Evaluation of ED-XRF spectra and MA-XRF datasets

Handling the very large spectral datasets resulting from MA-XRF scanning

P. Van Espen[1]
V. Osorio[1,2] & S. Legrand[1]

[1]AXES,University of Antwerpen
[2]Brightspec NV

MA-XRF scanning in Conservation,
Art and Archeology

Tutorial day

ICTP, Sept 24, 2017

MA-XRF has revolutionised the way we handle (store and evaluate) energy-dispersive X-ray spectra

revolution |ˌrevəˈlo͞oSH(ə)n|
noun

1 a forcible overthrow of a government or social order in favour of a new system.

- (the Revolution)the American Revolution.
- (often the Revolution)(in Marxism) the class struggle that is expected to lead to political change and the triumph of communism
- **a dramatic and wide-reaching change in the way something works or is organised or in people's ideas about it: *marketing underwent a revolution*.**

X-ray analysis underwent a revolution

Dealing with MA-XRF data

Outline

Storage and retrieval of spectra

Some ideas
An implementation

Evaluation of spectra

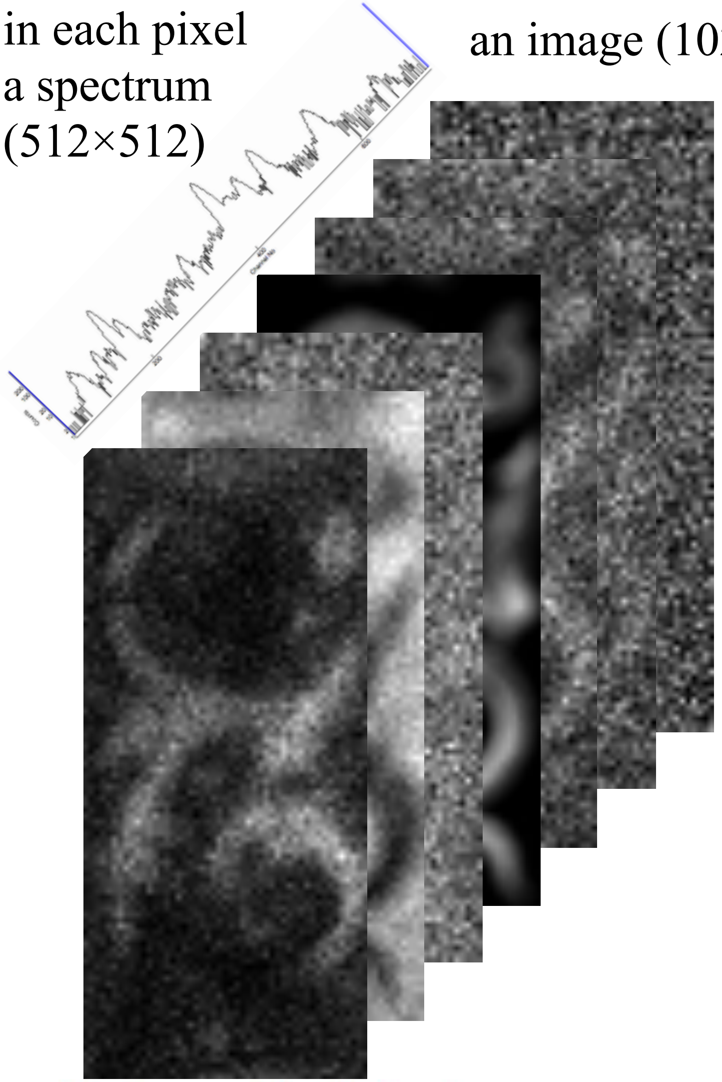
Some ideas
An implementation

Introduction

What do we have.

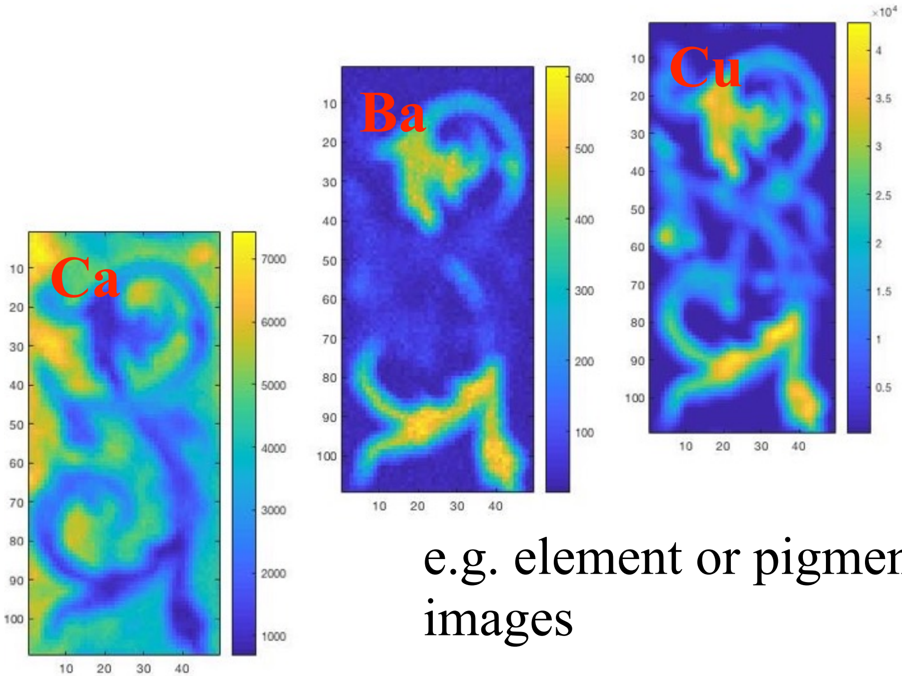
in each pixel
a spectrum
(512×512)

in each channel
an image (1024)



What do we want to achieve?

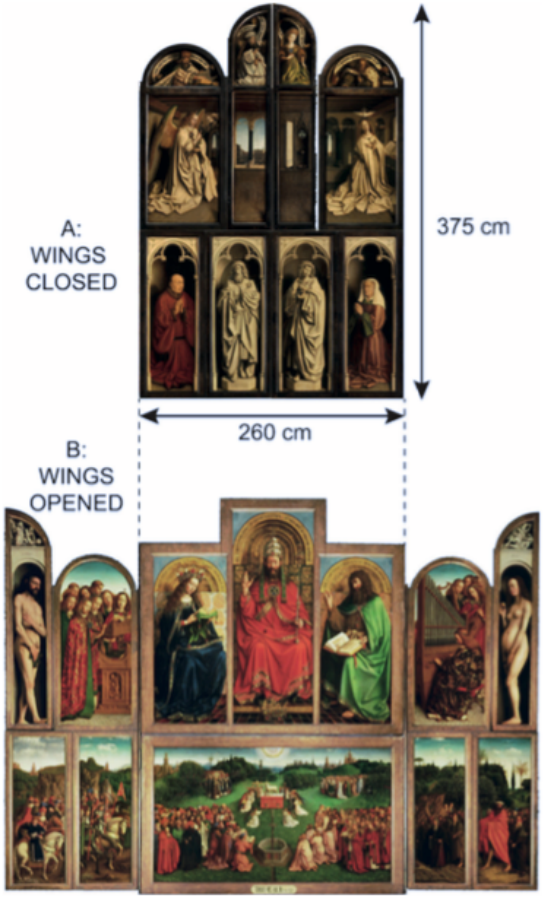
Make *informative* images based on the acquired spectrum in each pixel



e.g. element or pigment images

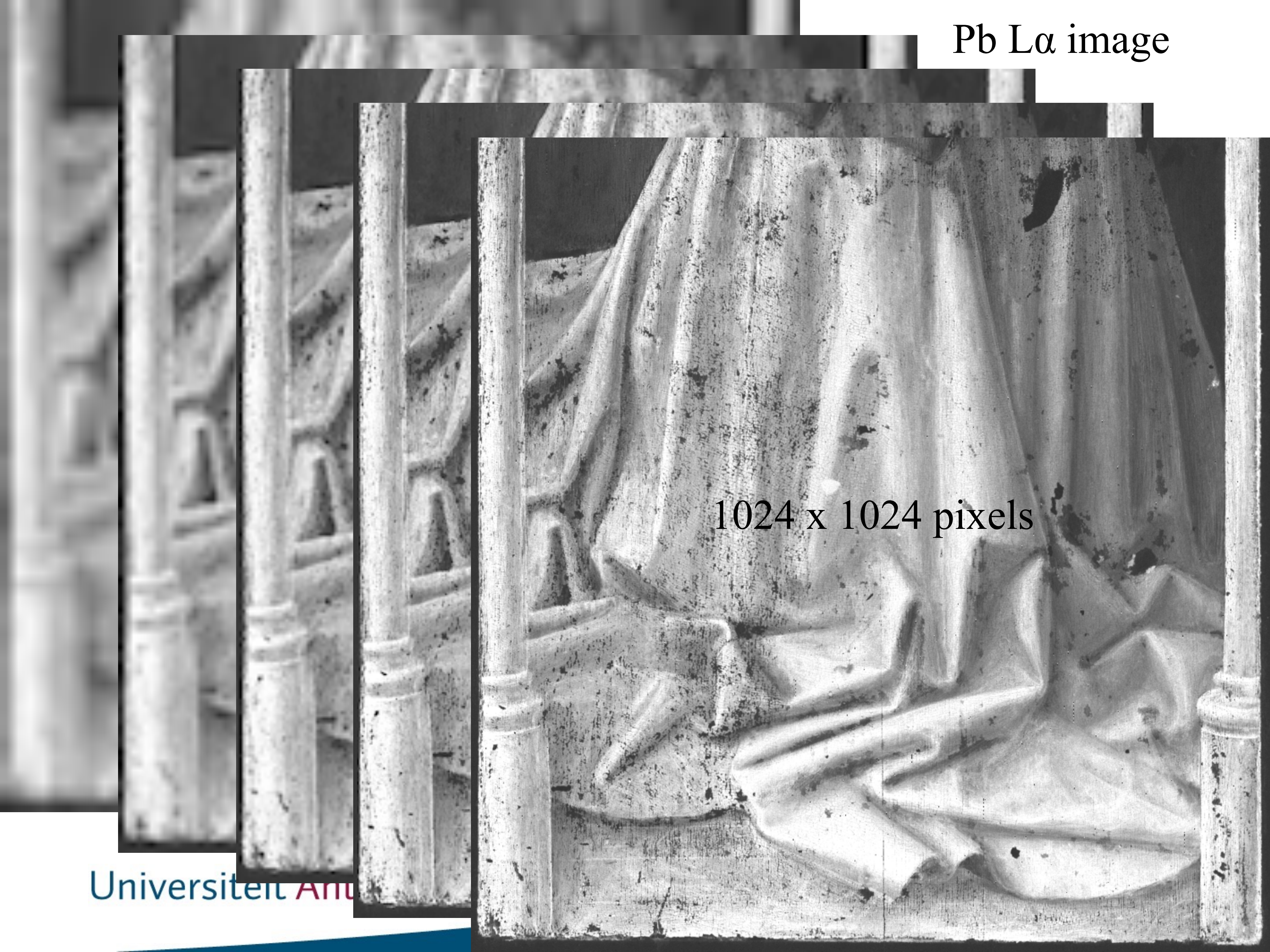
Example   Scanning "large" objects

Panel of the Ghent Altarpiece "The Adoration of the Lamb" painted by the brothers Van Eyck around 1432



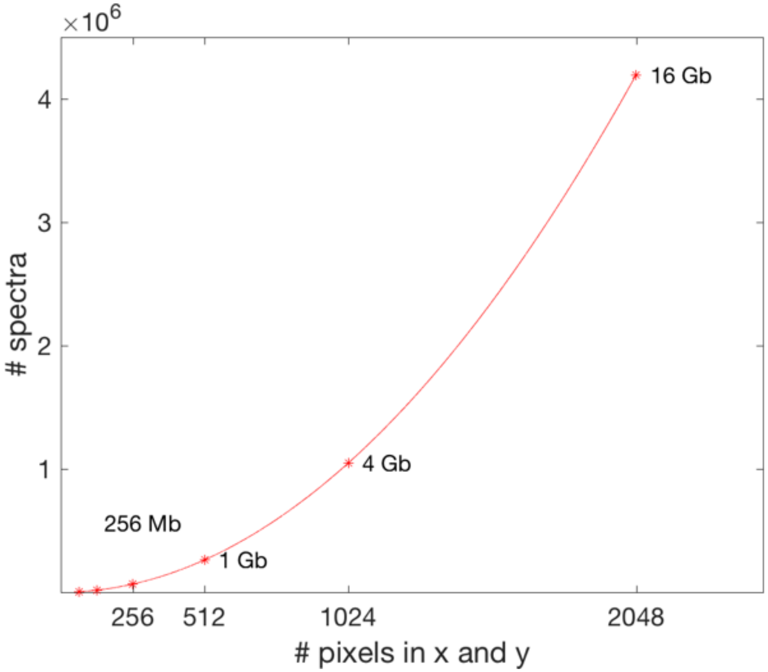
Ref: Geert Van der Snickt et. al.
Angew. Chem. Int. Ed. **2017**, 56, 4797 – 4801

Pb Lα image



The amount of data is HUGH

Number of spectra vs scan size



Scan size	# spectra	Amount of data*
64×64	4 096	16777216 = 16 Mb
128×128	16 384	67108864 = 64 Mb
256×256	65 536	268435456 = 256 Mb
512×512	262 144	1073741824 = 1 Gb
1024×1024	**1 048 576**	**4294967296 = 4 Gb**
2048×2048	**4 194 304**	**17179869184 = 16 Gb**

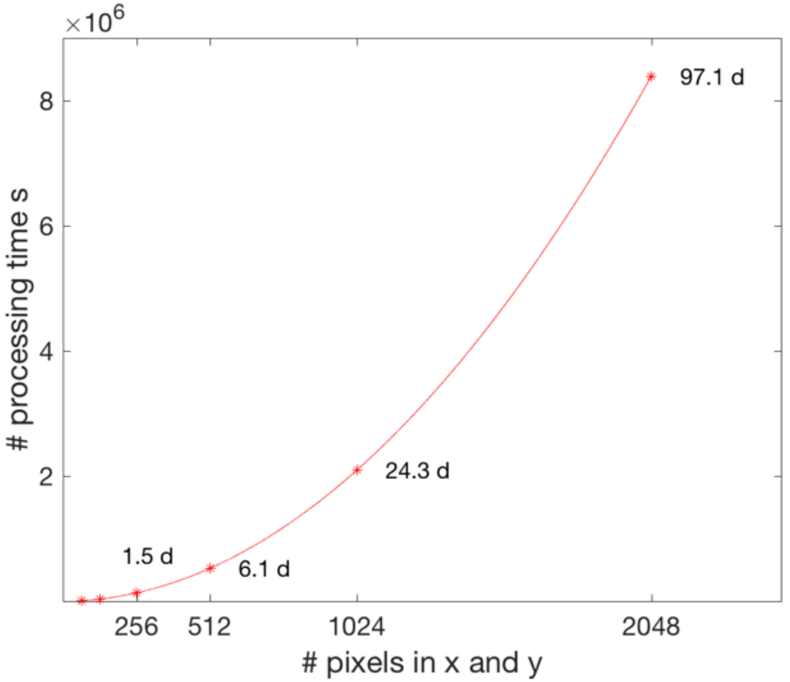
* assuming 1024 channels and 4 bytes/ch

storage and retrieval problem!!!

Universiteit Antwerpen

The time required to evaluate is Enormous

Time needed to analyse the spectra



Scan size	# spectra	time*
64×64	4 096	137 min = 2.4 h
128×128	16 384	546 min = 9.1 h
256×256	65 536	36.4 h = 1.5 d
512×512	**262 144**	**145 h = 6 d**
1024×1024	**1 048 576**	**582.5 = 24.3 d**
2048×2048	**4 194 304**	**97.1 d = 3.2 months**

*assuming 2 s per spectrum

evaluation problem!!!

We have a space-time problem

Unbinding space from time could solve the biggest conundrums in the Cosmos.

Anil Ananthaswany

conundrum |kəˈnəndrəm|

noun (pl. **conundrums**)

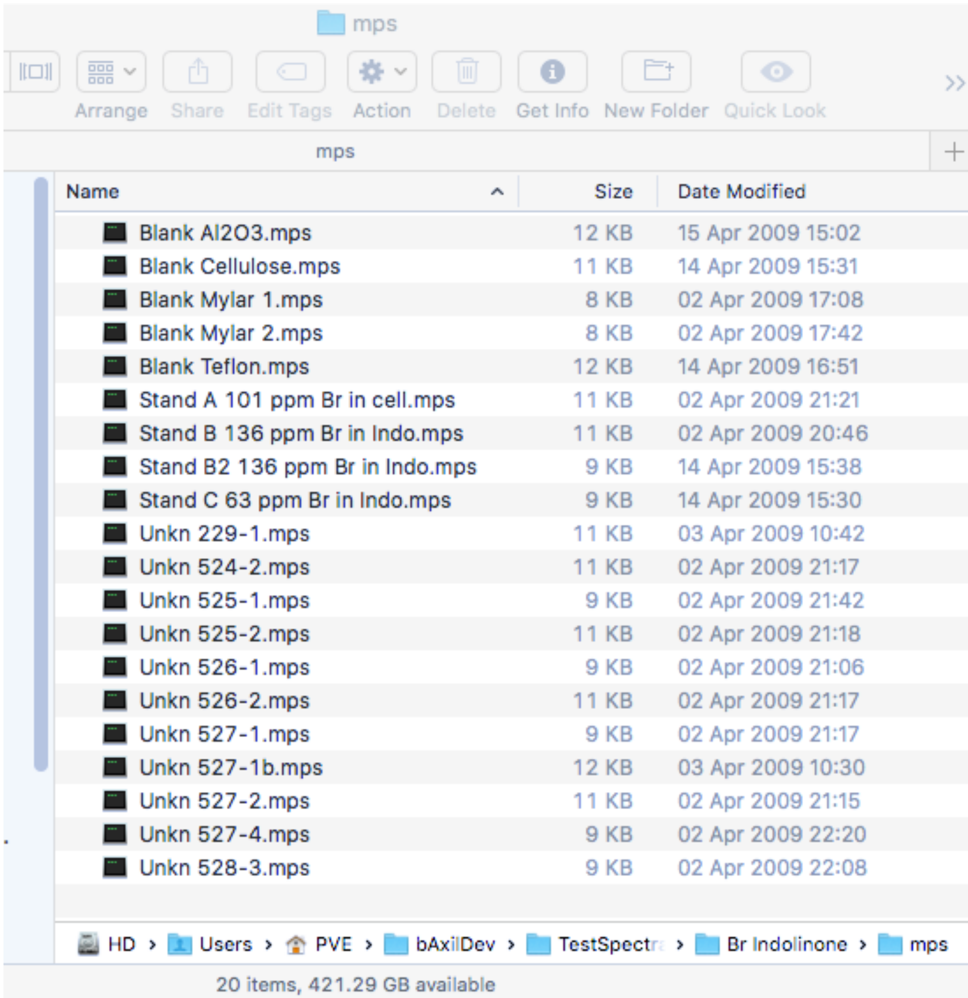
- a confusing and difficult problem or question: *one of the most difficult conundrums for the experts*.
- *a question asked for amusement, typically one with a pun in its answer; a riddle.*

Let's solve the conundrum!!!

Storage and retrieval

some ideas

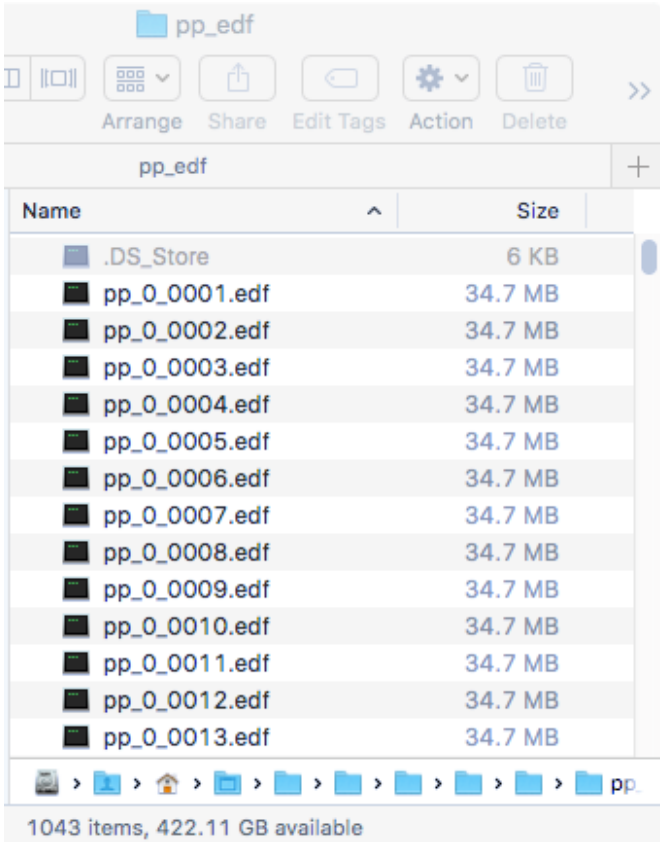
Storing the spectra



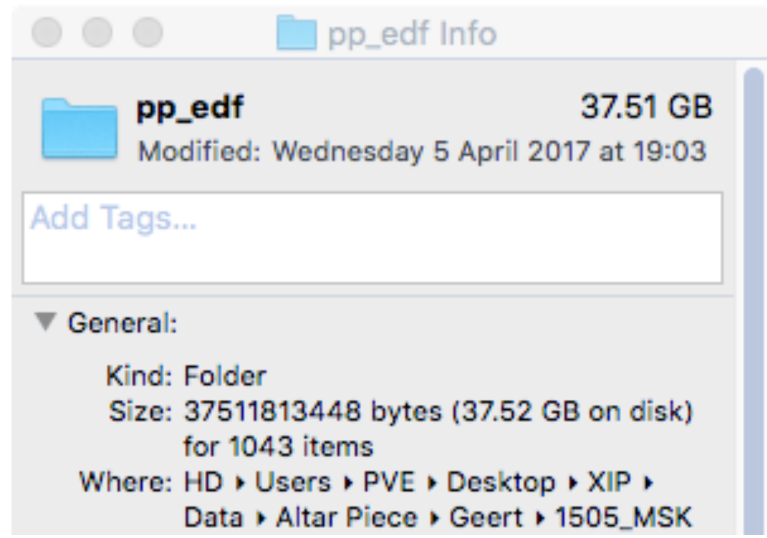
Historical: each spectrum one file

Serious file system performance degradation if more that 1000 files in one directory!!!

Better: one file per line scanned



~1000 spectra per file

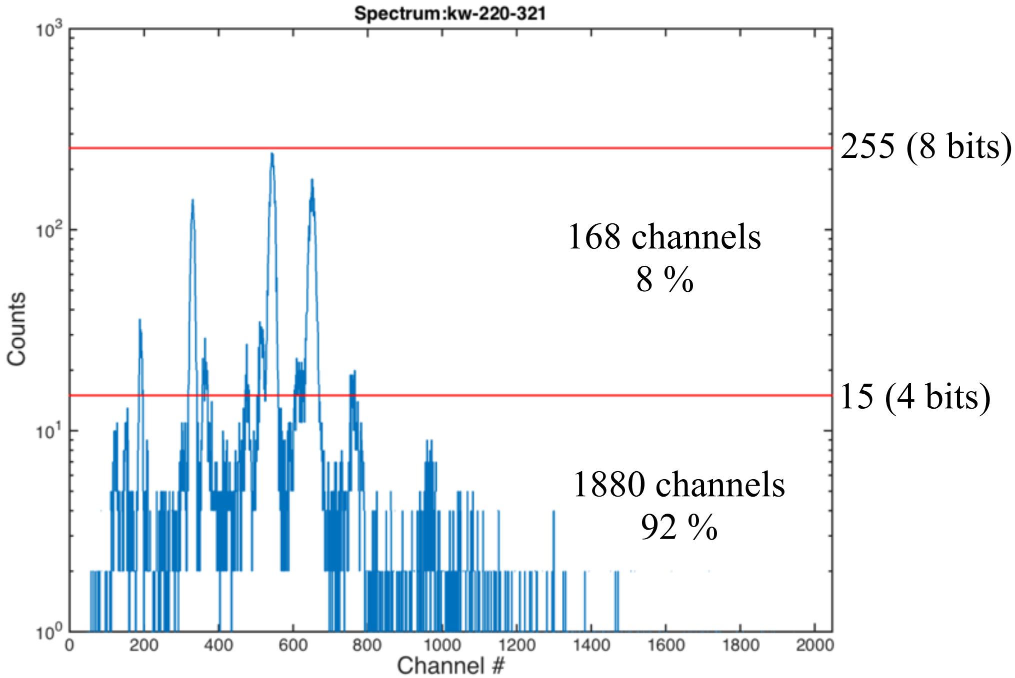


Retrieval of arbitrary spectrum (x,y) for inspection not trivial

Best ? Compressed in one file

Content of each channel stored in 4 bytes (values 0 — 4 294 967 295 counts)

One spectrum = 2048 channel = 8192 bytes = 65536 bits



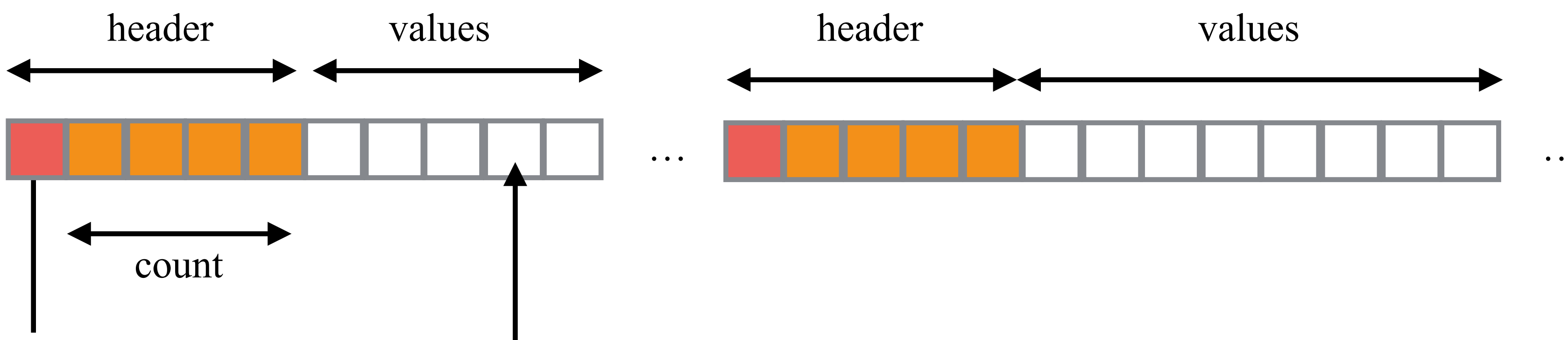
168×8 + 1880×4 = 1344 + 7520 = 8864 bits needed

factor: 65536 / 8864 = 7.4

Lossless spectrum compression

Original: constant length 2048 × 4 bytes

Variable length compression:
use ½ byte when possible, 1 byte if needed or 1½, or 2…



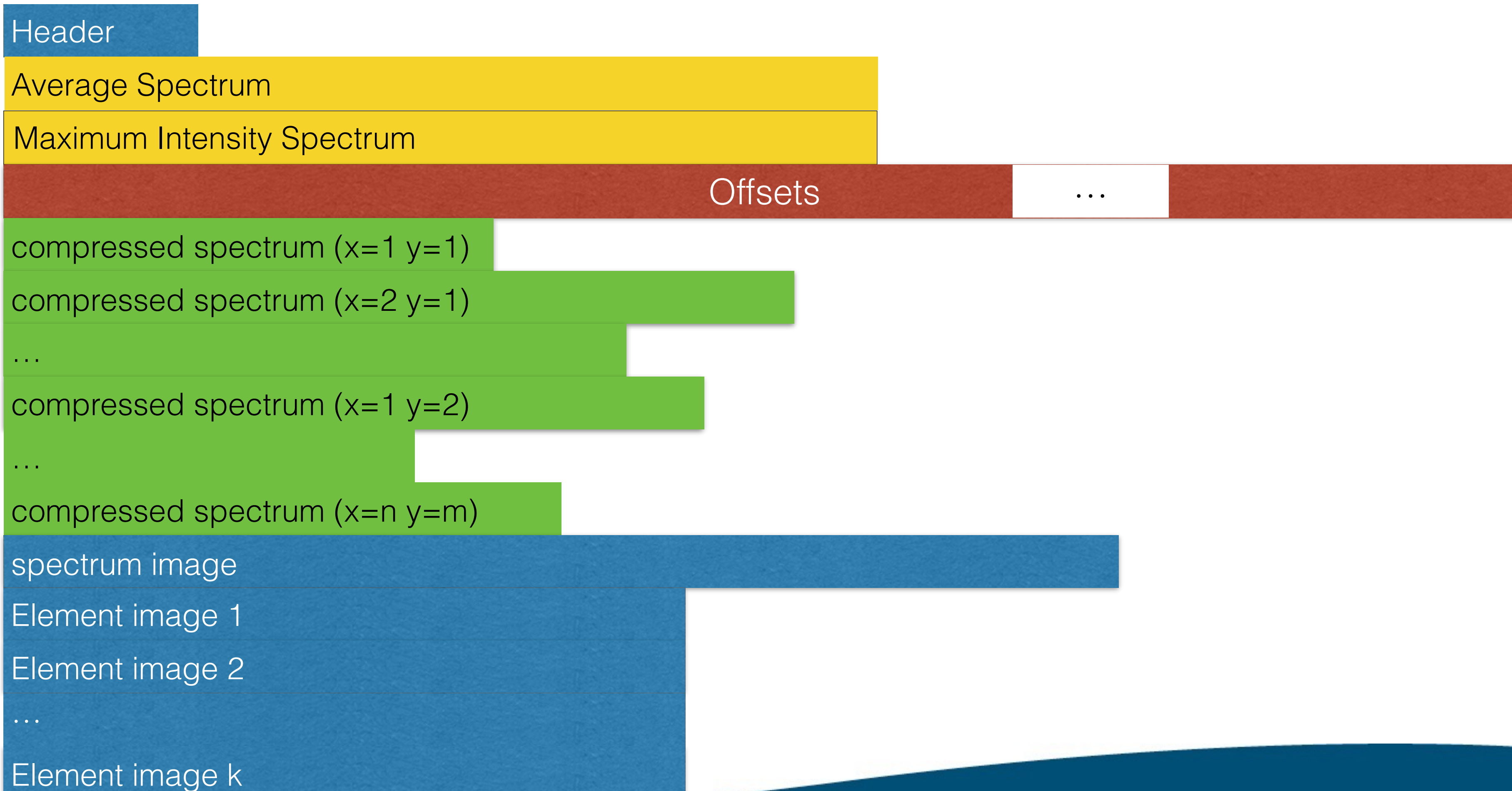
data size	values
0	0
1 nibble	1 — 15
2 nibbles	16 — 255
3 nibbles	
4 nibbles	1 — 65231
…	

nibble = ½ byte = 4 bits

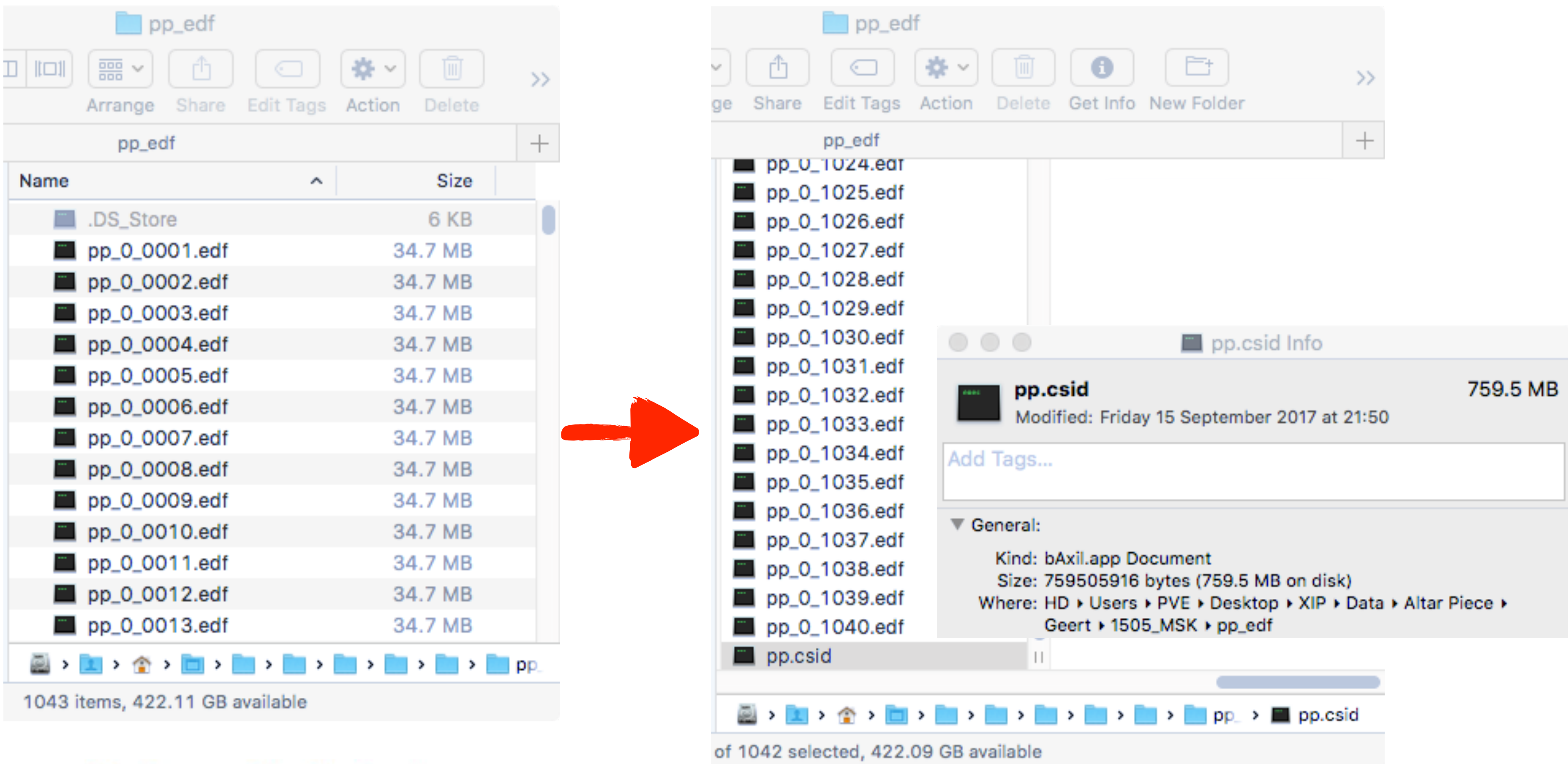
Storage of compressed spectra in one file

Original data (ascii, ESRF-EDF) → Compressed Spectroscopic Image Data (*.CSID)

special format for efficient storage and retrieval
makes **interactive** analysis possible



Directory of 37,511,813,448 bytes (37 Gb)
to a file of 759,505,916 bytes (759 Mb)
compression of a factor 47.5

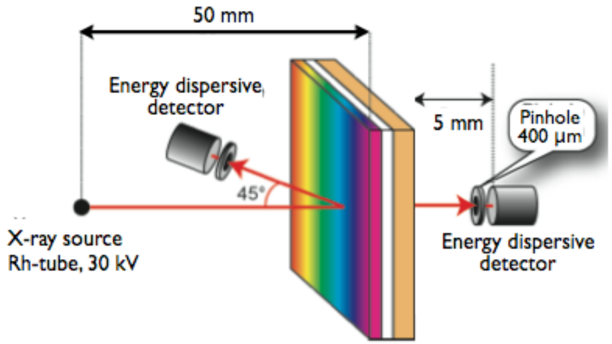


Storage and retrieval

implementation

Example: painting on oak panel by Pieter Eyskens

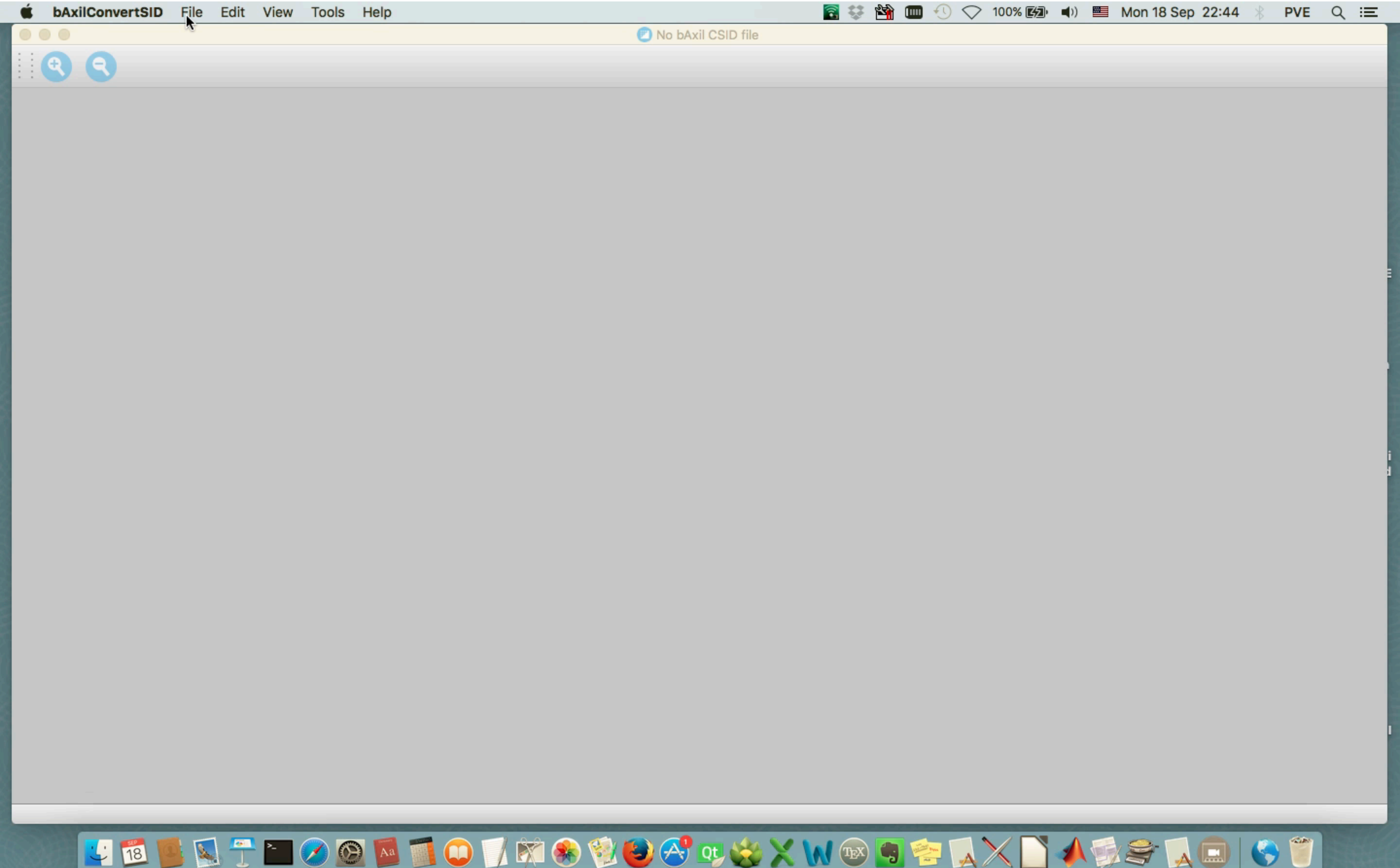
(mock-up)



Low Z white paint	Low Z coloured paint
Lead white in oil	
Lead white in oil	
Kaolin in hide glue	
Hide glue	
Panel in oak	

202 ×201 pixels
40602 spectra

Storage and retrieval as implemented in bAxilConvertSID



Evaluation

some ideas

Historical

Evaluation with user interaction of individual spectra

e.g. using computer programs like

- Axil, QXAS
- WinAxil, WinQXAS
- PyMCA
- bAxil
- …

Purpose relating spectral data to physical properties

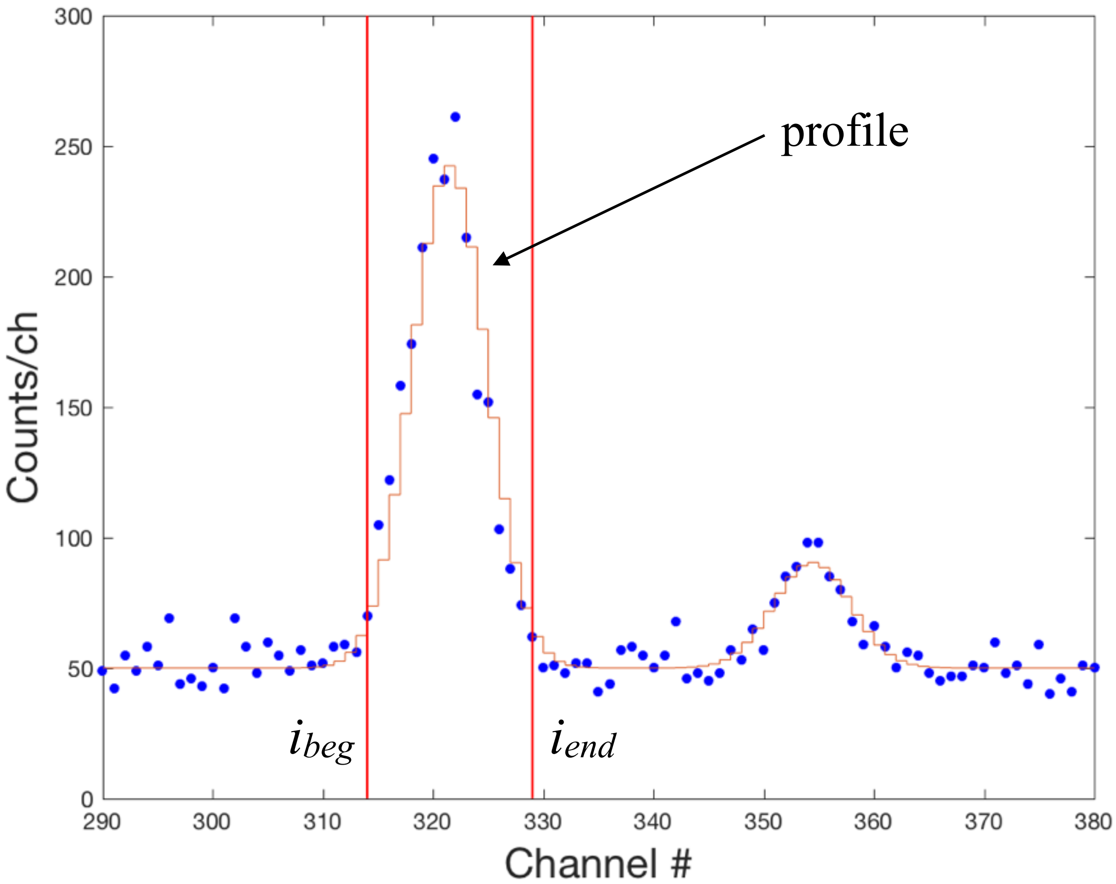
e.g. net peak area ⇔ amounts of the element

Different terms and techniques

Spectrum integration	Linear least squares
Fitting	Non-linear (iterative) least squares
~~deconvolution~~	Alternating least squares
…	

They all have something in common: use of a "profile"

Consider simple peak integration (region of interest, ROI)



$$A_{\mathrm{Fe}\,Ka} = \sum_{i=i_{beg}}^{i=i_{end}} y_i$$

integration limits are based an assumed profile

Characteristics of the profile

Shape	Gaussian	
Parameters	position (i_p) width (s_{ch})	non-linear parameters
	height (area, A)	linear parameter

mathematical

in channels

$$y(i) = \frac{A}{s_{ch}\sqrt{2\pi}} \exp\left[-\frac{(i-i_p)^2}{2s_{ch}^2}\right]$$

s_{ch} width (sigma) of peak in channels

$$s_{ch} = \frac{\text{FWHM}}{2\sqrt{2\ln 2}} = \frac{\text{FWHM}}{2.35}$$

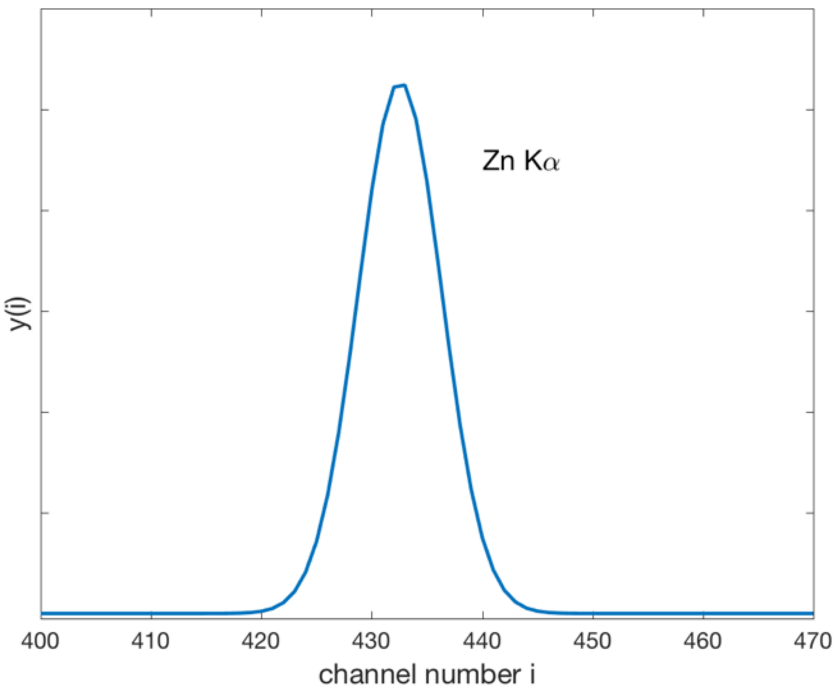
in keV

$$y(i) = \frac{A \times Gain}{s_E\sqrt{2\pi}} \exp\left[-\frac{(E_i-E_p)^2}{2s_E^2}\right]$$

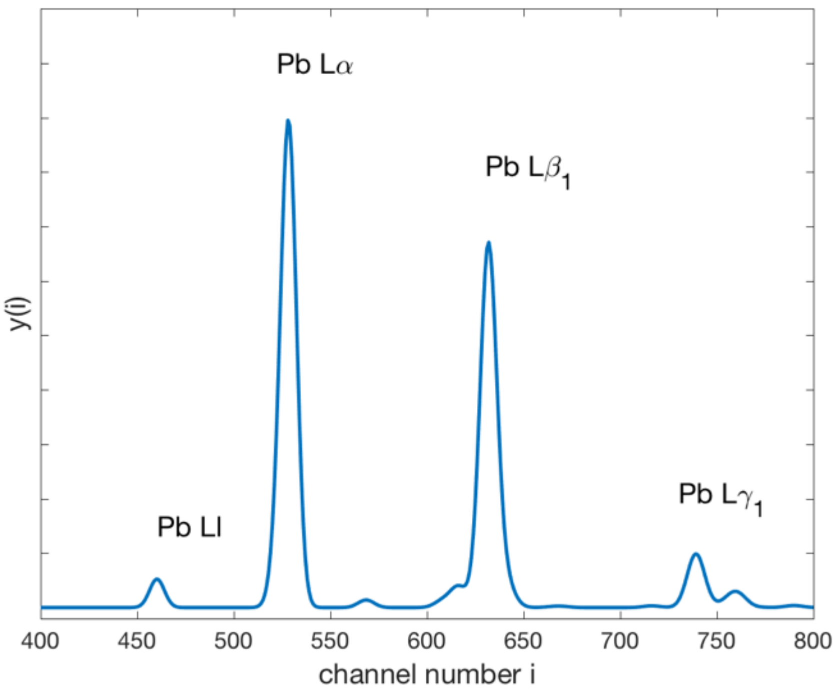
s_E width of peak in keV

E_i energy of channel i, E_i = zero + gain×i

Profile of an X-ray line (Zn Kα)



Profile of an element (Pb)



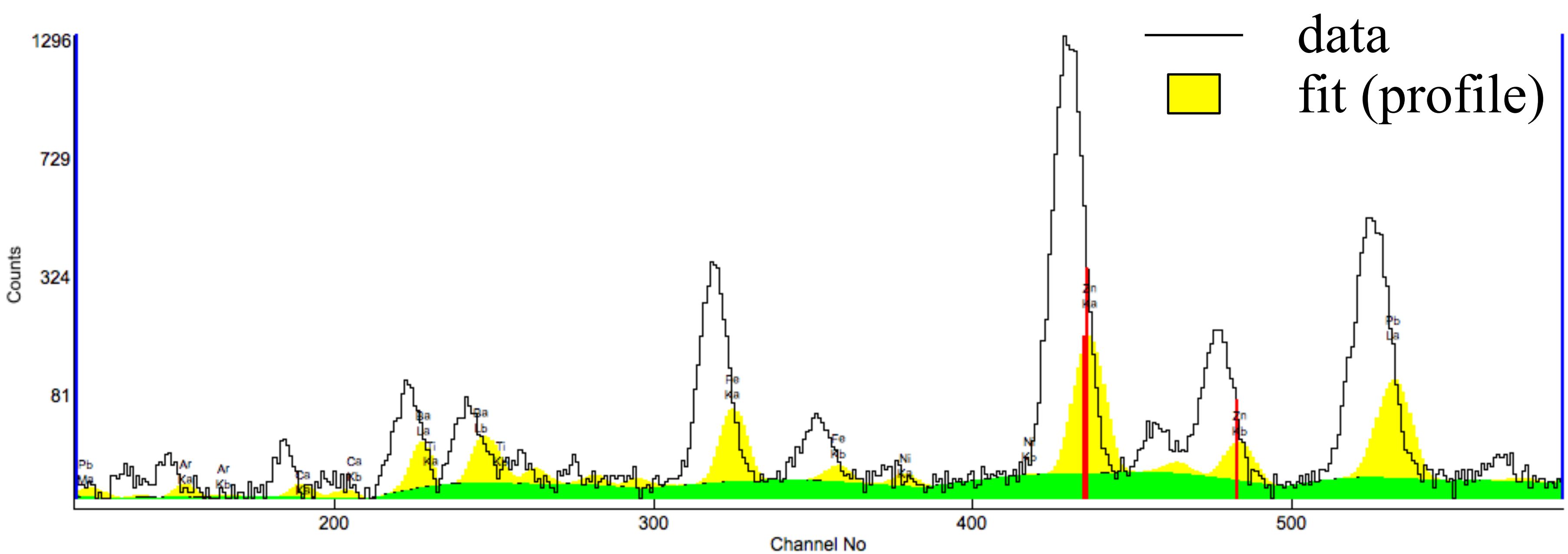
different spectrum evaluation methods = different applications of the profiles

1 "Traditional" non-linear least squares

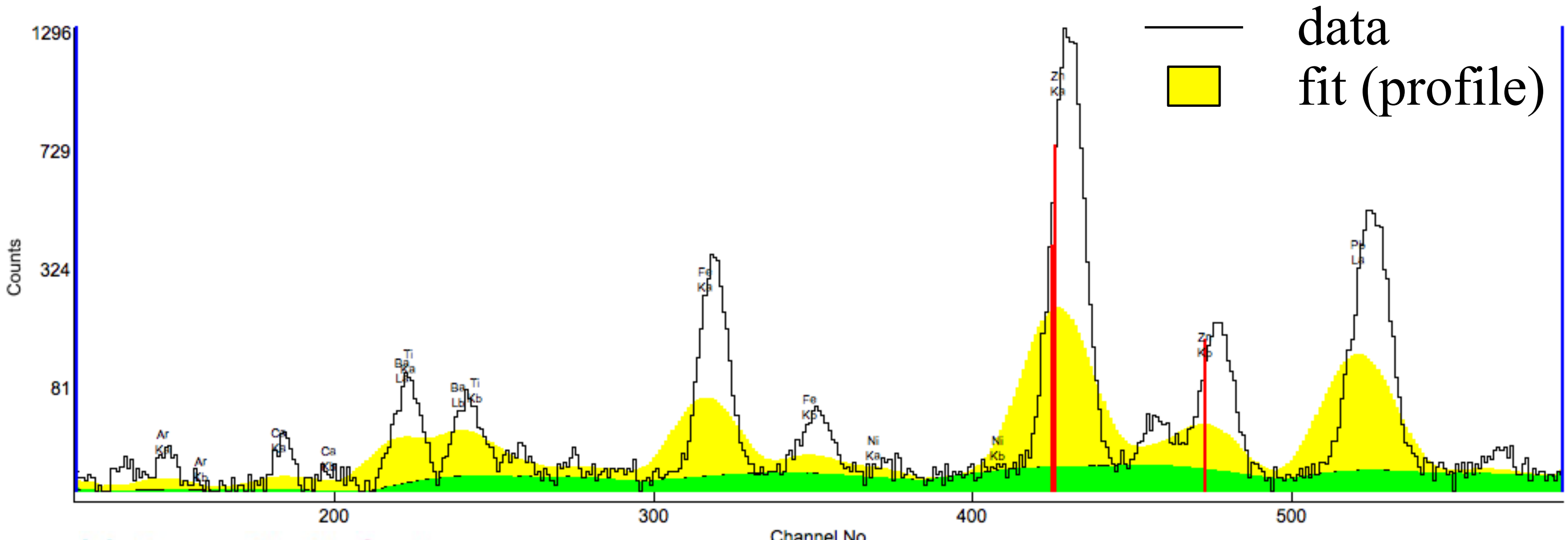
Iterative proces

Adjusting the profiles

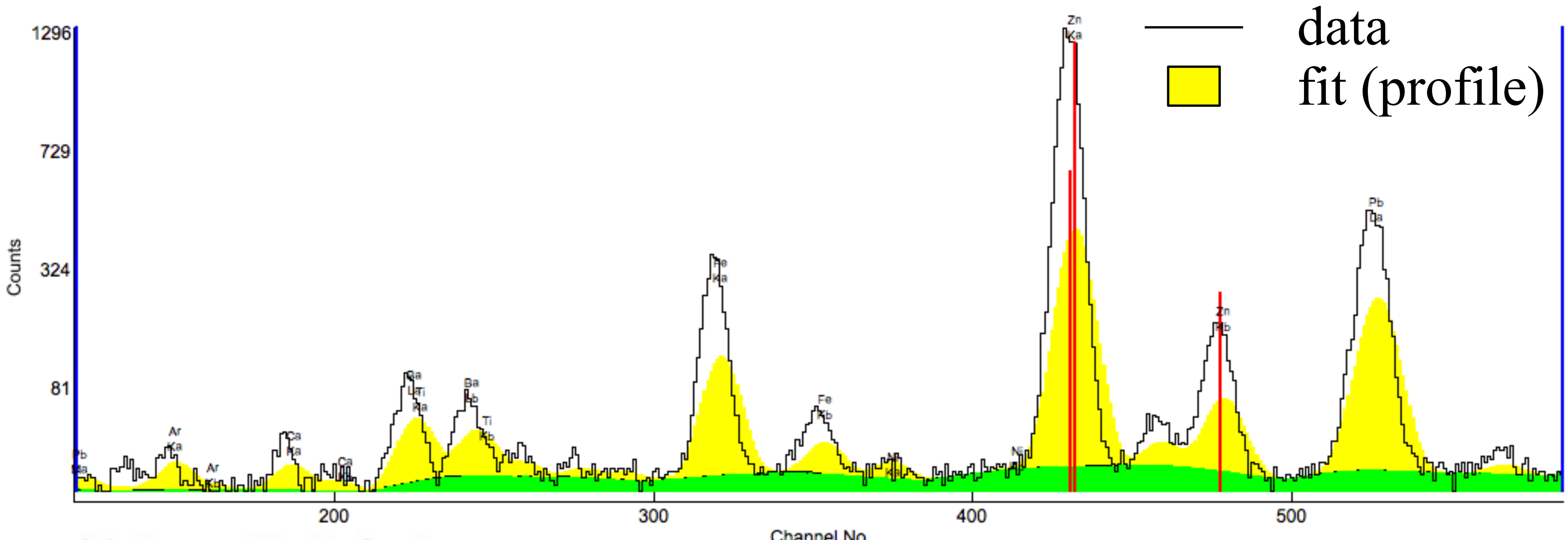
n	χ^2	Zero eV	A_{Fe}	A_{Zn}
1	44.6	-100.00	469	1602



n	χ^2	Zero ev	A_{Fe}	A_{Zn}
1	44.6	-100.00	469	1602
2	44.5	-0.98	1224	4495

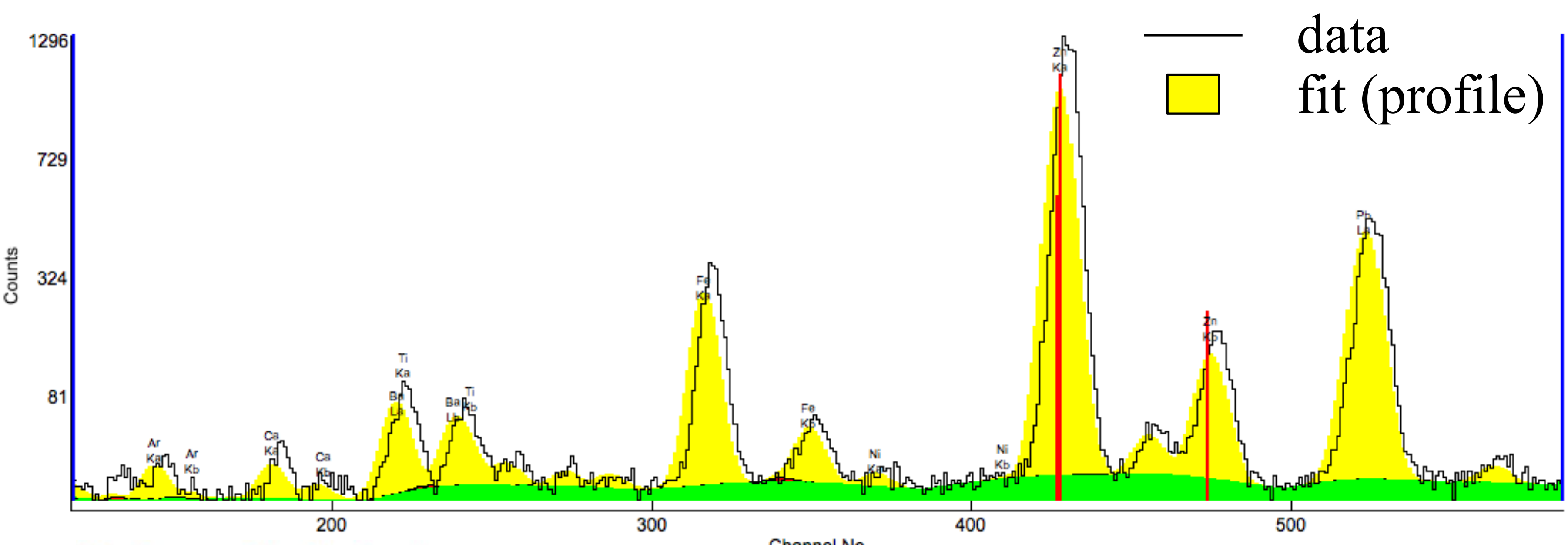


n	χ^2	Zero ev	A_{Fe}	A_{Zn}
1	44.6	-100.00	469	1602
2	44.5	-0.98	1224	4495
3	22.6	-62.62	1500	5607

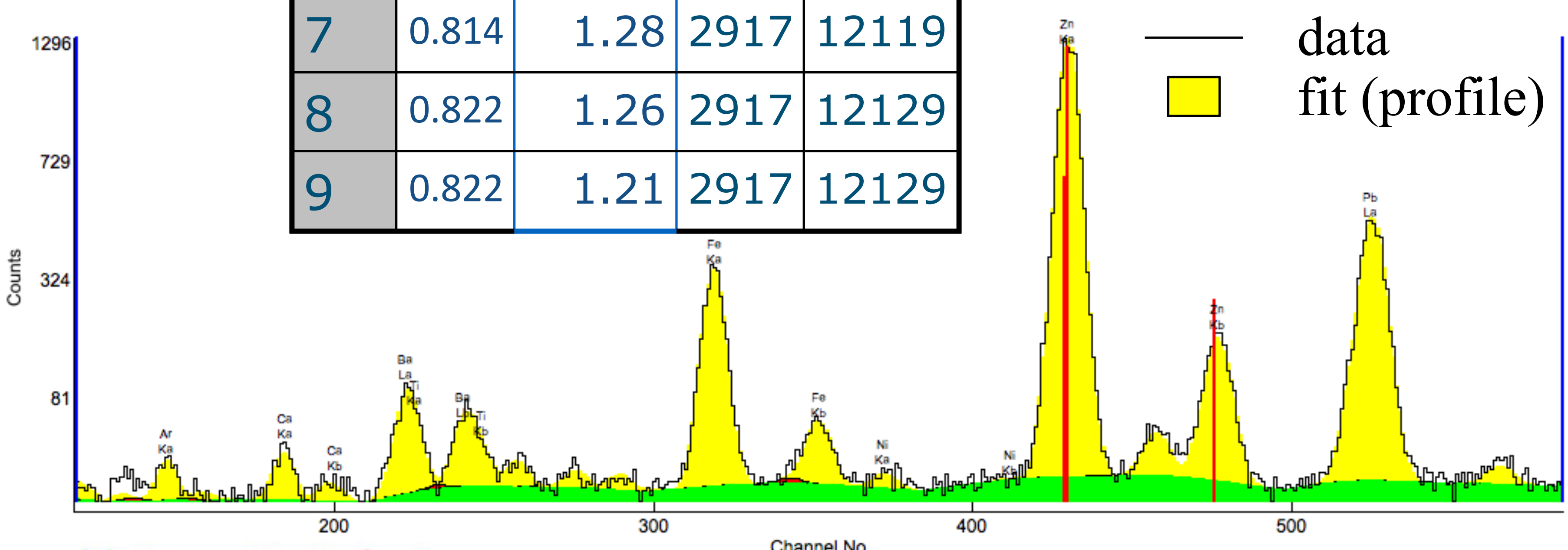


Universiteit Antwerpen

n	χ^2	Zero ev	A_{Fe}	A_{Zn}
1	44.6	-100.00	469	1602
2	44.5	-0.98	1224	4495
3	22.6	-62.62	1500	5607
4	18.9	71.03	2680	10922



n	χ^2	Zero	A_{Fe}	A_{Zn}
1	44.6	-100.00	469	1602
2	44.5	-0.98	1224	4495
3	22.6	-62.62	1500	5607
4	18.9	71.03	2680	10922
5	4.97	-3.64	2883	12005
6	1.22	3.81	2867	12013
7	0.814	1.28	2917	12119
8	0.822	1.26	2917	12129
9	0.822	1.21	2917	12129



Non-linear Least Squares

- Takes profile changes into account
  (energy and resolution calibration)
- Account for the continuum (background) present
- Allows to model other artefacts (sum peaks)

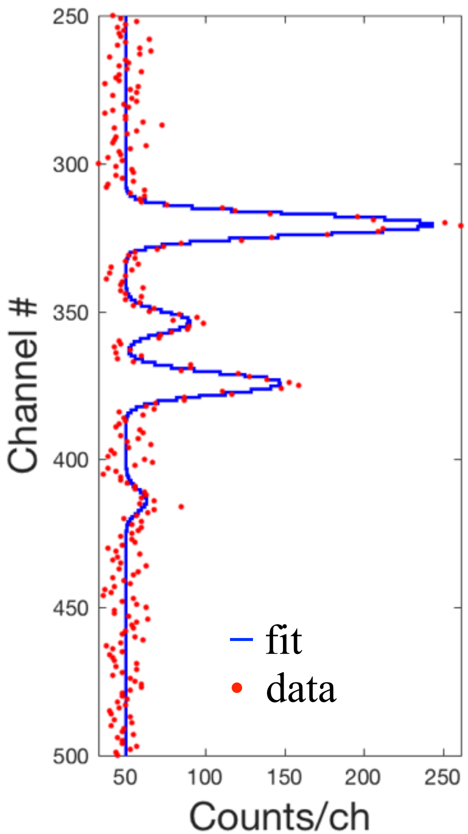
BUT

- Is inherently slow
  mathematical complexity and iterative
  (Marquardt - Leverberg algorithm)

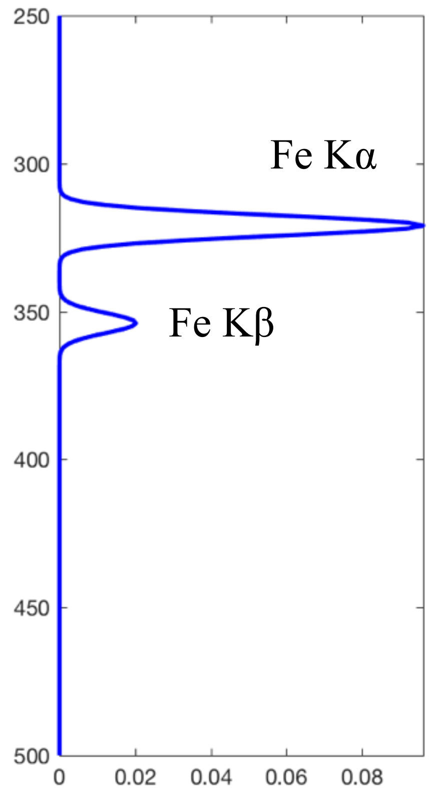
2. Linear Least Squares

position and width of peaks know
estimate the area

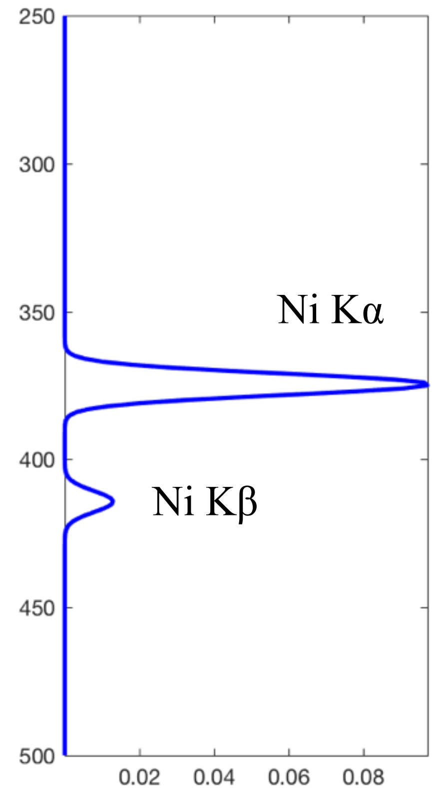
$$y(i) = A_{\mathrm{Fe}}\, p_{\mathrm{Fe}}(i) + A_{\mathrm{Ni}}\, p_{\mathrm{Ni}}(i)$$



$= a_{\mathrm{Fe}}$



$+\ a_{\mathrm{Ni}}$



is like "fitting a straight line"

"Simple" matrix algebra

$$y(i) = A_1 P_1(i) + A_2 P_2(i) + \ldots + A_n P_m(i)$$

$$\mathbf{y}_{n\times 1} = \mathbf{P}_{n\times m}\mathbf{A}_{m\times 1}$$

General linear model

Least squares solution

Finding the A_j's

$$\mathbf{A} = (\mathbf{P}^T\mathbf{P})^{-1} \times \mathbf{P}^T\mathbf{y}$$

Direct solution using matrix multiplication

- Fast

BUT

- Systematic errors if energy or resolution calibration changes
- Cannot take continuum (background) into account
- Cannot deal with non-linear effects (absorption, sum peaks)

Summary

Level of knowledge about the profile	Need to estimate	Method	Comment
I know: shape	Area, position and width	Non-linear least squares	Slow
I know: shape, position and width	Area	Linear least squares	Vulnerable to systematic errors
"I know nothing"	Shape and area	Multivariate statistics PCA, PLS, ALS	Results need interpretation

position and width = energy and resolution calibration

Combining speed with accuracy

Hybrid least squares fitting HLS

At start — Use non-linear least squares to determine the profiles $p_j(i)$
(is energy and resolution calibration)

Calculate the matrix $$\mathbf{M} = \left(\mathbf{P}^T\mathbf{P}\right)^{-1}\mathbf{P}^T$$

For each spectrum $\boldsymbol{y}$

Remove the non-additive continuum (stripping)

Calculate the elemental contributions A_j $$\mathbf{A} = \mathbf{M}\mathbf{y}$$

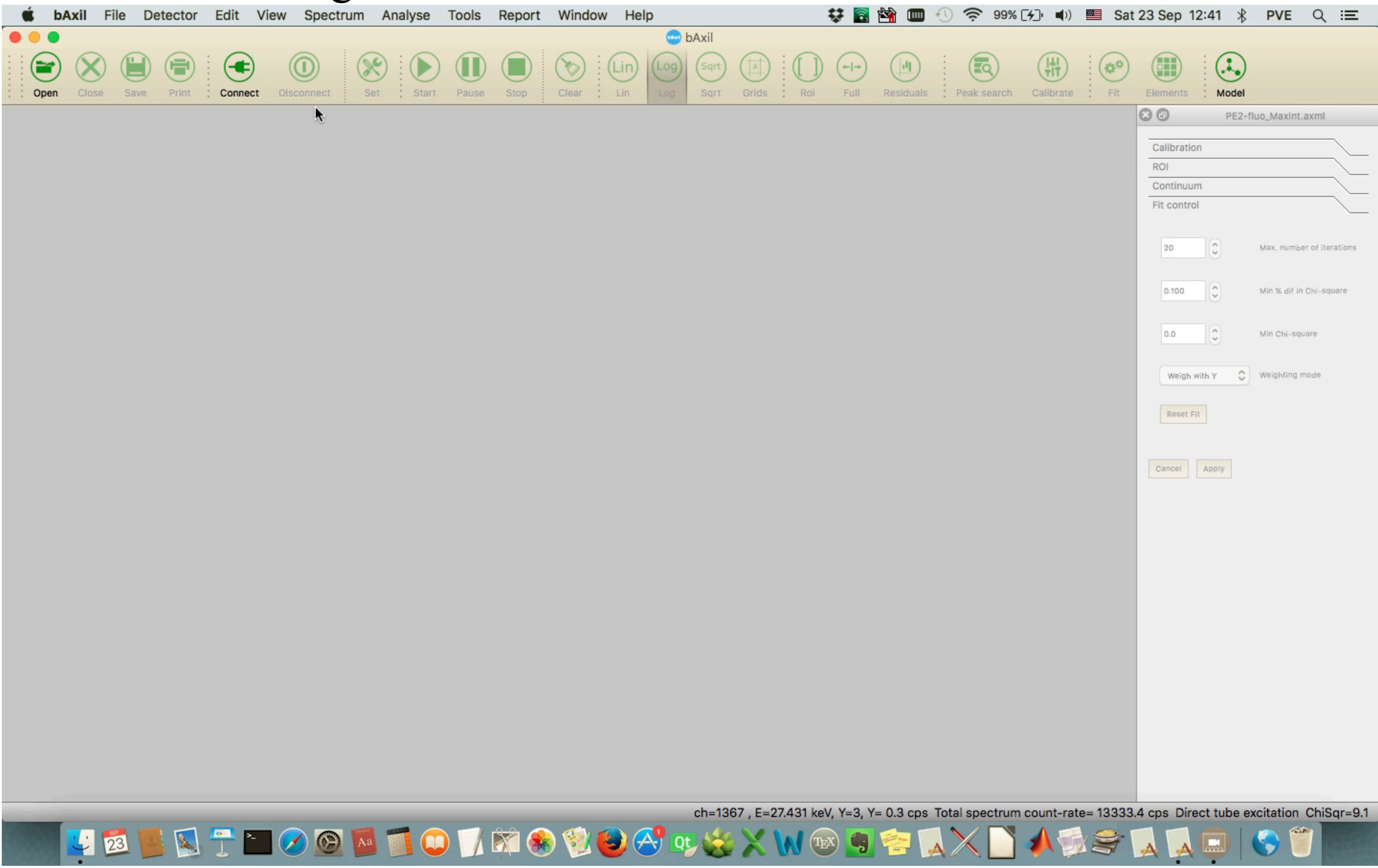
Repeat non-linear fitting if necessary

ONE matrix multiplication per spectrum!!!

Evaluation

implementation

Evaluation using bAxil

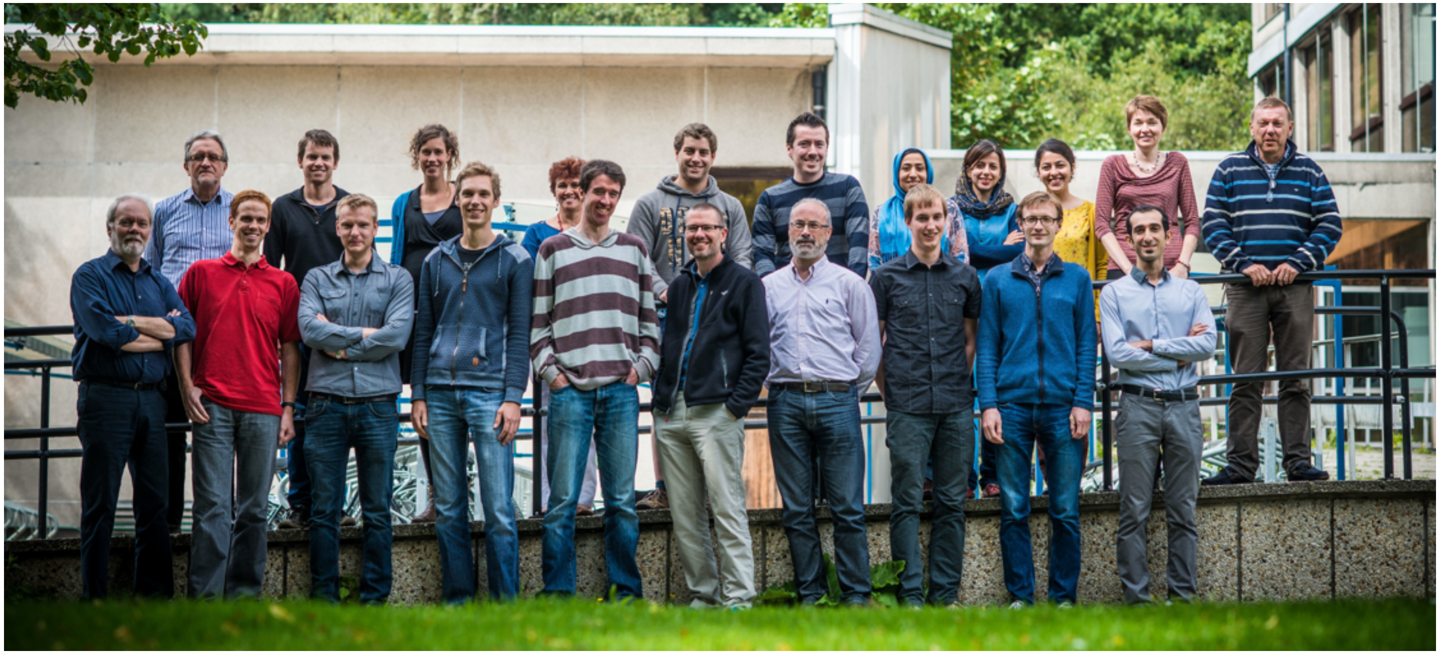


Conclusions:

- ✤ Reduced the data size on disk by a factor of ~ 10
- ✤ Reduced the spectrum evaluation time by a factor of ~ 1000
- ✤ While maintaining the interactive capabilities of spectrum evaluation
- ✤ While obtaining results very comparable with non-linear least squares fitting

A useful method for the processing of large MA-XRF datasets

Thanks AXES Antwerp X-ray Analysis, Electrochemistry & Speciation University of Antwerp https://www.uantwerpen.be/en/rg/axes/



Thanks to BRIGHTSPEC http://www.brightspec.be/

Special thanks for your attention