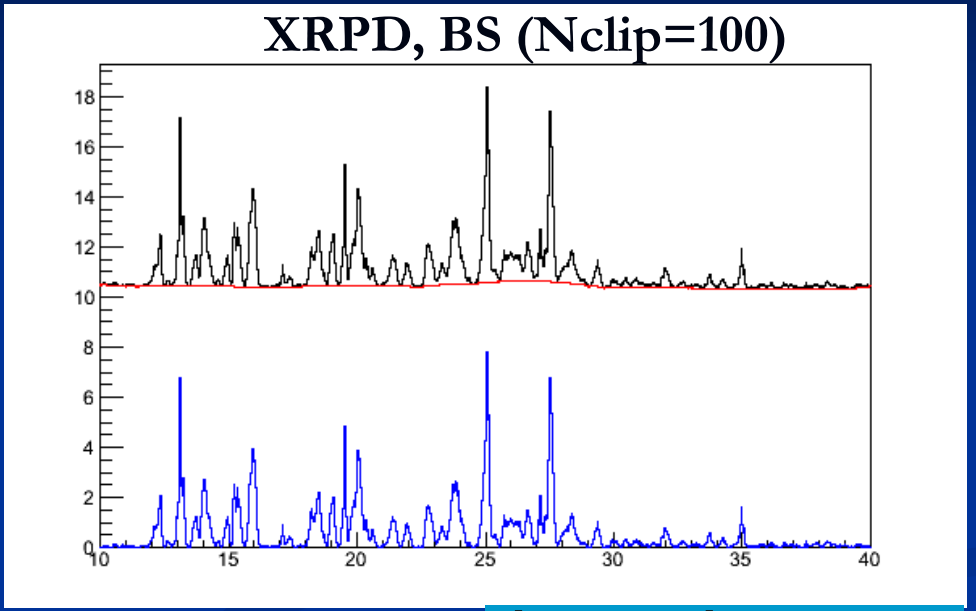
$$\hat{y}_2(i) = Min\left[\hat{y}_1(i), \frac{\hat{y}_1(i+2) + \hat{y}_1(i-2)}{2}\right]$$

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

Nclip: width of clipping window & number of iterations

Pre-processing on XRPD





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

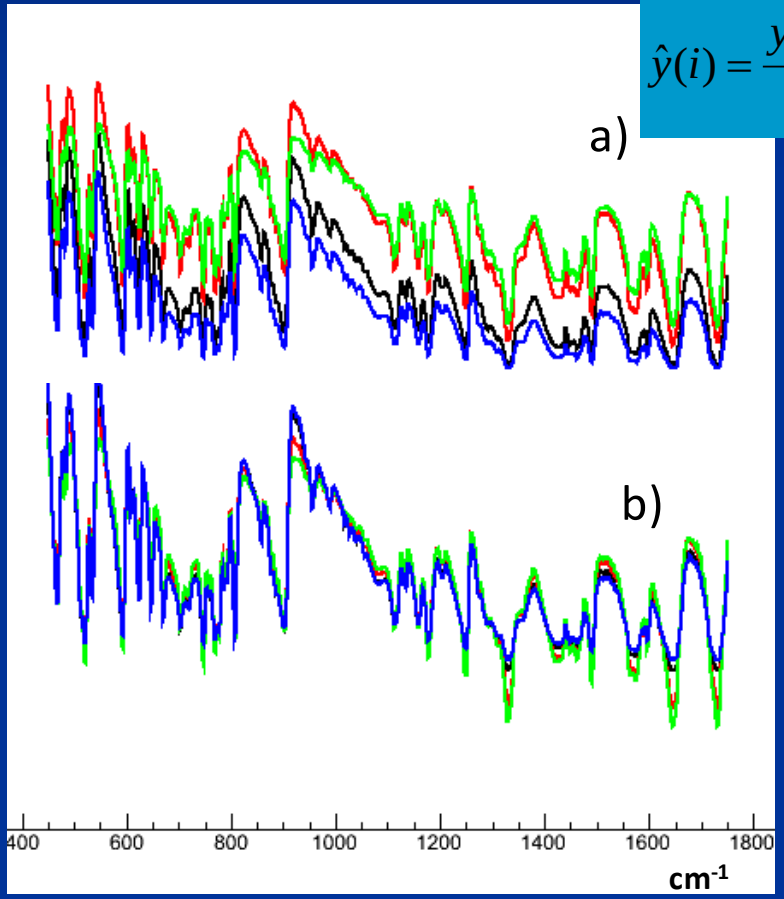
$$\hat{y}(i) = \frac{y(i)}{\sum_{j=1}^{N} y(j)}$$

Pre-processing on FTIR

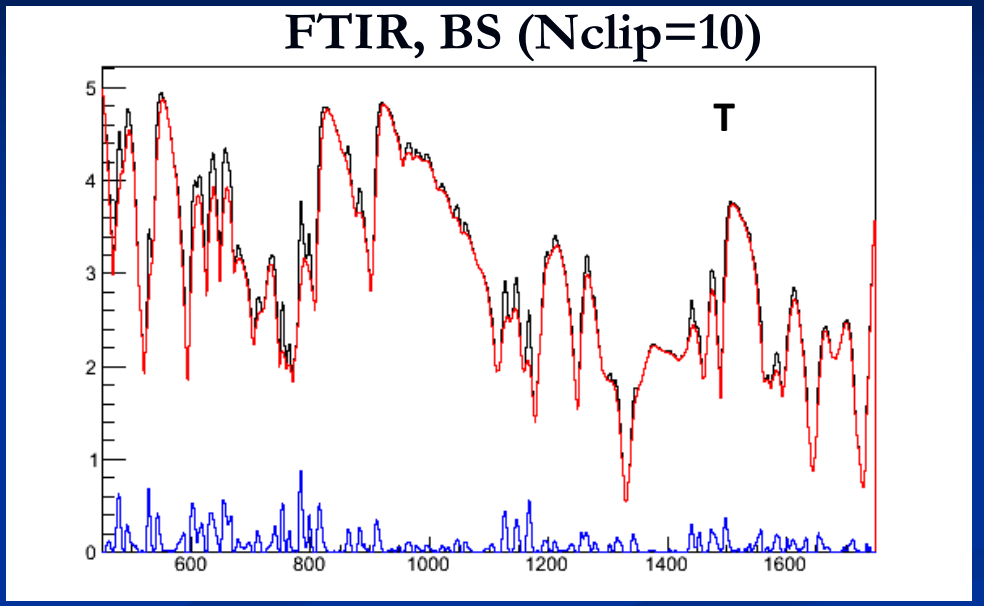
FTIR, SNV

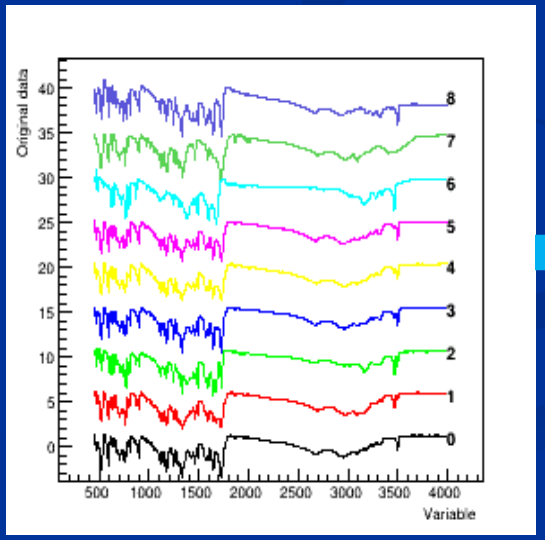
same spectrum with different amount of KBr pellet

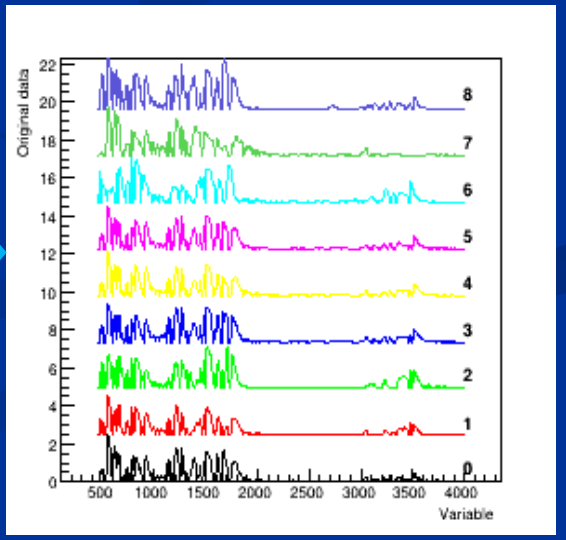
$$\hat{y}(i) = \frac{y(i) - \langle y(i) \rangle}{\sigma_y}$$



a) Before → b) After

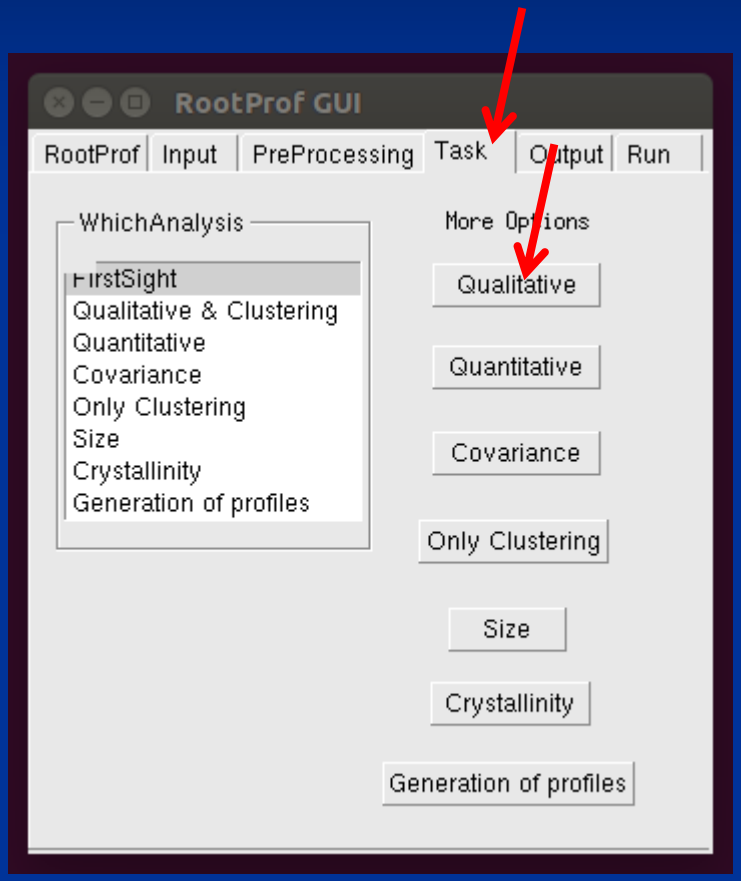


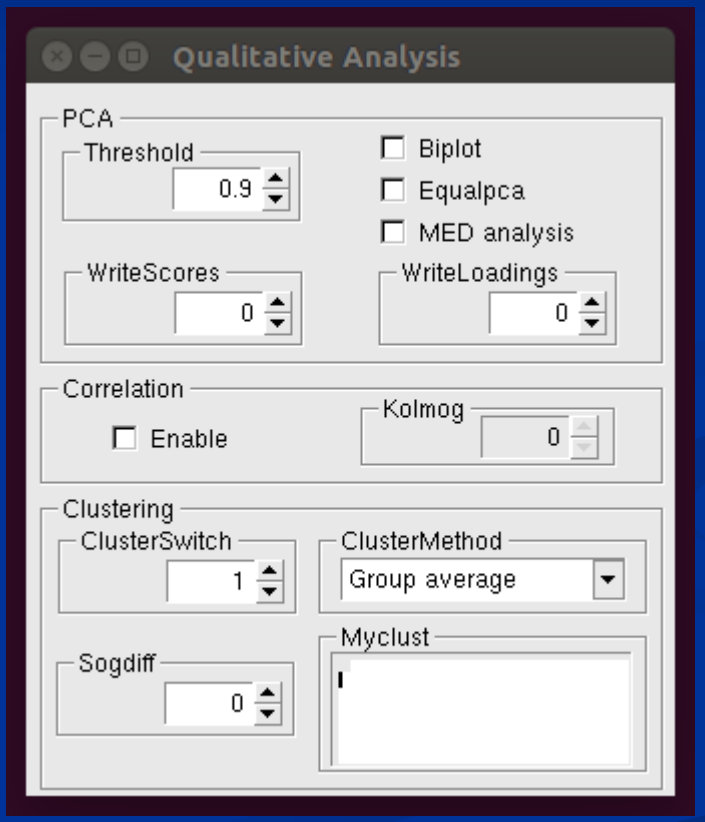




Step 3: Task

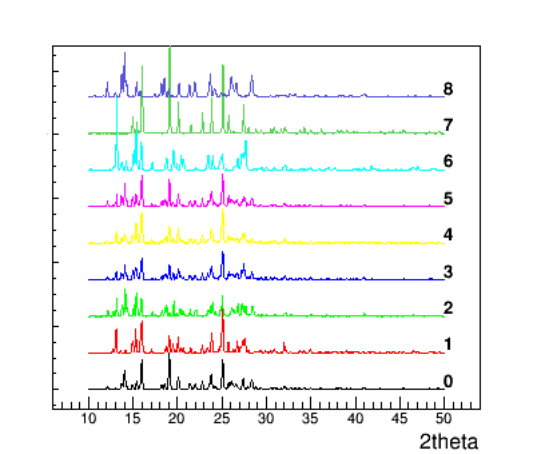
Qualitative analysis

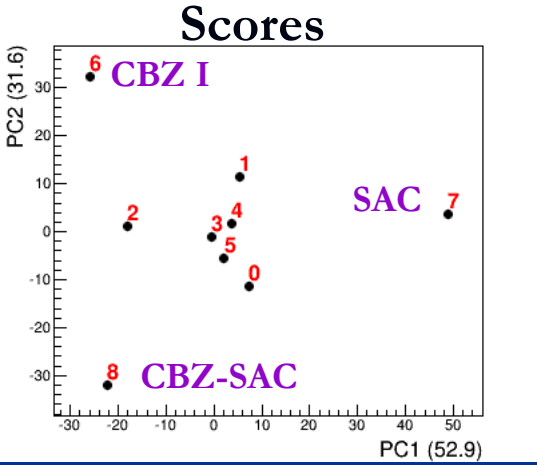


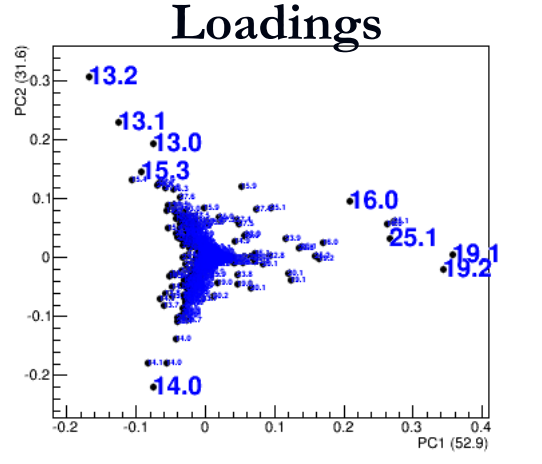


Qualitative analysis by PCA

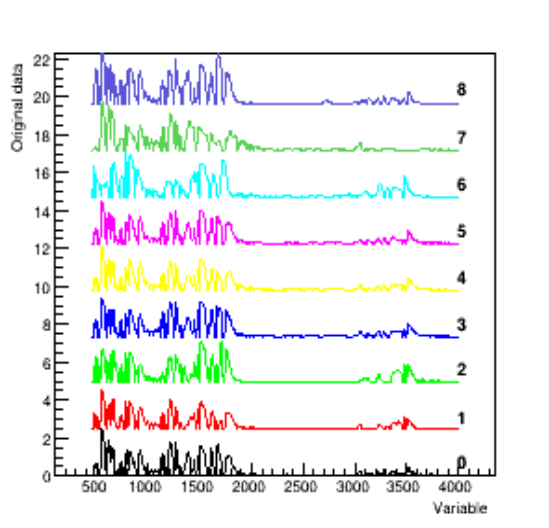
XRPD

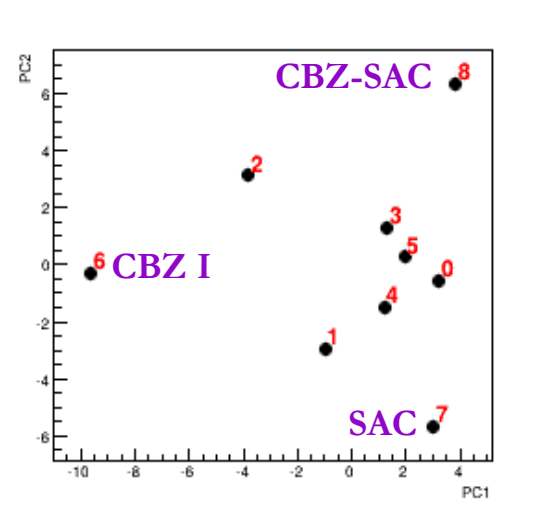






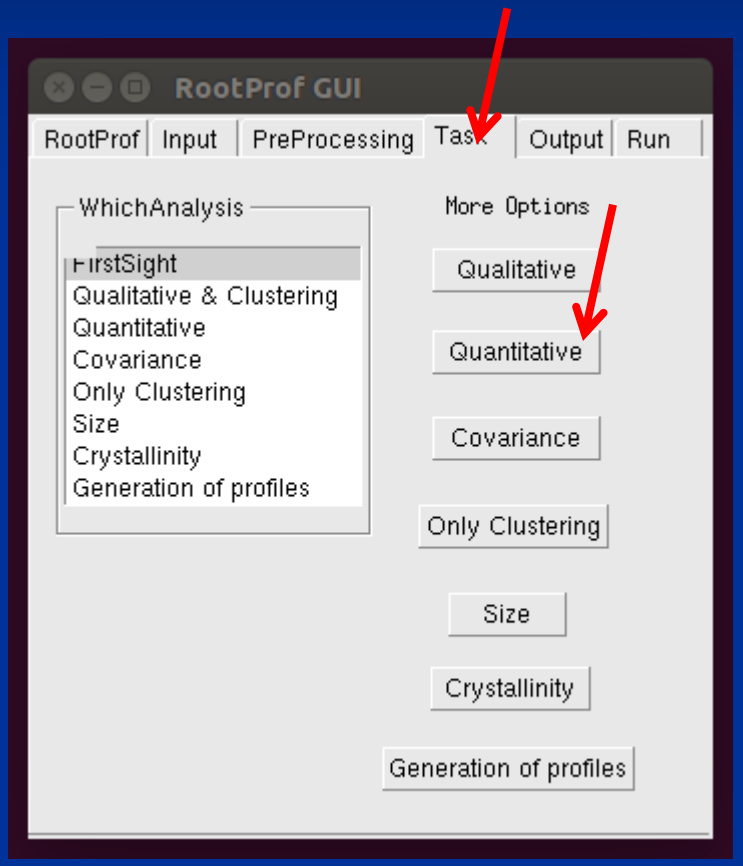
FTIR

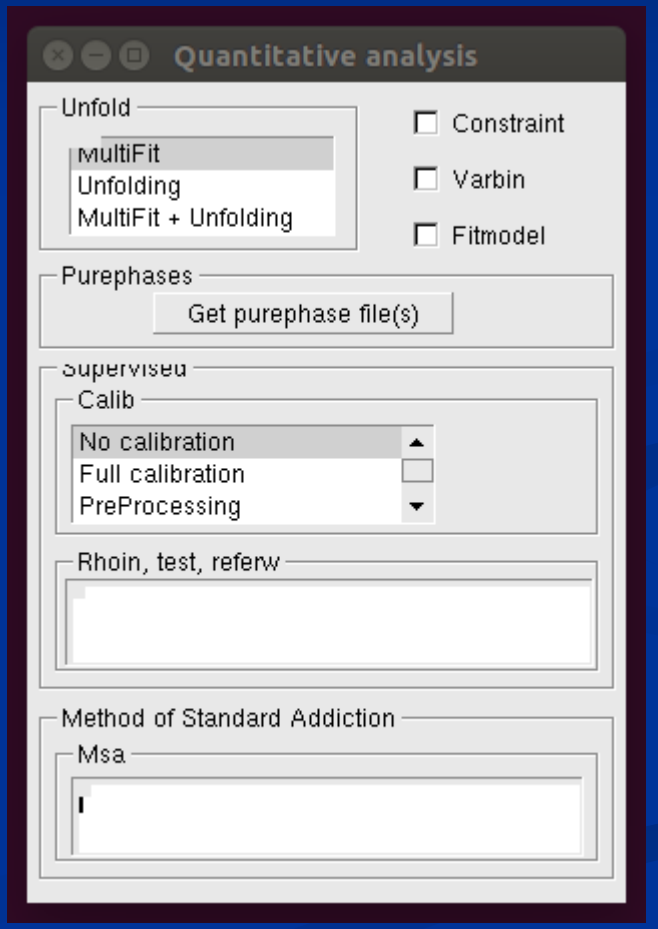




Step 3: Task

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

$w_j = p_j$ **For IR**

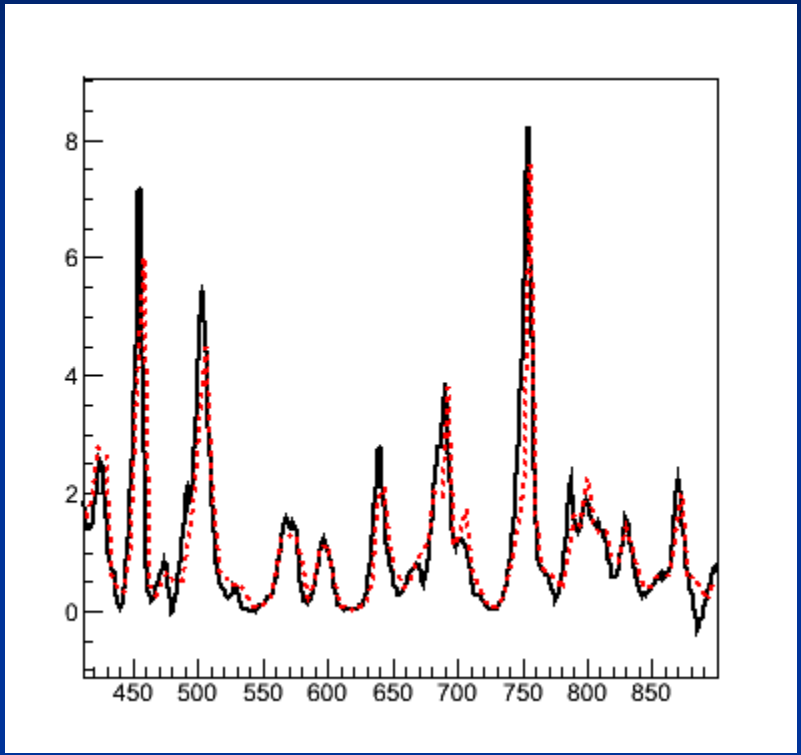
$$w_j = \frac{p_j}{\sum_{t=1}^{M} p_t} \quad \text{or} \quad w_j = \frac{\frac{p_j}{\mu_i^*}}{\sum_{t=1}^{M} \frac{p_t}{\mu_t^*}}$$ **For XRPD**

$$Sum = \sum_{t=1}^{M} p_t$$

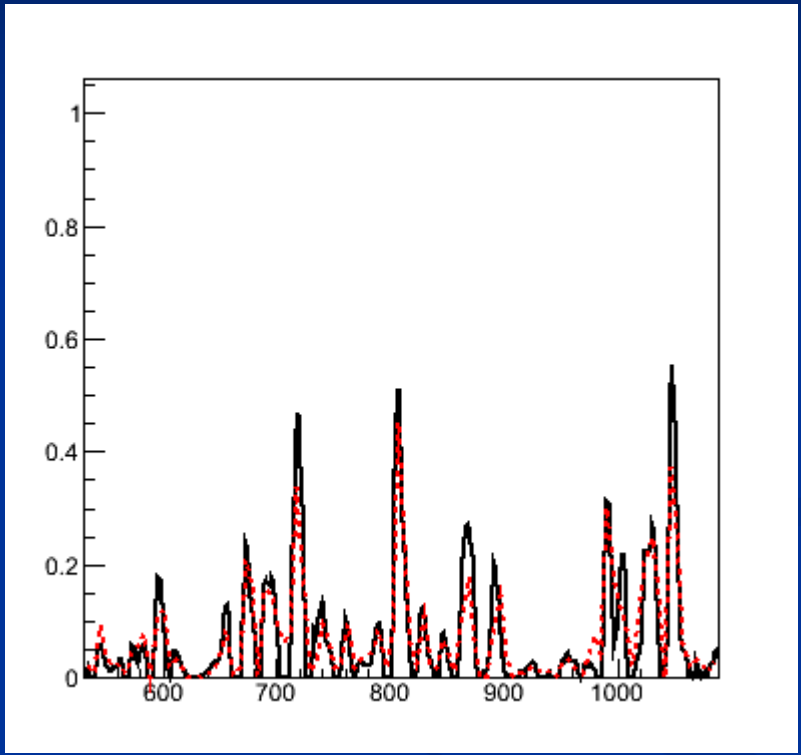
Deviation from 1 indicates phases not included in the fit or amorphous content

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

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

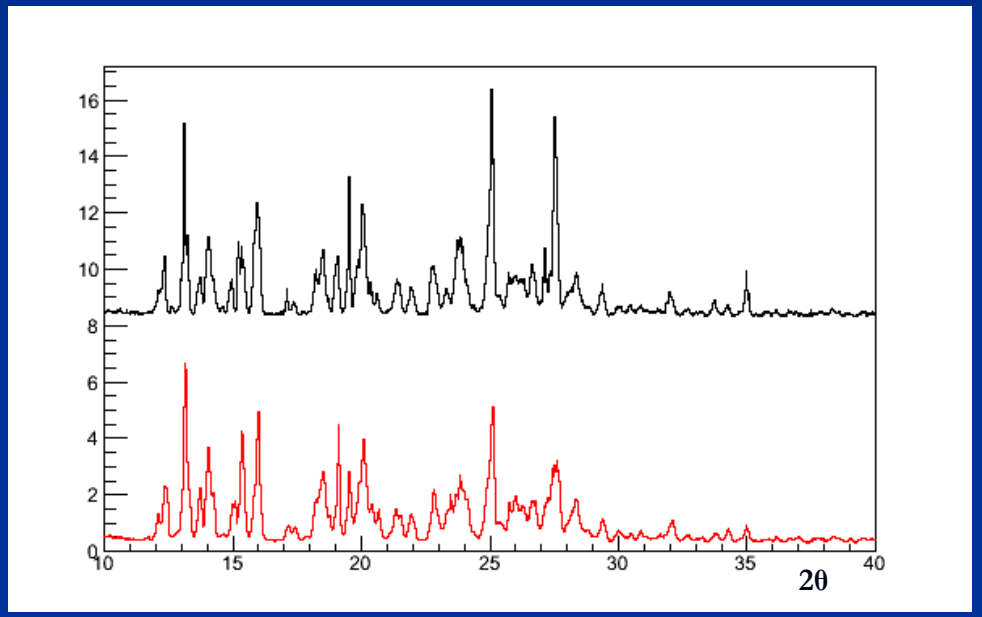
free parameter

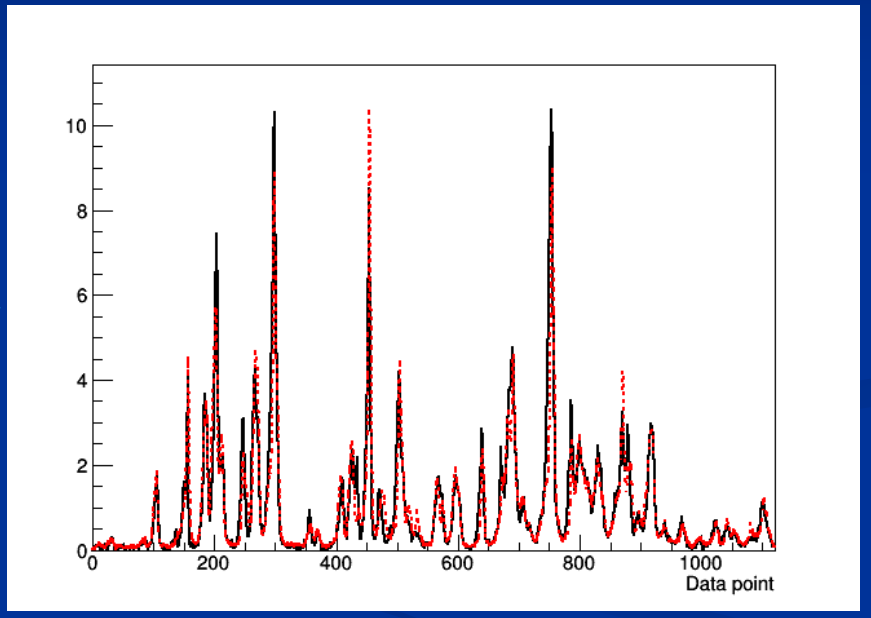
$$Sum = \sum_{t=1}^{M} w_t$$

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

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

Iterative LSQ

$\hat{f}_j(i)$ Pre-processed profile of the j-th pure phase

M= number of pure phases in the sample

N= number of 2theta values

Unfolding

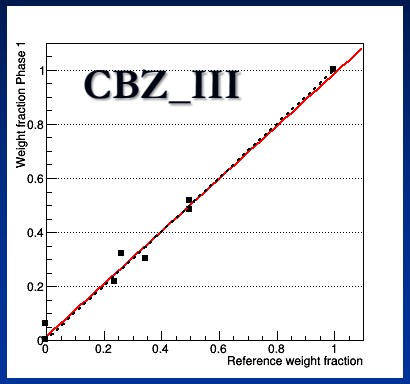
$$p_j = \sum_{i=1}^{N} \hat{f}_j(i)\hat{y}(i)$$

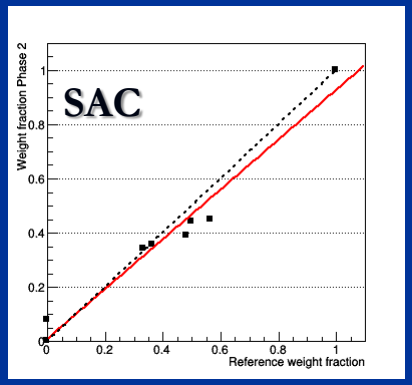
Iterative Gold deconvolution

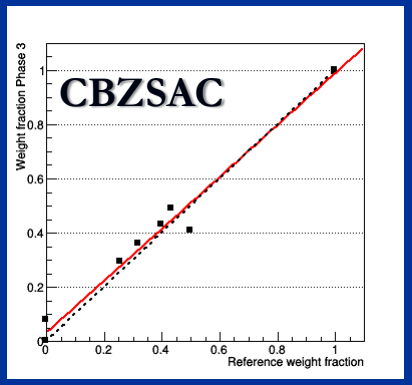
$\hat{f}_j(i)$ Response matrix formed by the pure phase profiles,

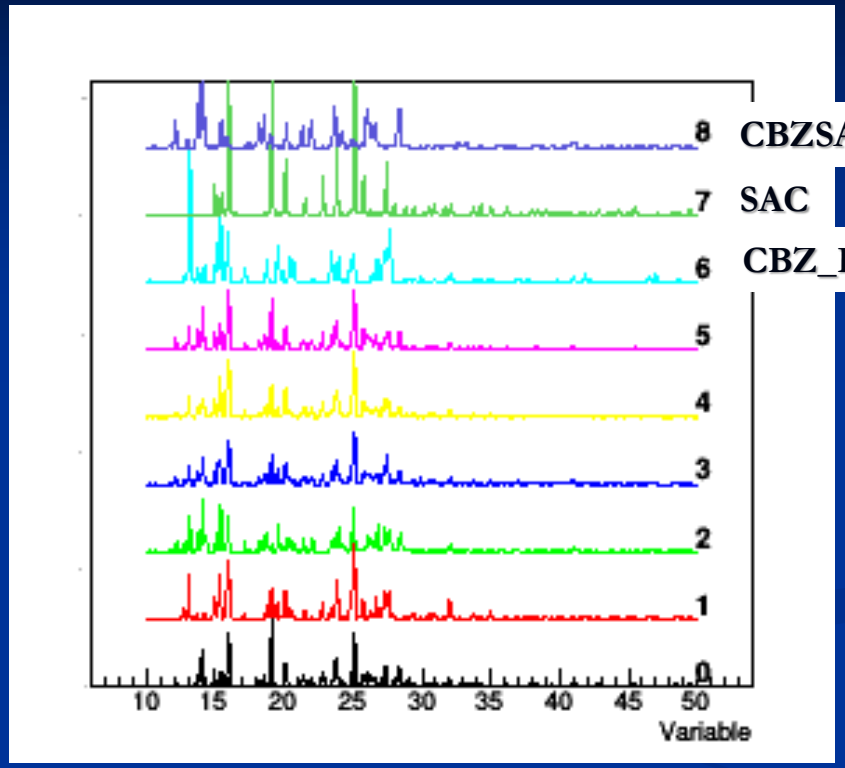
Less accurate, but faster and less dependent on pre-processing

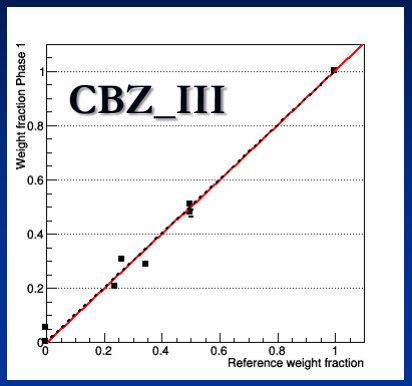
Unfolding vs Multifit

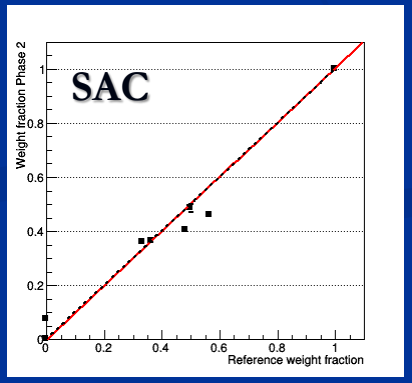


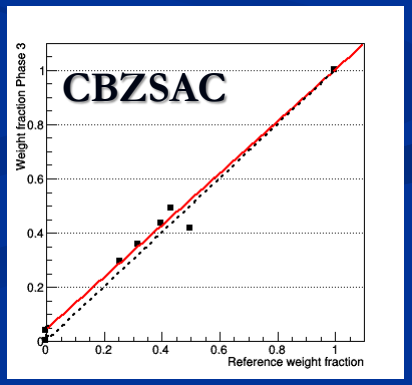












AKLD=0.121
Time=0.4 s

AKLD=0.110
Time=54.4 s

Kullback-Leibler distance

$$AKLD = \left| \sum_{i=1}^{K} w_i' \ln \frac{w_i'}{w_i} \right|$$

Supervised quantitative analysis

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

Automatic choice of best pre-processing option

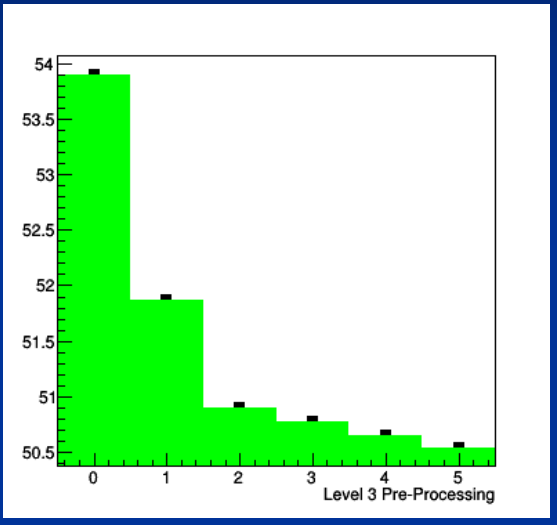


Table 1: Pre-processing values considered for automatic calibration

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

Best-fit determination of pure phase rescaling

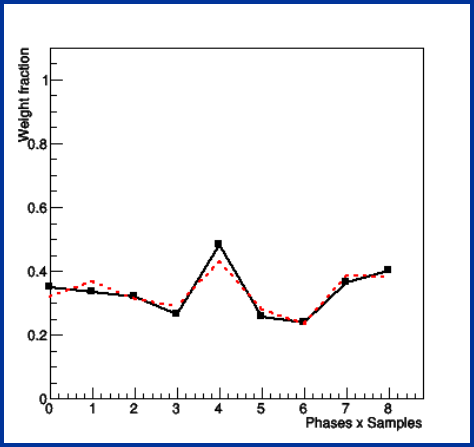
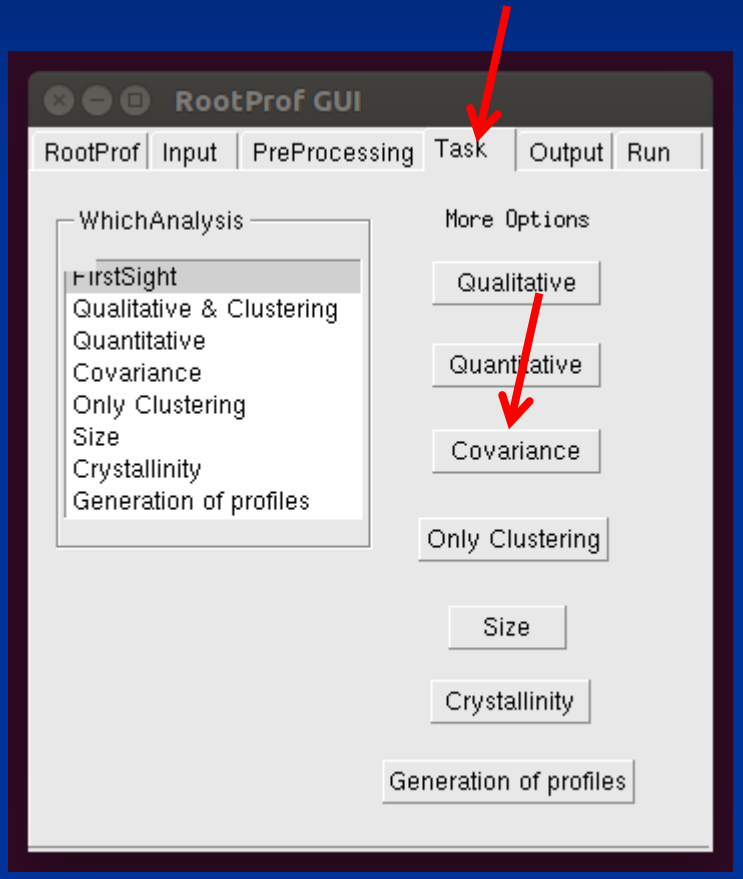


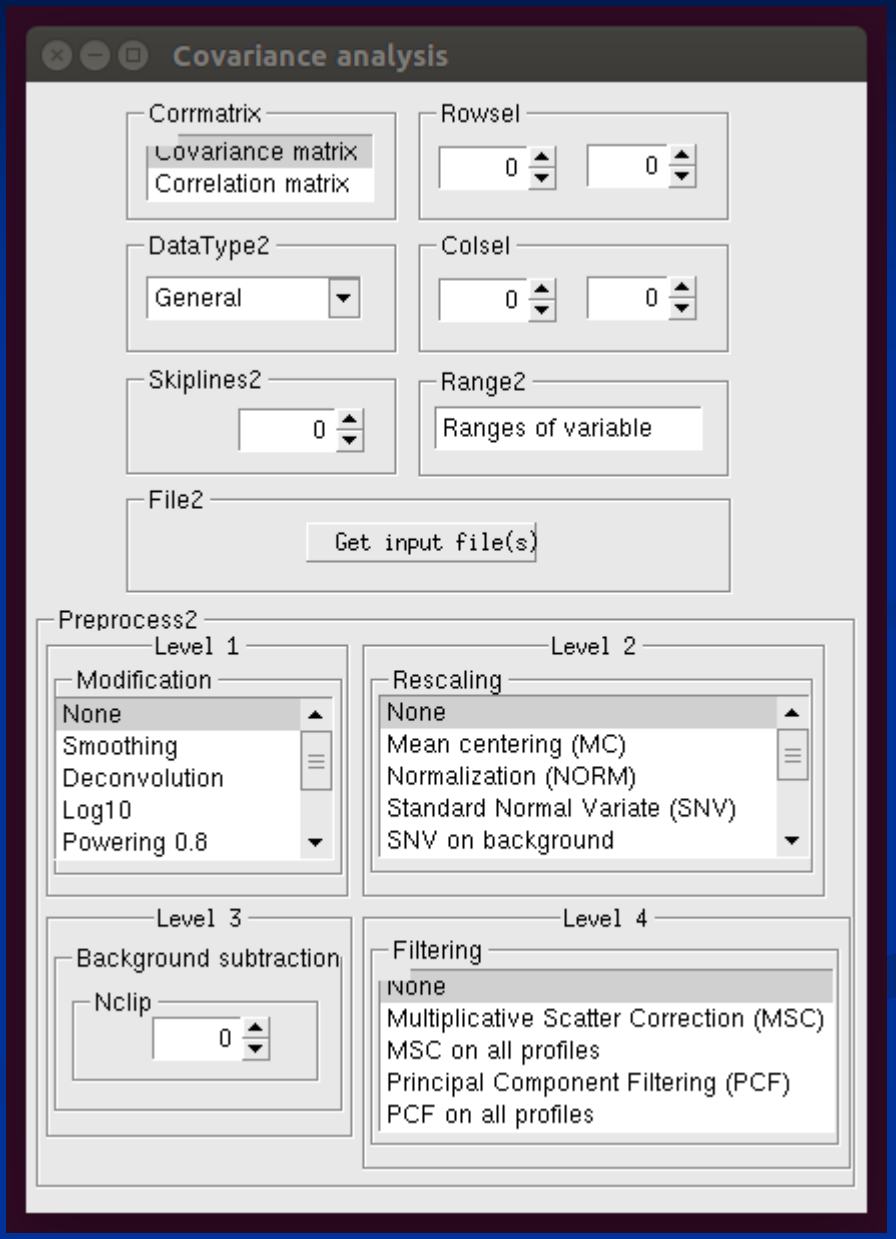
Table 1: Results of the calibration procedure by using different compositions of the calibration set.

Samples in Calibration Set	Calibration Parameters	Overall AKLD
--	1.00, 1.00, 1.00	0.108
0,1,2	1.00, 0.88, 0.98	0.104
3,4,5	1.00, 1.02, 1.16	0.094
0,1,2,3,4,5	1.00, 0.93, 1.05	0.097
6,7,8	1.00, 1.00, 1.00	0.108
0,1,2,6,7,8	1.00, 0.87, 0.97	0.104
3,4,5,6,7,8	1.00, 1.02, 1.16	0.094
0,1,2,3,4,5,6,7,8	1.00, 0.93, 1.05	0.097

Step 3: Task

Covariance analysis

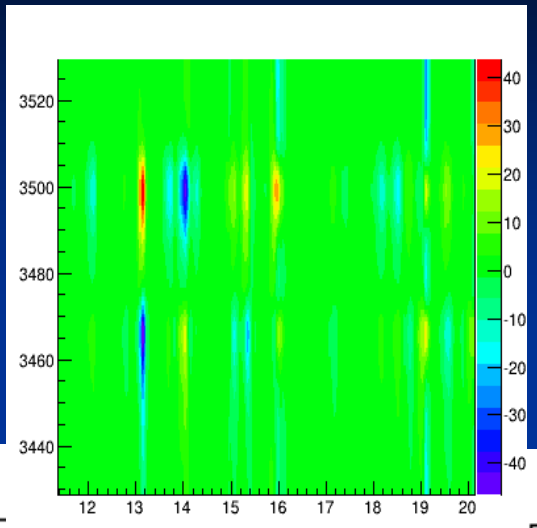




Covariance analysis

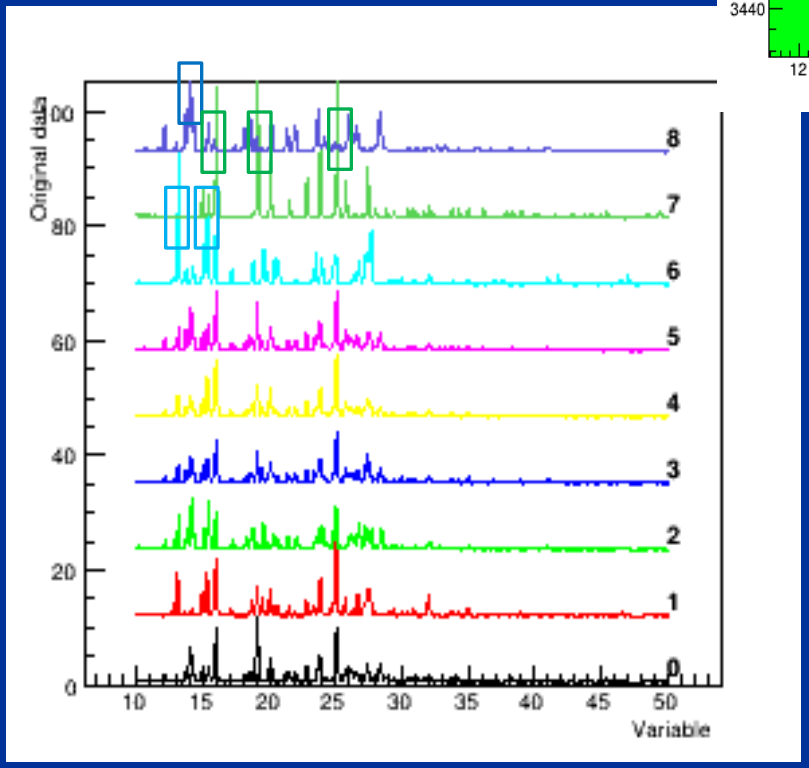
Averages calculated over the samples

$$COV_{AB}(i,j) = \frac{\sum_{t=1}^{K} \left(\hat{y}_t(i) - \langle \hat{y}(i) \rangle\right)_A \left(\hat{y}_t(j) - \langle \hat{y}(j) \rangle\right)_B}{K-1}$$



A=XRPD

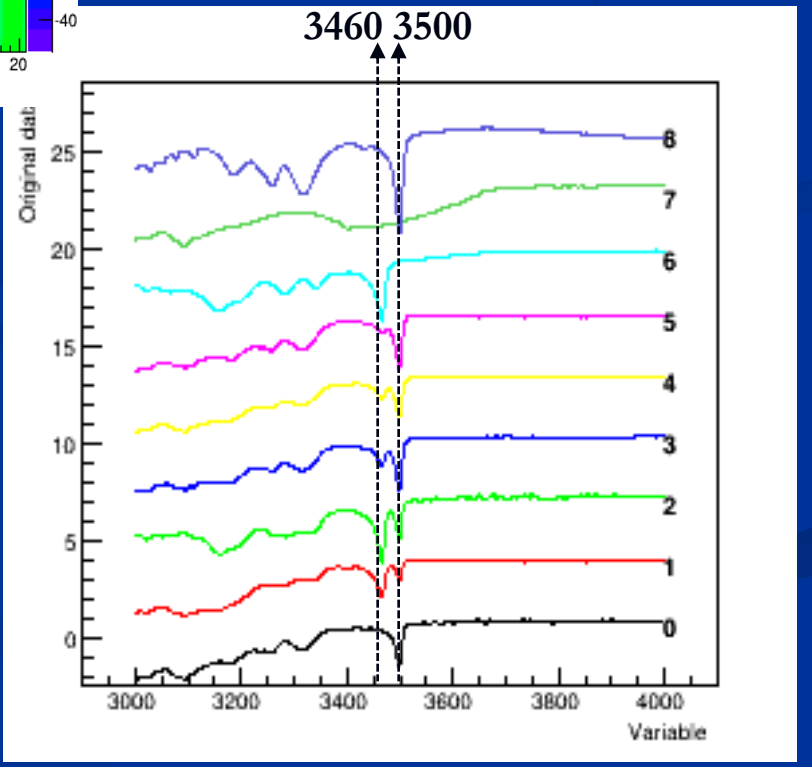
B=FTIR



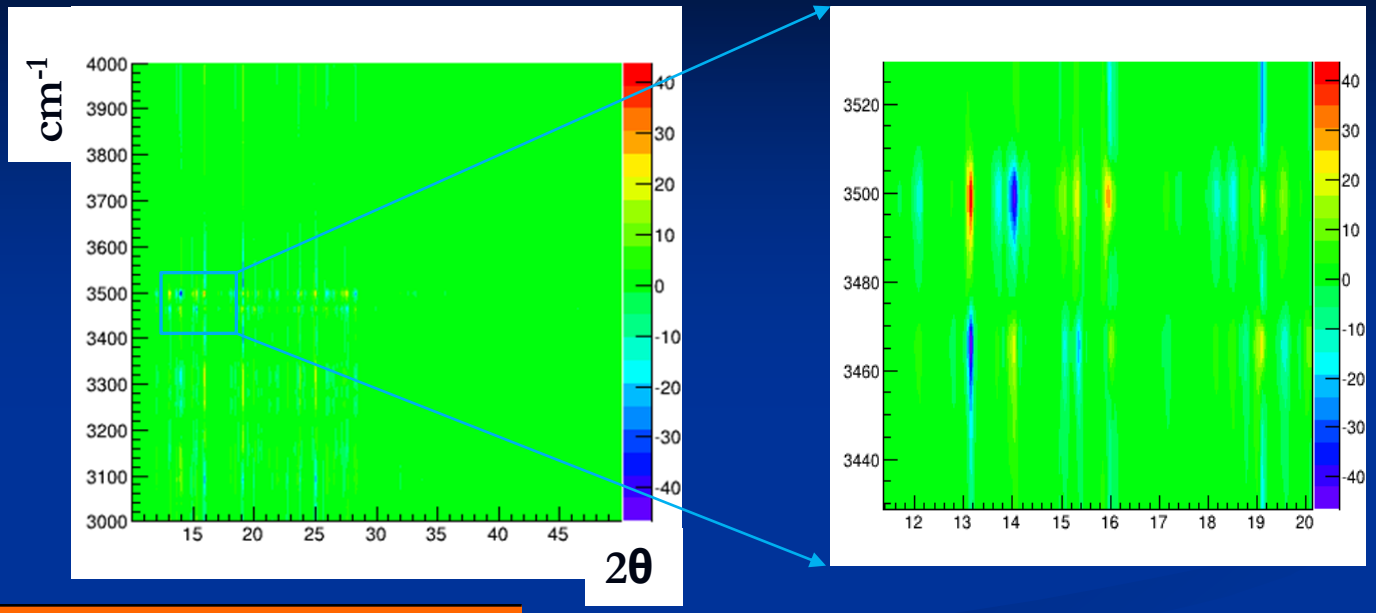
CBZ-SAC

SAC

CBZ III



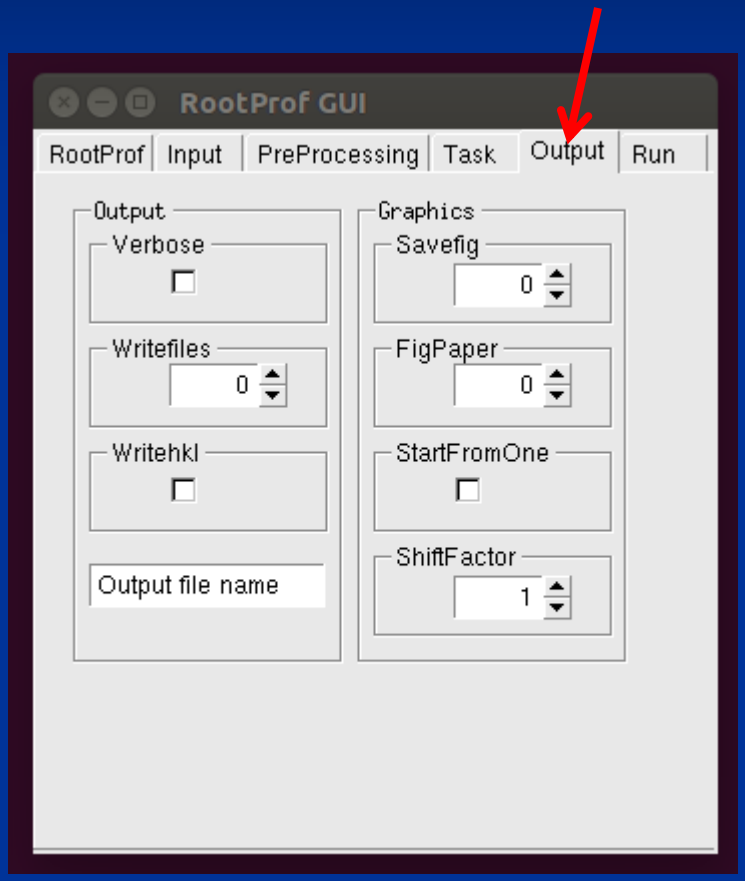
Covariance matrix interpretation

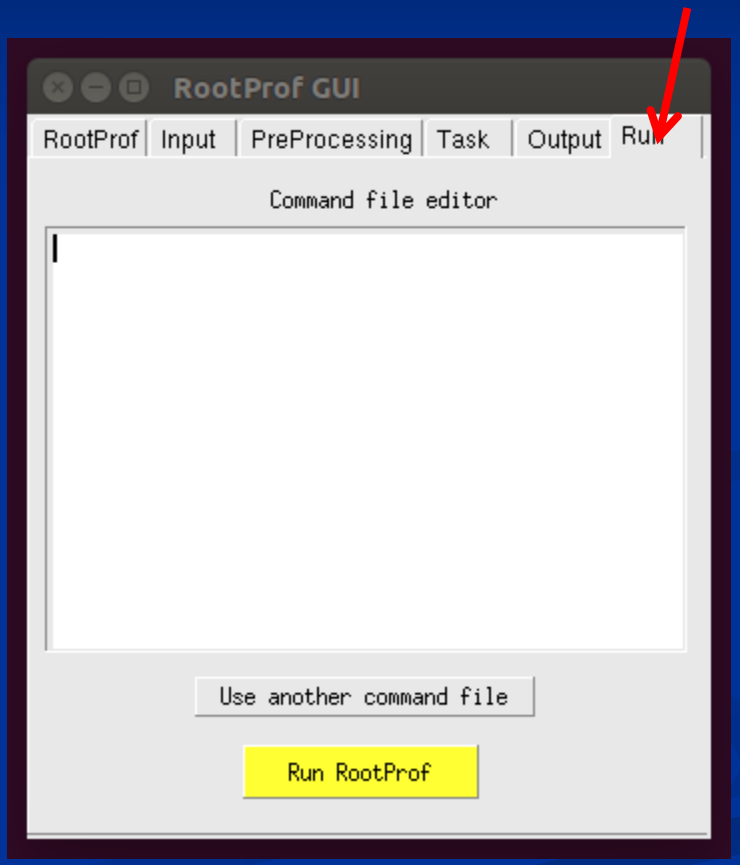


Crystal phase	XRPD signal (2θ)	FTIR signal (cm^{-1})	Correlation
CBZ-SAC	14.0	3460	Positive
		3500	Negative
SAC	16.0, 19.5	3460	Positive
		3500	Positive
CBZ III	13.2, 15.5	3460	Negative
		3500	Positive

3460 and 3500 cm^{-1} FTIR peaks very sensitive for discrimination of CBZ III, SAC and CBZ-SAC

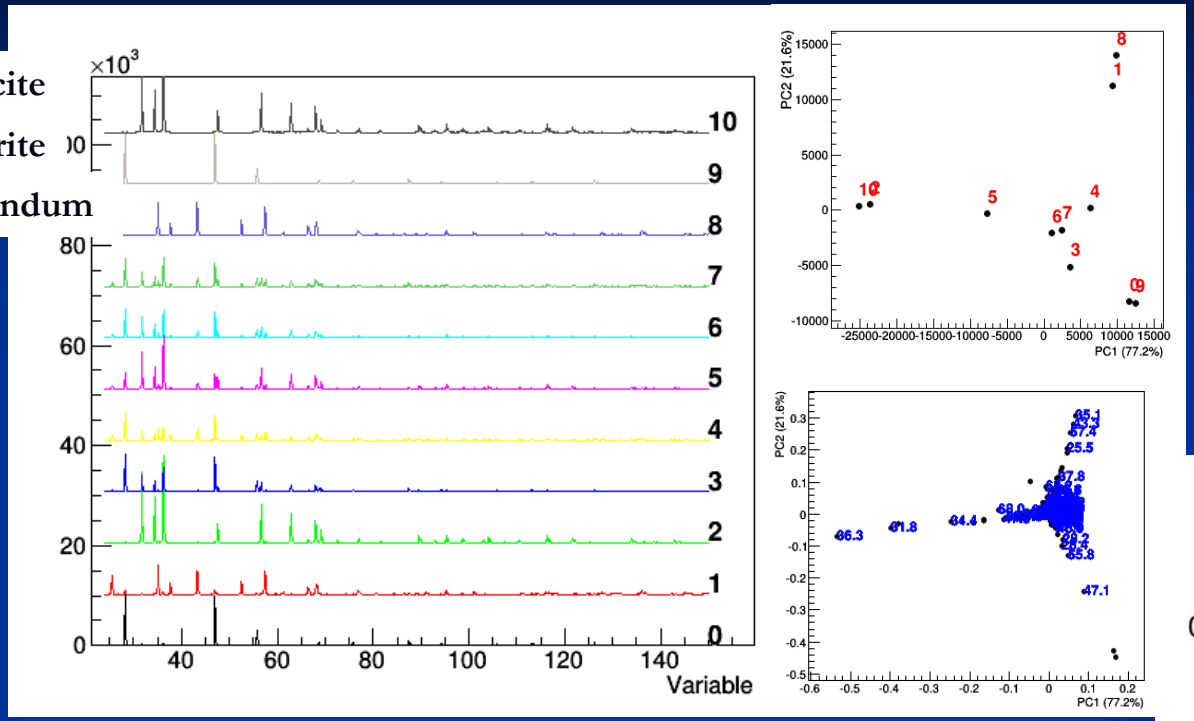
Step 4: Output & Step 5: Run



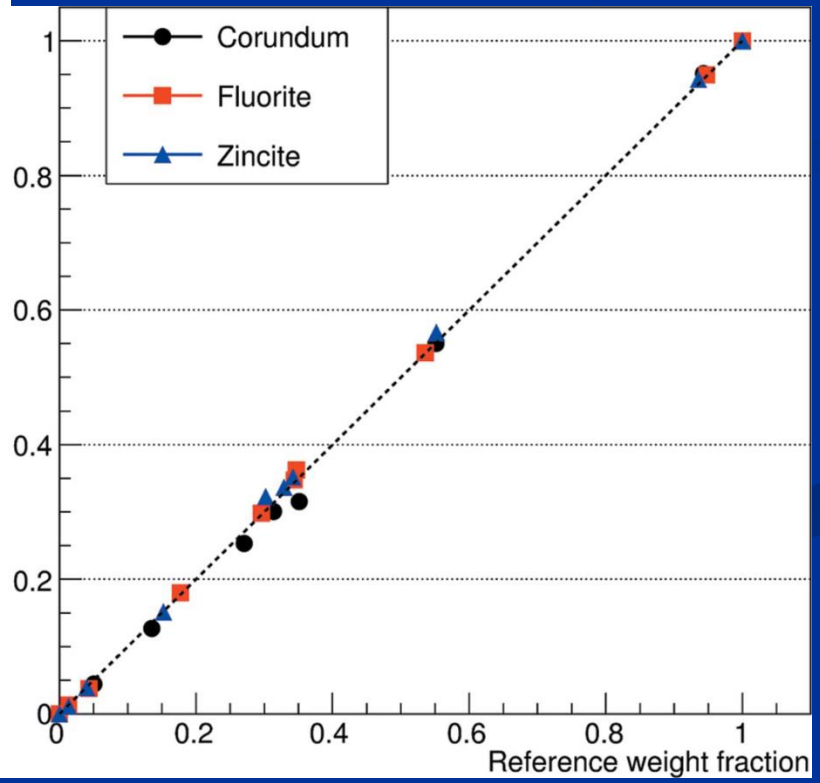


Round-robin mixtures

Zincite
Fluorite
Corundum



Known x-ray attenuation coefficient

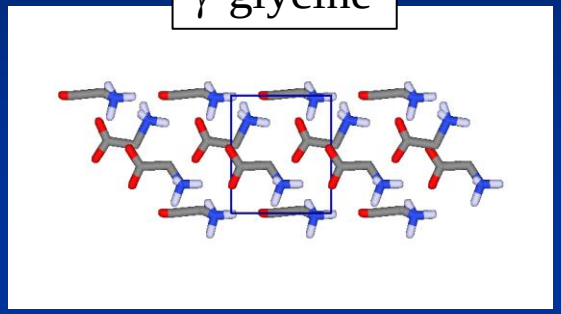


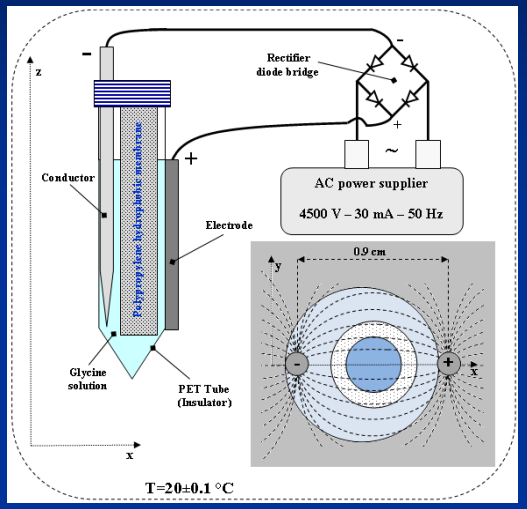
AKLD = 0.021

comparable results obtained by the Rietveld method

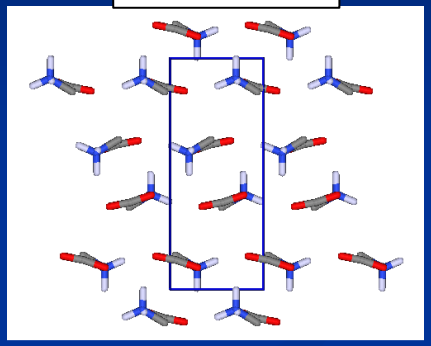
Polymorphism of glycine

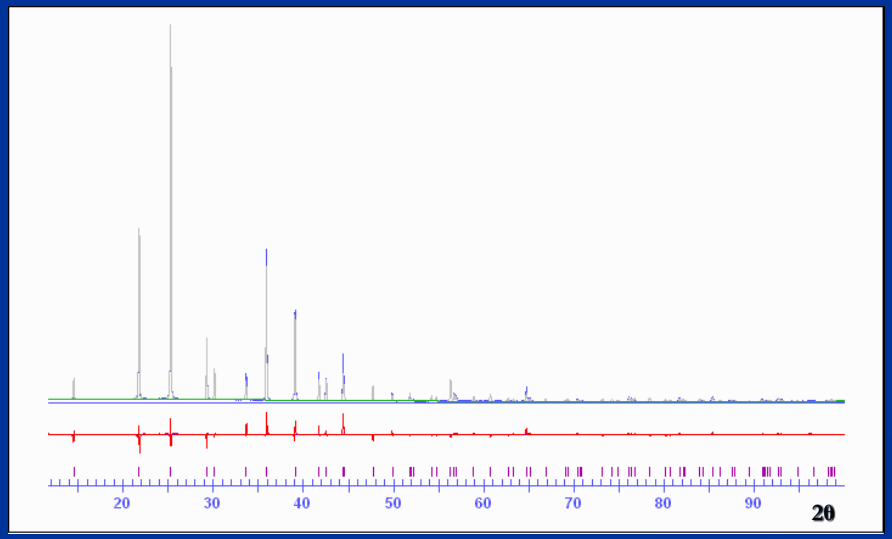
γ-glycine

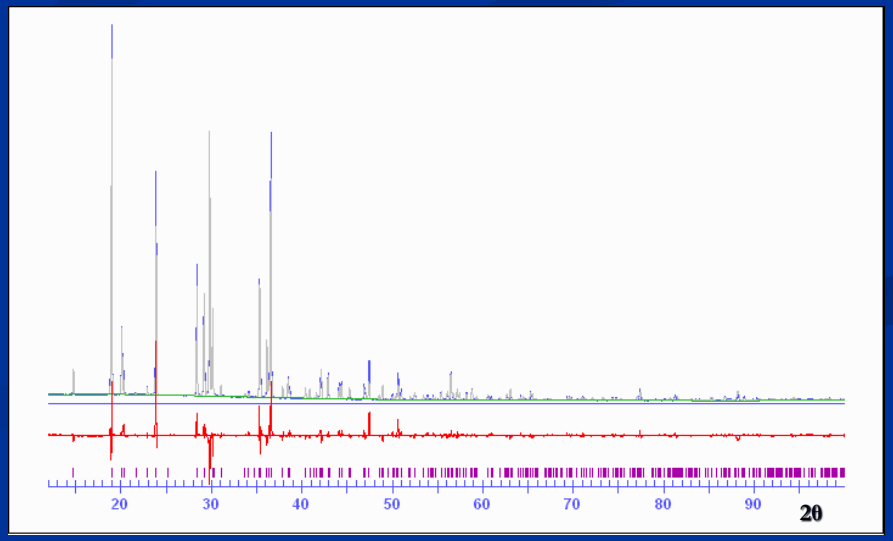




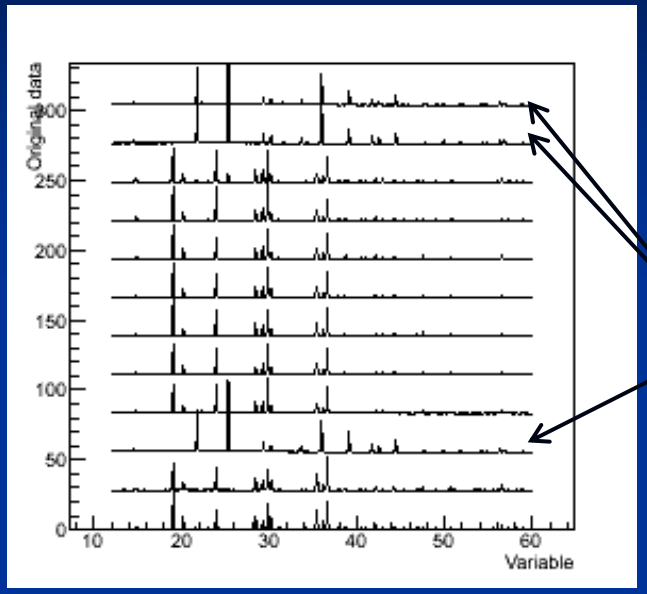
α-glycine

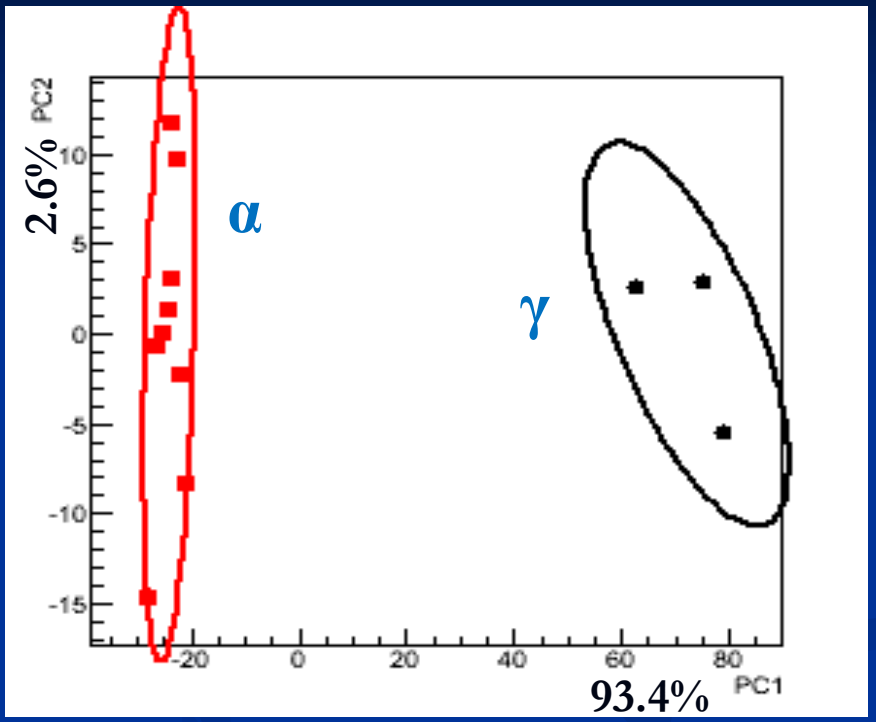






Classification of XRPD patterns





Mahalanobis distance

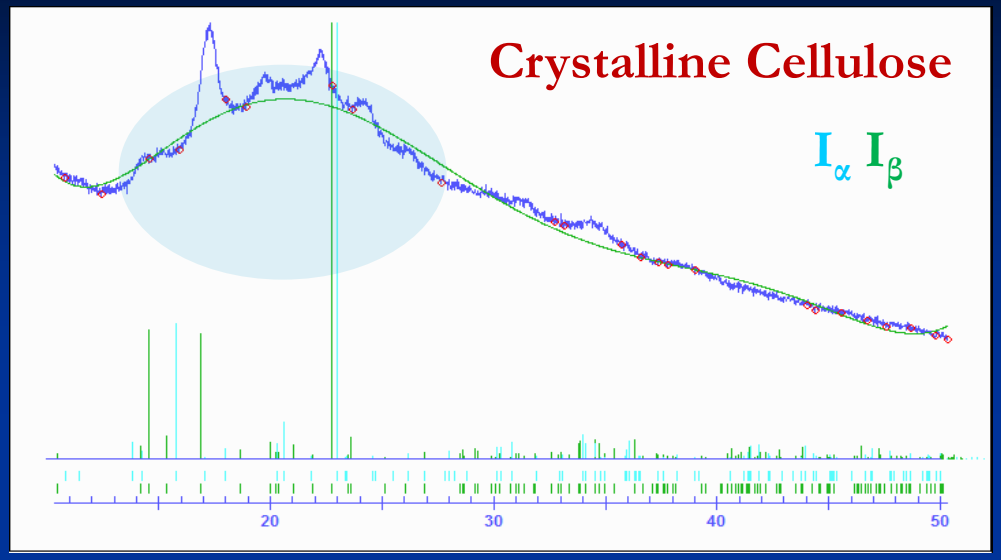
$$dist = \sqrt{\left(\mathbf{u}_j - \mathbf{u}_i\right)^T COV \left(\mathbf{u}_j - \mathbf{u}_i\right)}$$

u_i=sample mean of group i

95% confidence ellipses and P-values

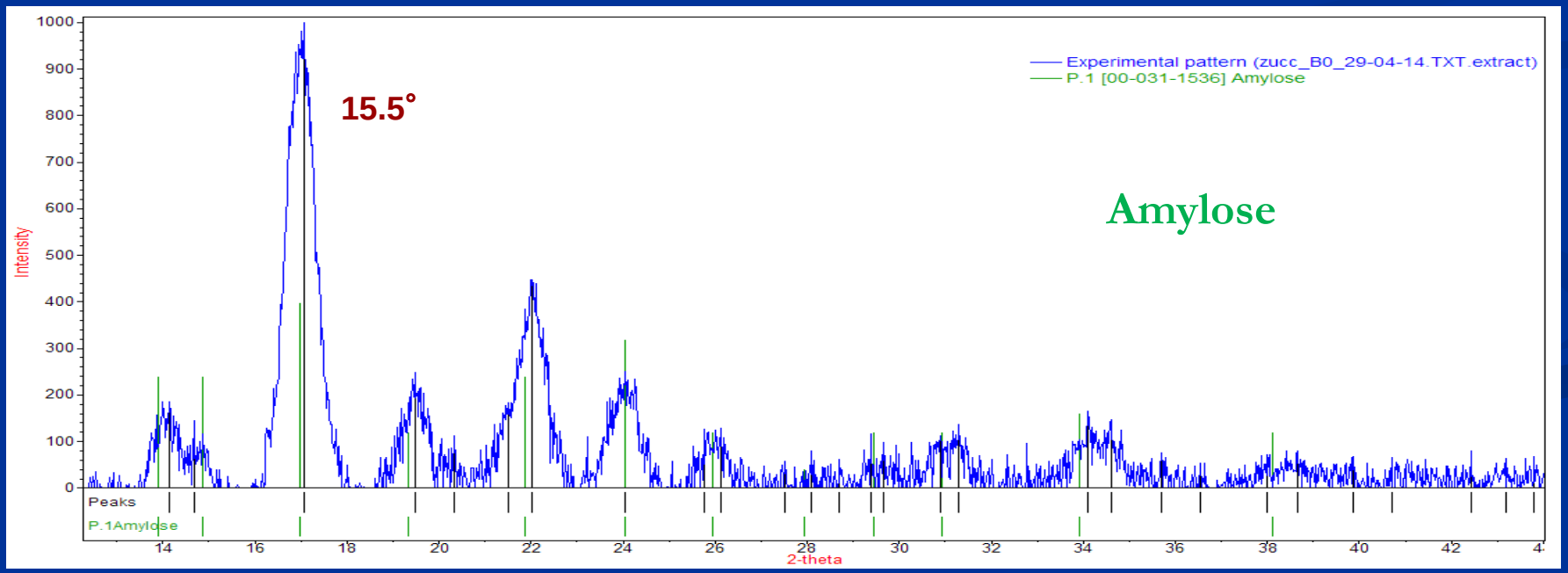
P-value = 5.2×10^{-10}

Characterization of zucchini fruits

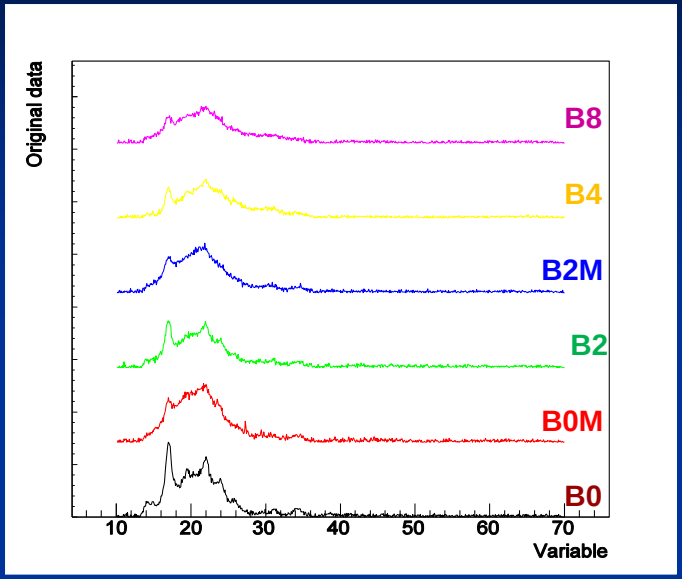


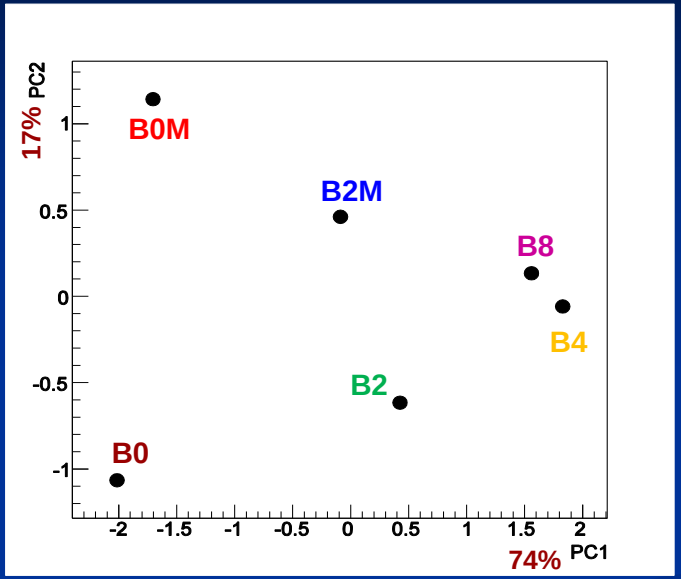
Biochar added to soil:
0% (B0), 2% (B2), 4% (B4), 8% (B8)

Mychorrhizial product addition:
B0M, B2M



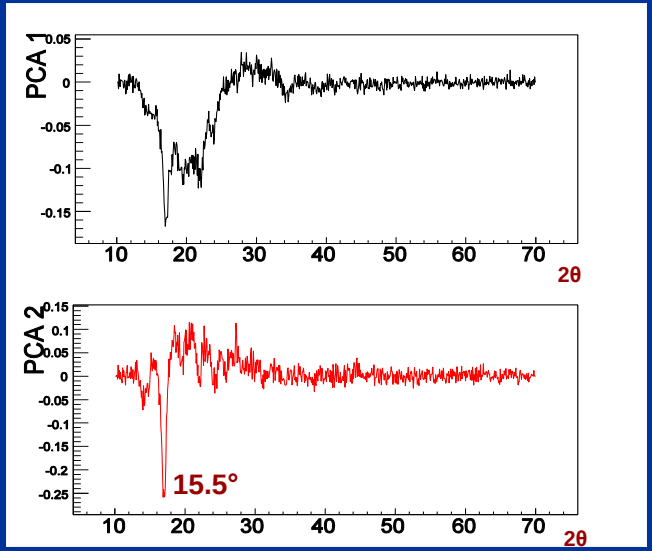
PCA analysis of fruits





Sensitive to biochar addiction (PC1)

Recognize mycorrhizial product addition (PC2)



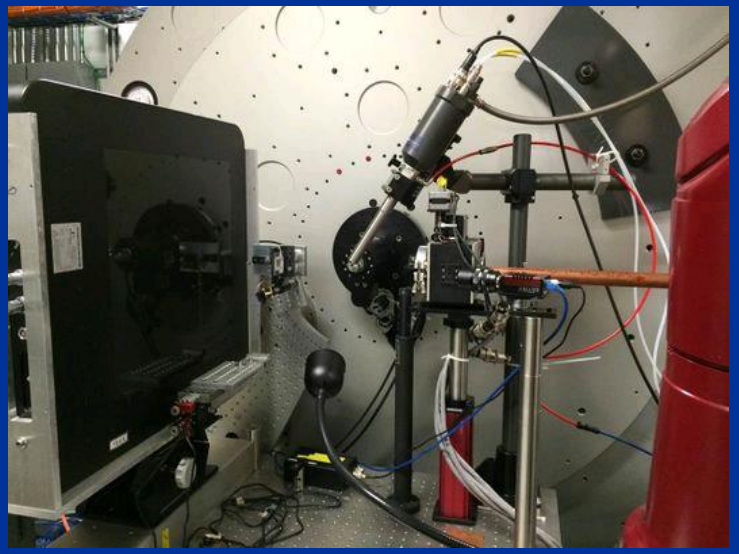
PC1: relative abundance of crystalline cellulose

PC2: relative abundance of amylose (15.5°)

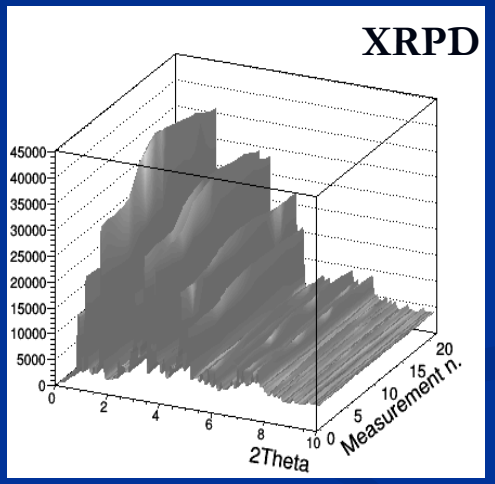
Structural characterization of halide perovskites

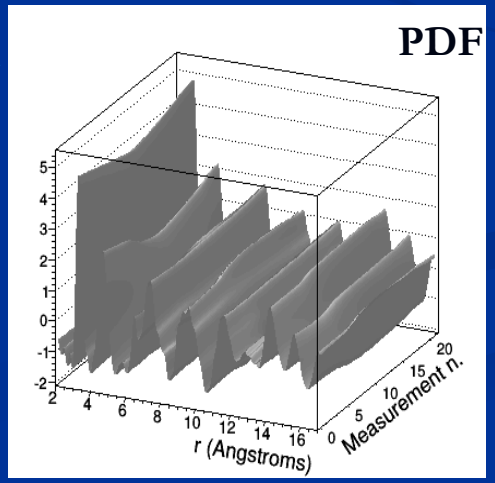
Temperature varied in situ 300 → 400 → 300 ° C

Tetragonal-to-cubic phase transition monitored

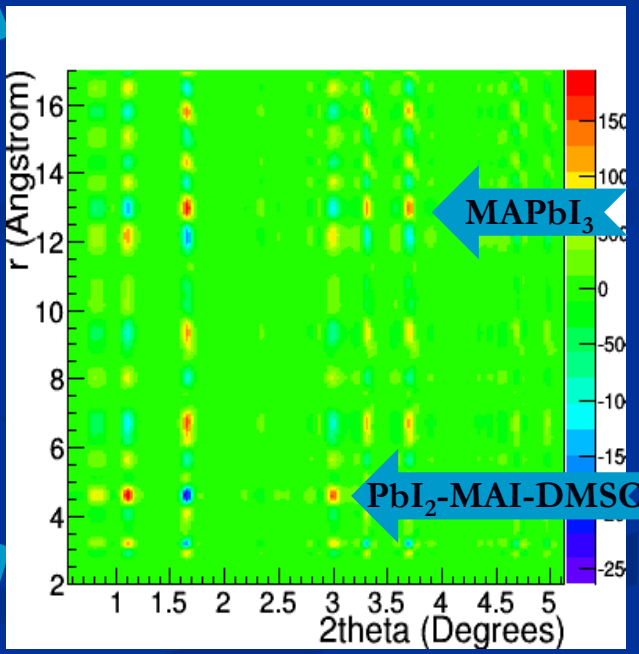


R. Caliandro, D. Altamura, B.D. Belviso, A. Rizzo, S. Masi, C. Giannini
J. Appl. Cryst. 2019, 52





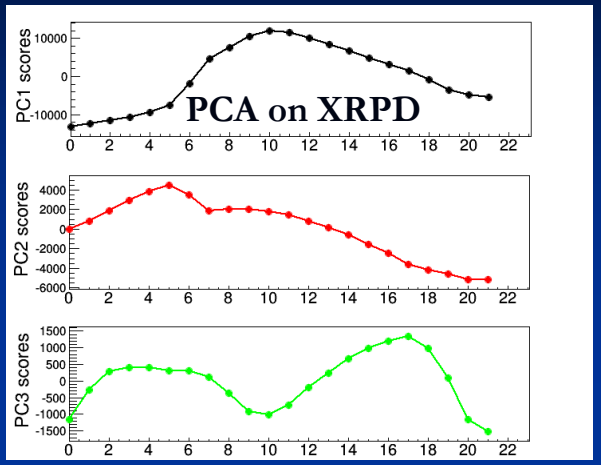
Covariance matrix calculated by RootProf

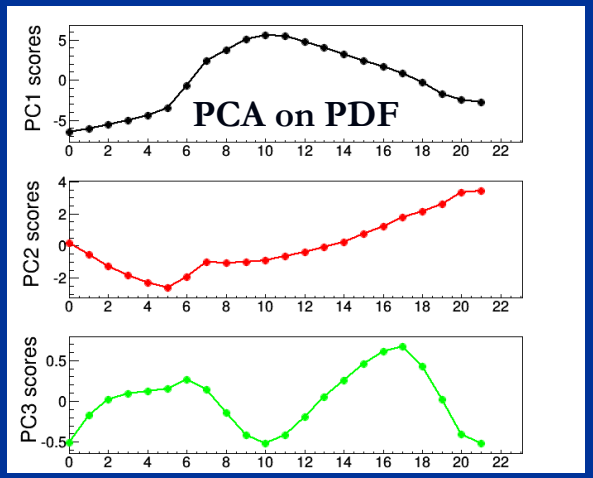


$MAPbI_3$ and PbI_2-MAI-DMSO peaks anticorrelated

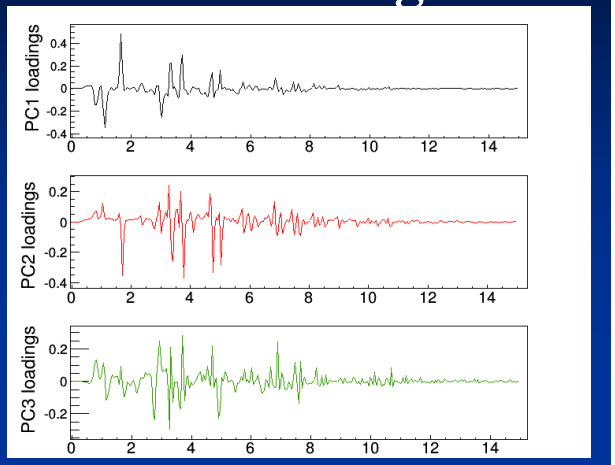
Trends in time and space domains

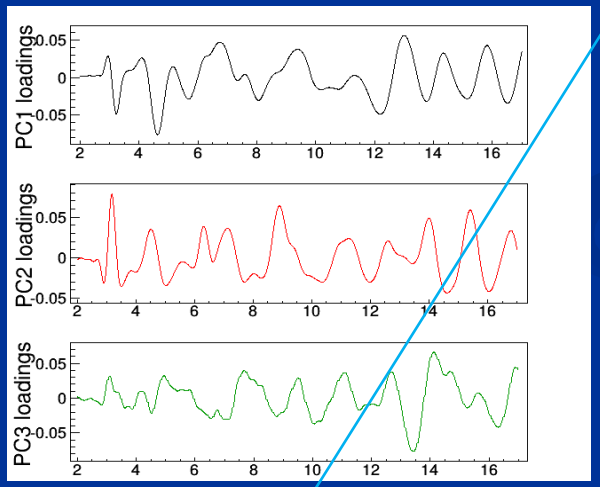
Scores



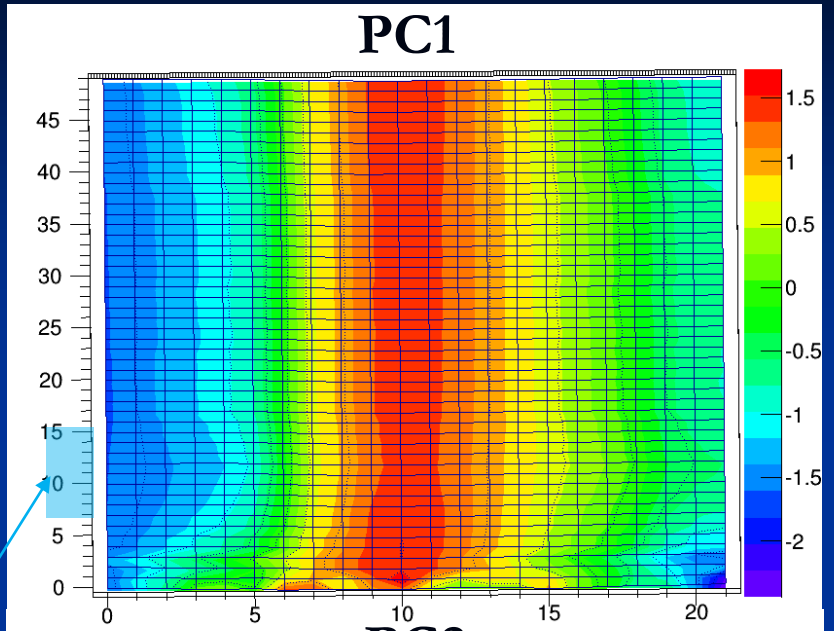


Loadings

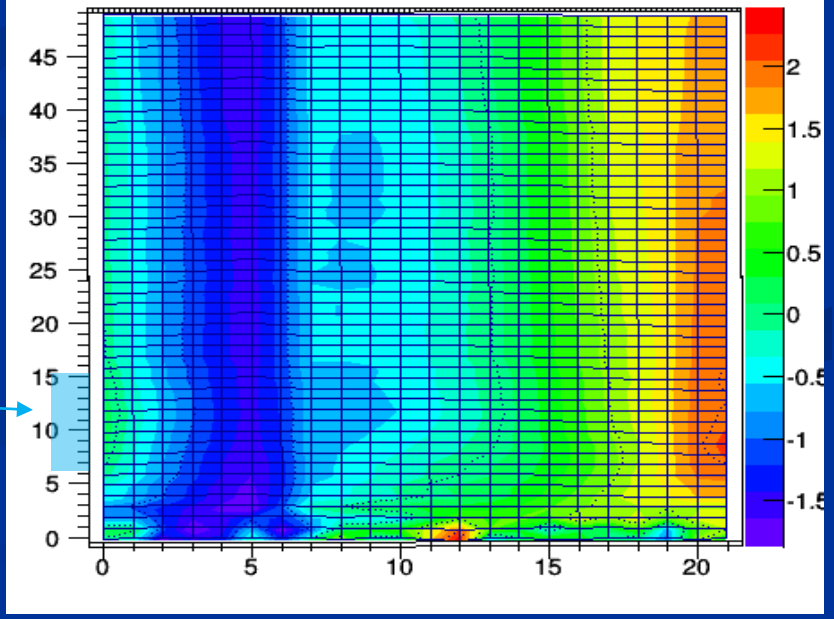




PC1



PC2

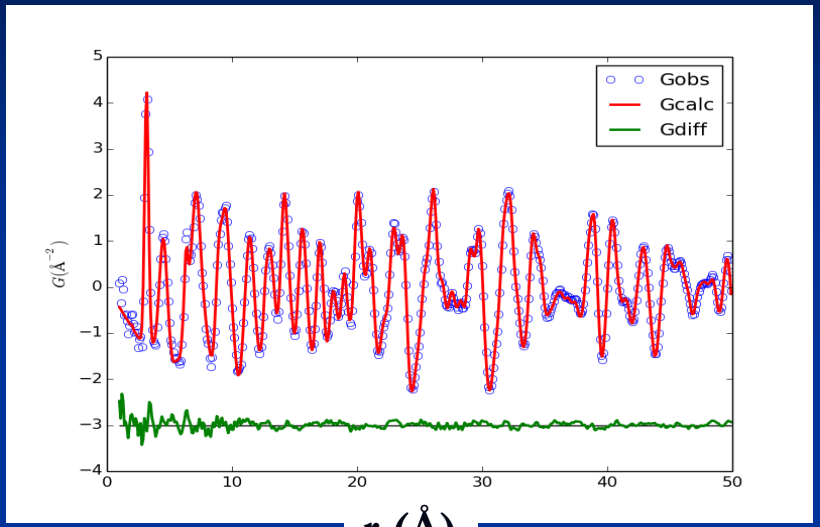


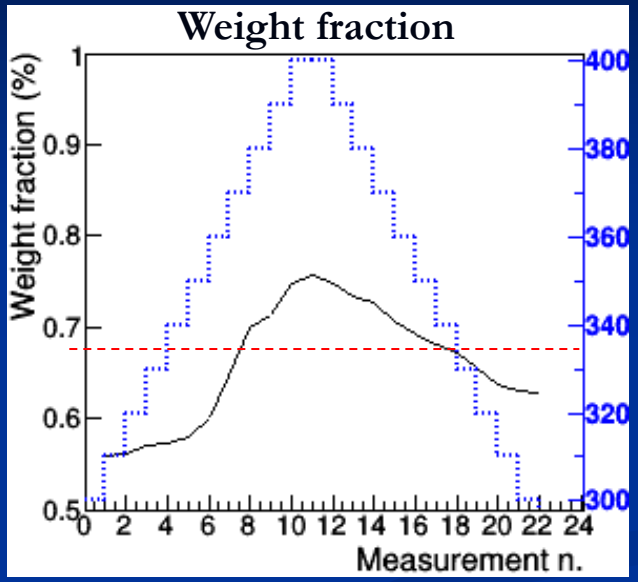
Local disorder from misaligned octahedra

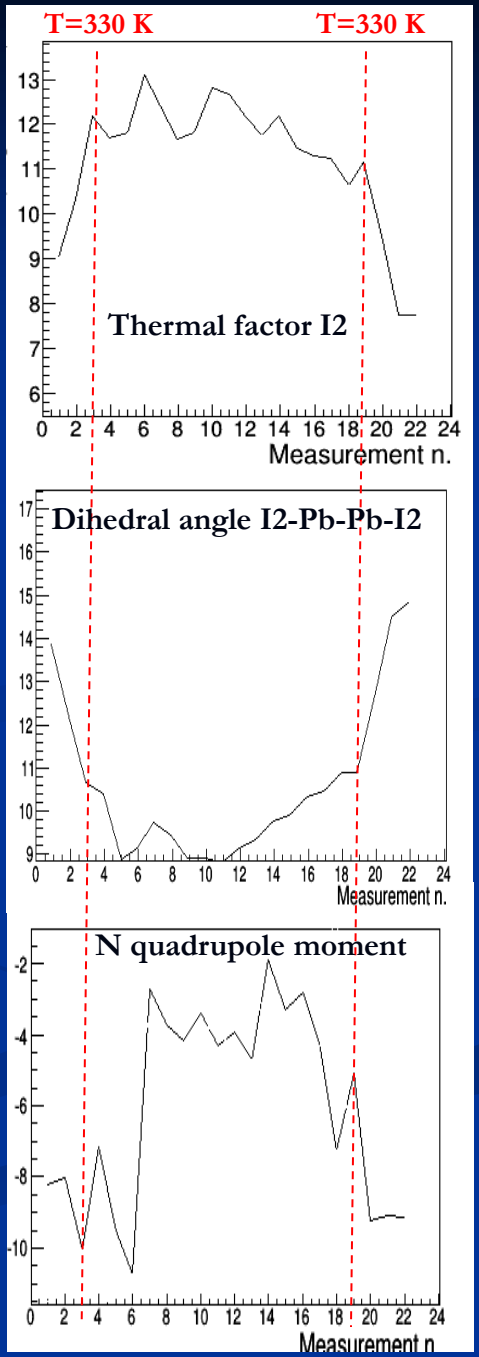
Anomalous contribution from nearest inter-octahedral ditstances

Fitting of individual profiles

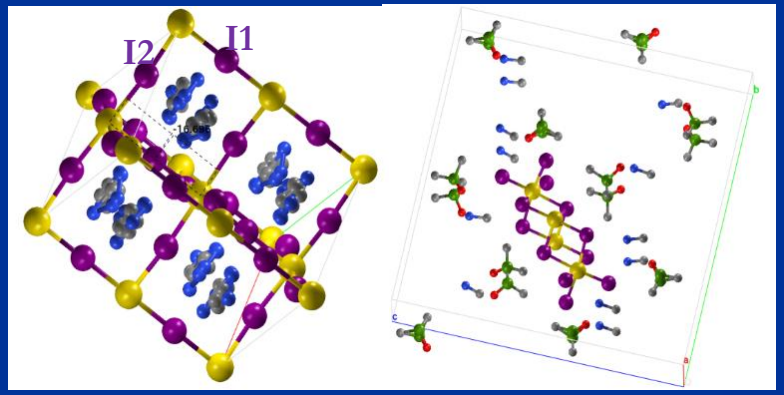
Automatic fitting of individual PDF profiles

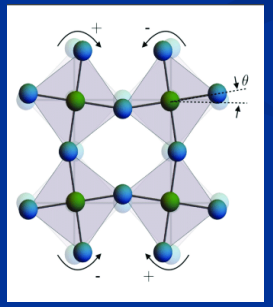




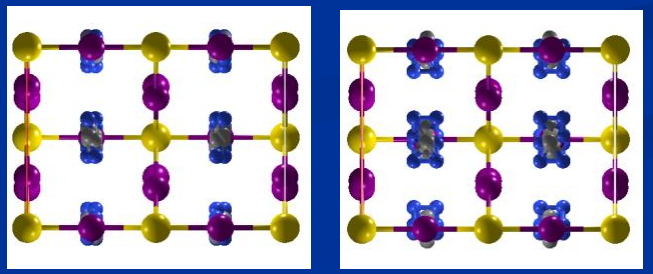


$MAPbI_3$ PbI_2-MAI-DMSO

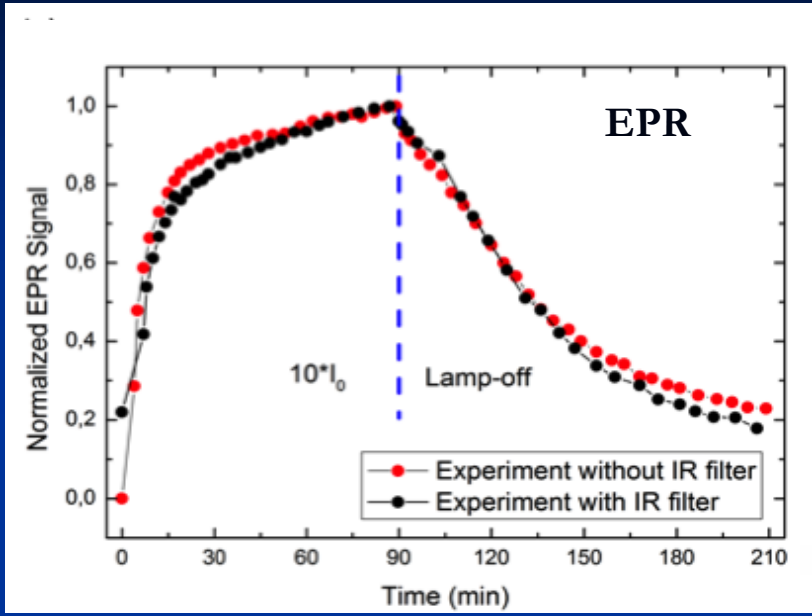




300K 400K



Dynamic investigations: illumination variations



ACS Energy Letters
Energy Letters 2018, 3 1840–1847
Cite This: ACS Energy Lett. 2018, 3, 1840–1847

Light-Induced Formation of Pb^{3+} Paramagnetic Species in Lead Halide Perovskites

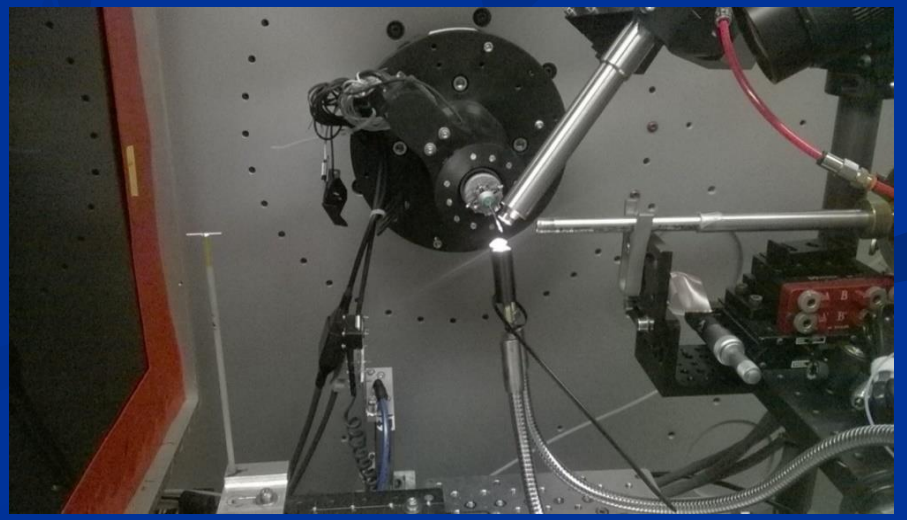
Silvia Colella,[†,‡] Michela Todaro,[#] Sofia Masi,[†,‡] Andrea Listorti,[†,‡] Davide Altamura,[∥] Rocco Caliandro,[*,∥] Cinzia Giannini,[∥] Elisa Carignani,[§] Marco Geppi,[§] Daniele Meggiolaro,[¶,▽] Gianpiero Buscarino,[*,#] Filippo De Angelis,[*,¶,▽] and Aurora Rizzo[‡]

Reversible generation, under illumination, of electron paramagnetic resonance signal from $MAPbI_3$ perovskite polycrystalline powder

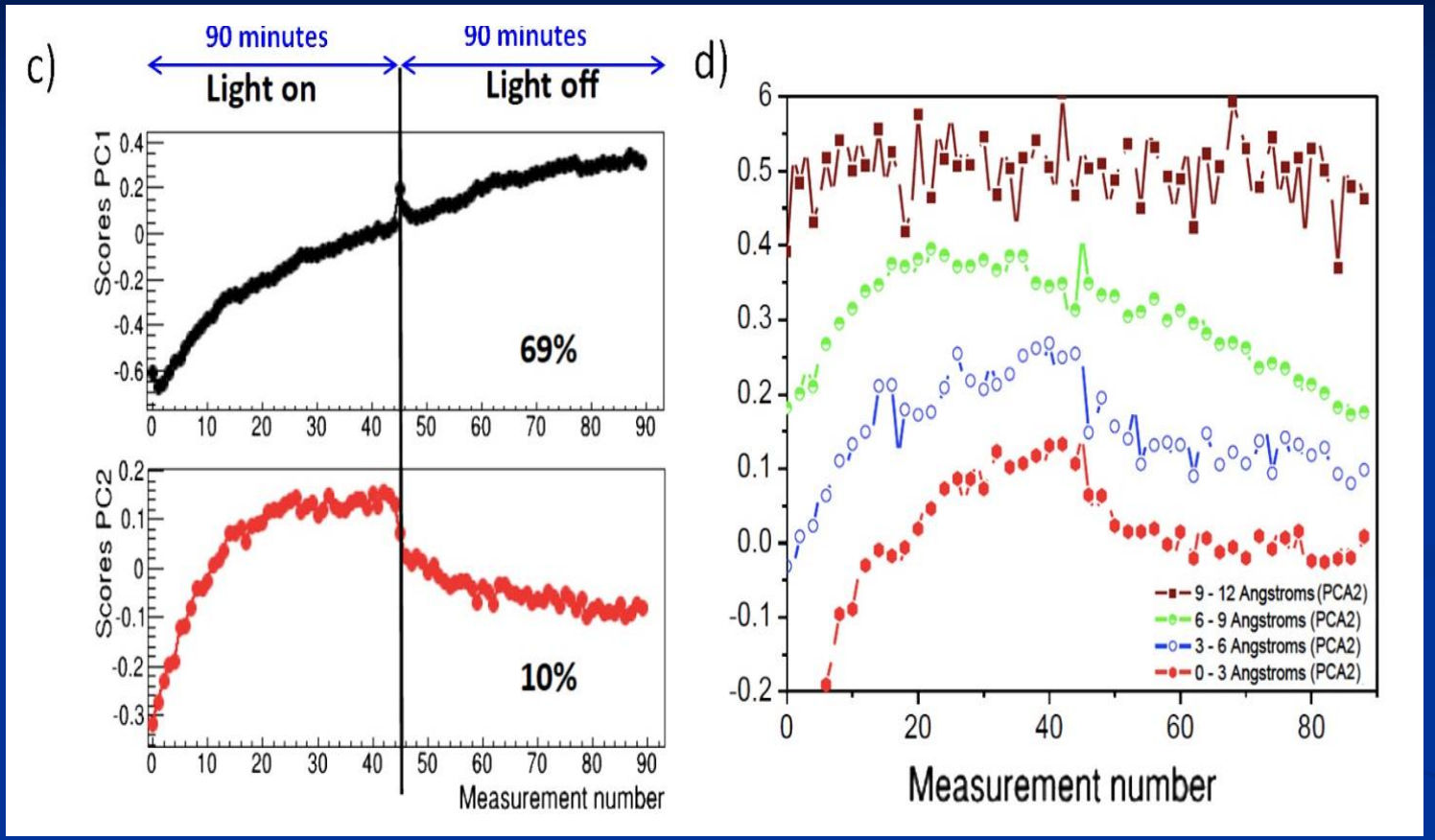
Sample illuminated

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

- Caliandro et al. Investigating temperature-induced structural changes of lead halide perovskites by in situ X-ray powder diffraction (2019) *J. Appl. Cryst.* 52
- Guccione et al. Improved multivariate analysis for fast and selective monitoring of structural dynamics by in situ X-ray powder diffraction (2018) *Phys. Chem. Chem. Phys* 20
- Guccione et al. Principal component analysis for automatic extraction of solid-state kinetics from combined in situ experiments (2018) *Phys. Chem. Chem. Phys* 29
- Palin et al. Rational design of the solid-state synthesis of materials based on poly-aromatic molecular complexes (2016) C*rystEngComm* 31
- Palin et al. Chemical selectivity in structure determination by time dependent analysis of in situ XRPD data: a clear view of Xe thermal behavior insdie a MFI zeolite (2015) *Phys. Chem. Chem. Phys.* 26
- Caliandro et al. Multivariate analysis of quaternary carbamazepine–saccharin mixtures by X-ray diffraction and infrared spectroscopy (2013) *J. Pharm. Biom. Anal.* 78–79
- 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