The Abdus Salam International Centre for Theoretical Physics

Joint ICTP-IAEA Workshop on Advances in X-ray Instrumentation for Cultural Heritage Applications

Evaluation and interpretation of large datasets from scanning macro X-ray fluorescence (MA-XRF) experiments: application to historical paintings

Piet Van Espen
piet.vanespen@uantwerpen.be

Tuesday, 14 July 2015

AXES
Antwerp X-ray Analysis, Electrochemistry & Speciation
University of Antwerp

Joint ICTP-IAEA School on Novel Experimental Methodologies for Synchrotron Radiation Applications in Nano-science and Environmental Monitoring

Evaluation of XRF Spectra

from basics to advanced systems

Piet Van Espen
piet.vanespen@uantwerpen.be

AXES
Antwerp X-ray Analysis, Electrochemistry & Speciation
University of Antwerp

20 Nov 2014

Some final remarks: The future

Non-linear least-squares works

if you have a good parsimonious model

if you have TIME

X-ray fluorescence imaging:

256 x 256 image = 65536 x-ray spectra

@ 1 s / spectrum

= 65536 seconds

= 1092 minutes

= 18 hours !!!!

Need to explore new methods

Linear models?

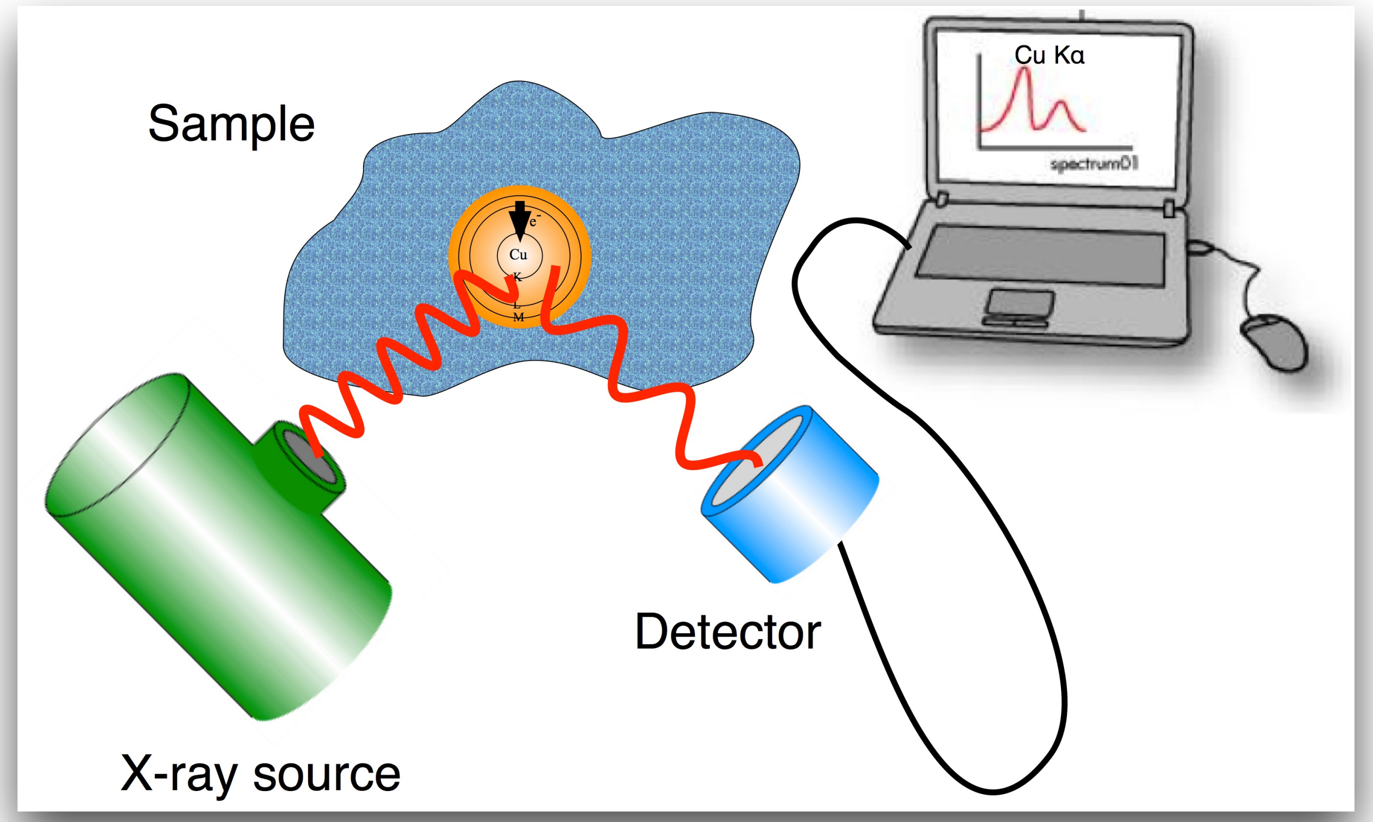
Multivariate models?

Outline

1. The context
2. A space - time problem
3. Solving the “space” problem
4. Solving the “time” problem
5. Results
6. Application
7. Conclusions

1. The context X-ray fluorescence

analytical technique for element analysis



Energies of the fluorescence x-rays are characteristic for each element

range: 1 - 100 keV

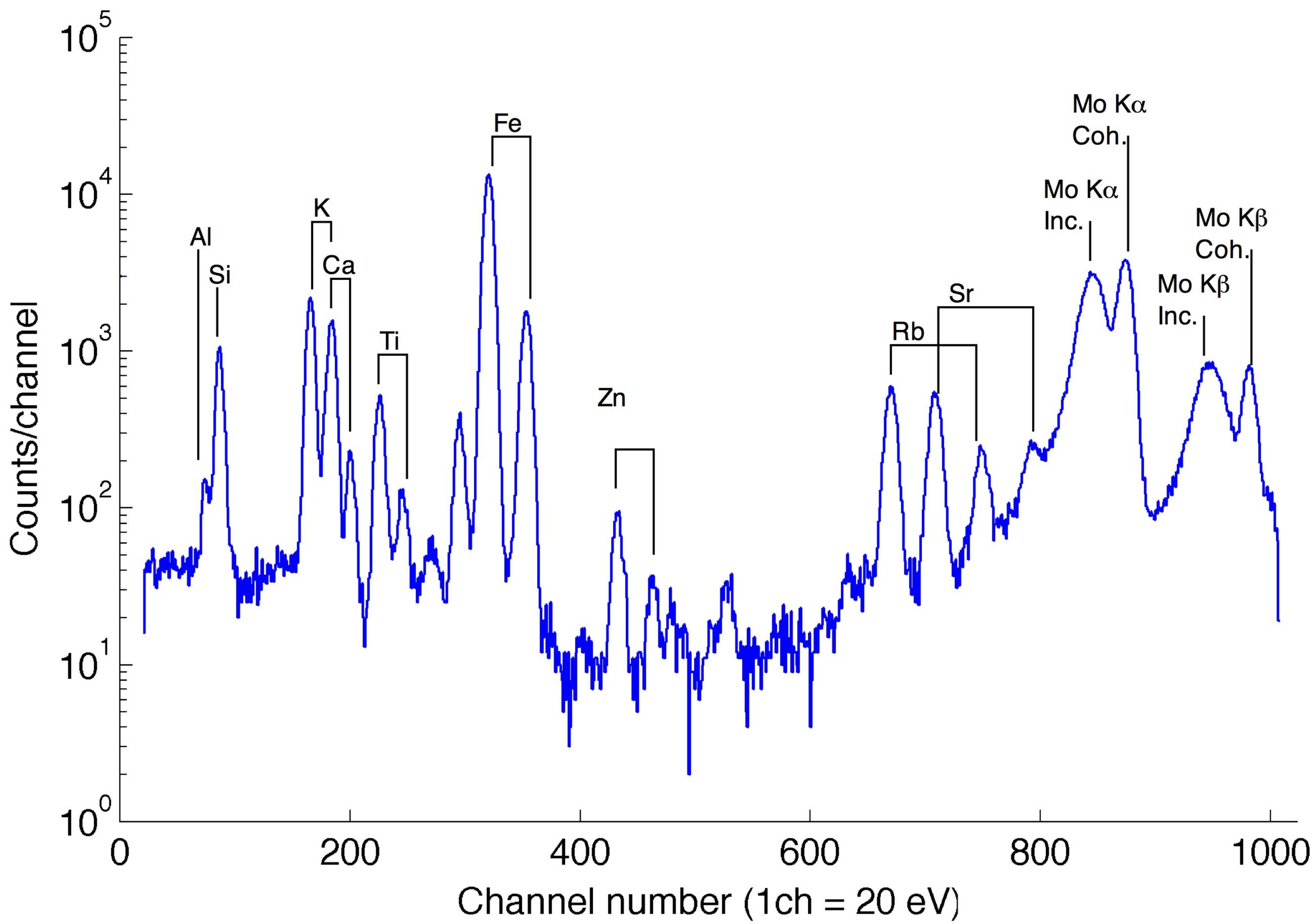
qualitative analysis
which element?
Na - U

Intensity
(number of counts in characteristic line)

range: 100% - ppm

quantitative analysis
how much of each element

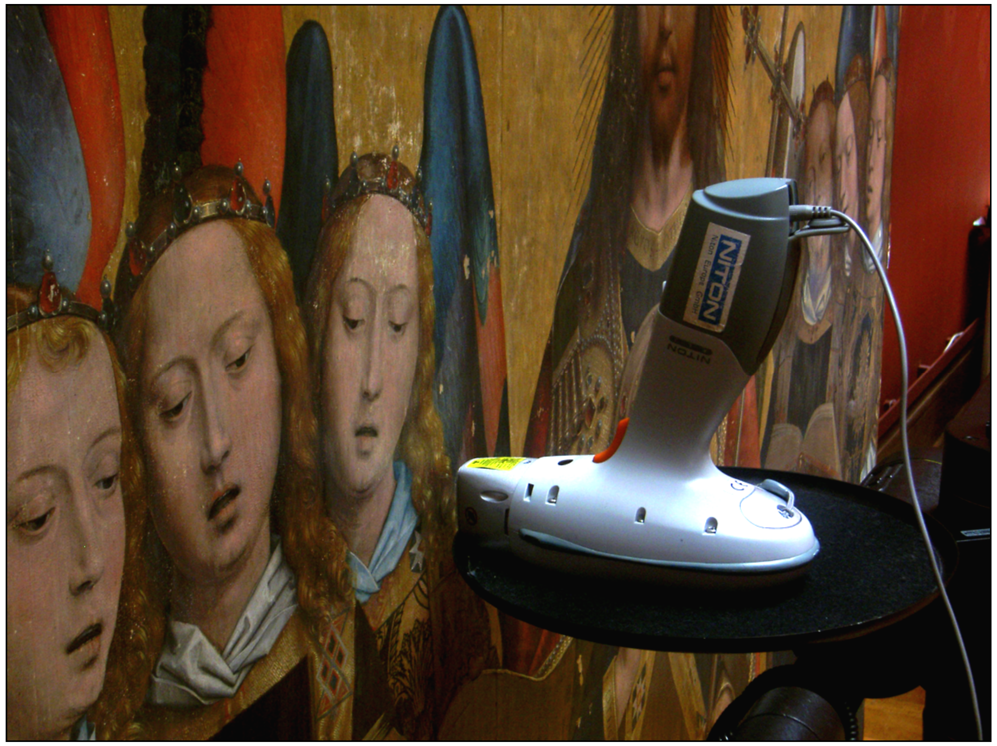
Result of a measurement: X-ray spectrum



Characteristic lines (Kα, Kβ, Lα, Lβ,…) with frequent interference (peak overlap) sitting on a non-specific continuum (background)

Data: 1 - 4 K, 1024 - 4096 channels

X-ray fluorescence analysis (XRF) and cultural heritage



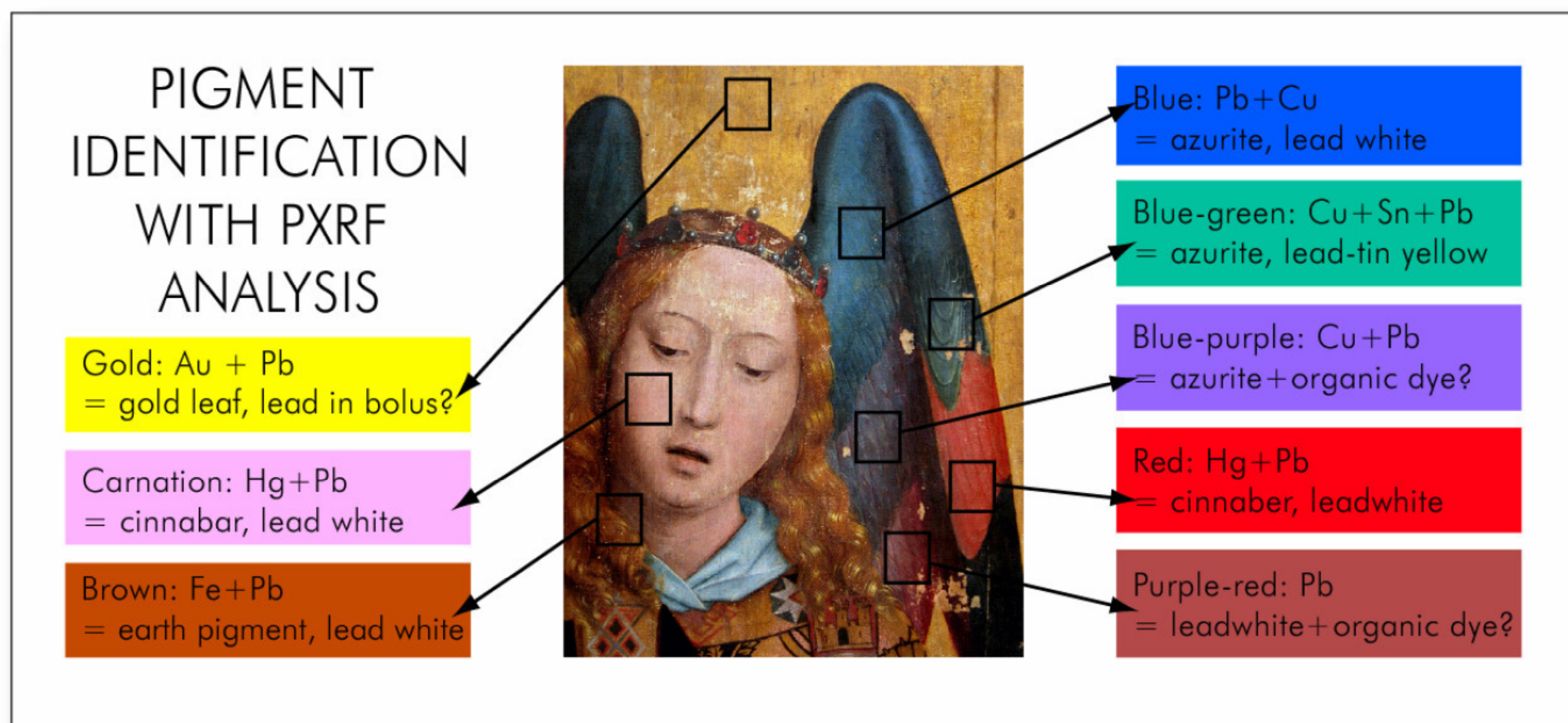
It is **non-destructive**!!!

pigment analysis using portable XRF instruments.

measuring and analysing 10 — 100 spots on a painting

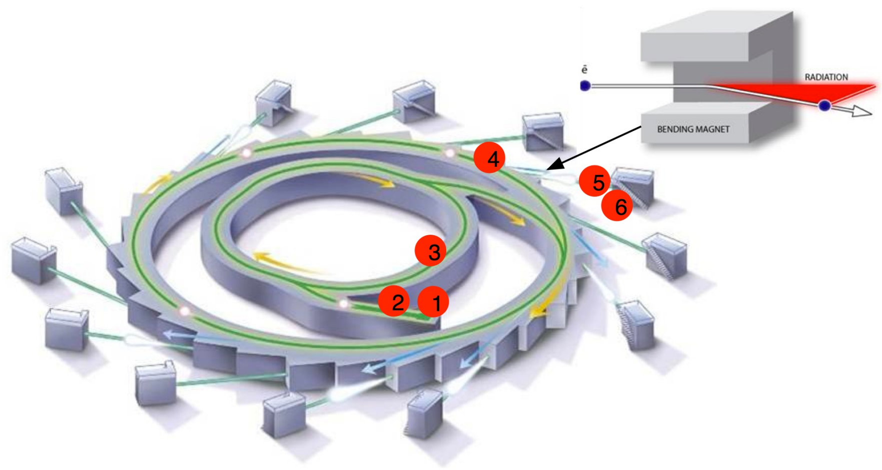
information for:

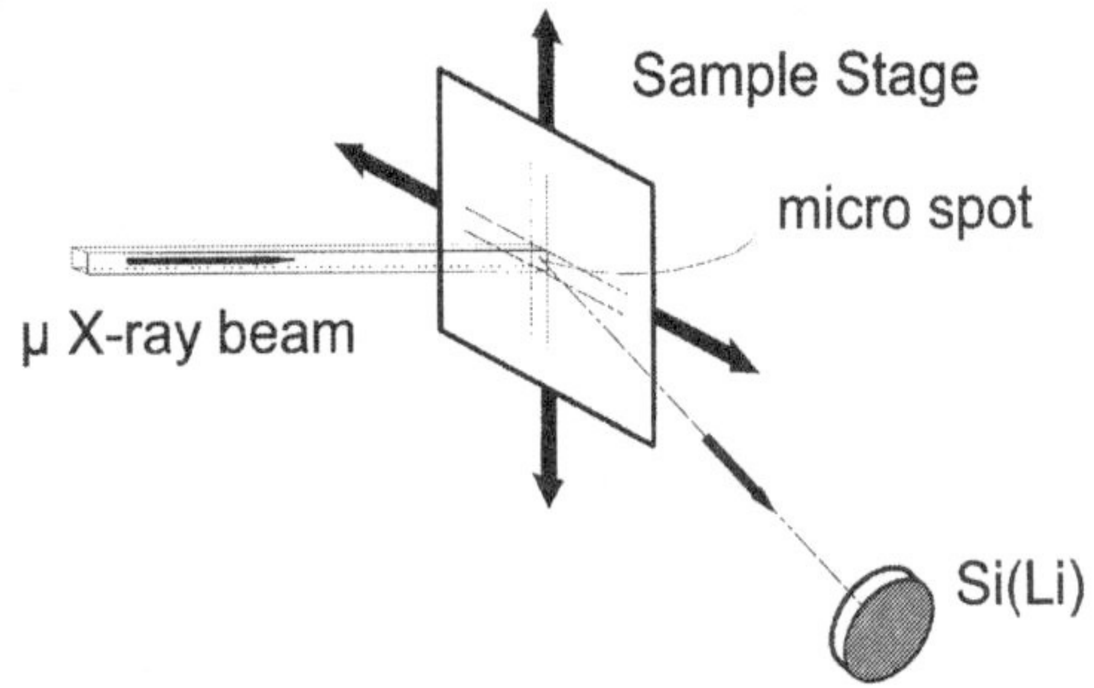
- restoration
- preservation
- authentification
- …



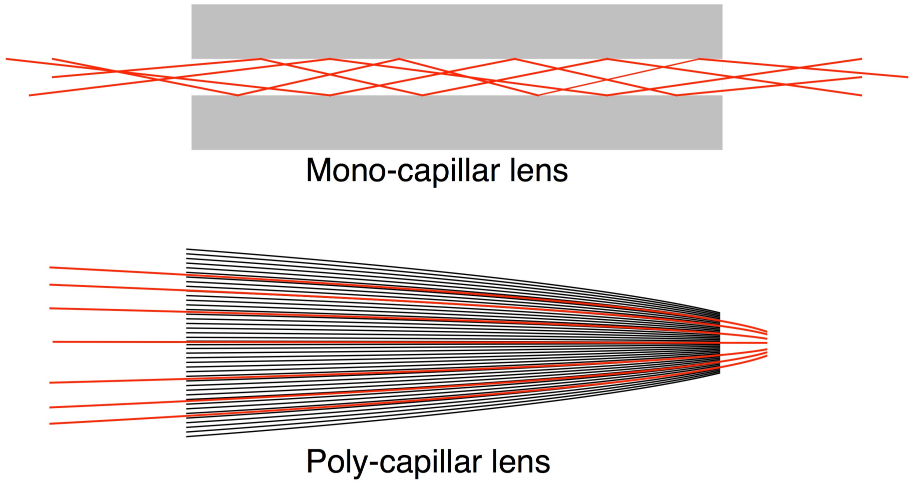
X-ray fluorescence imaging

Use a pencil x-ray beam
and take a spectrum at each pixel





using capillary optics or collimators



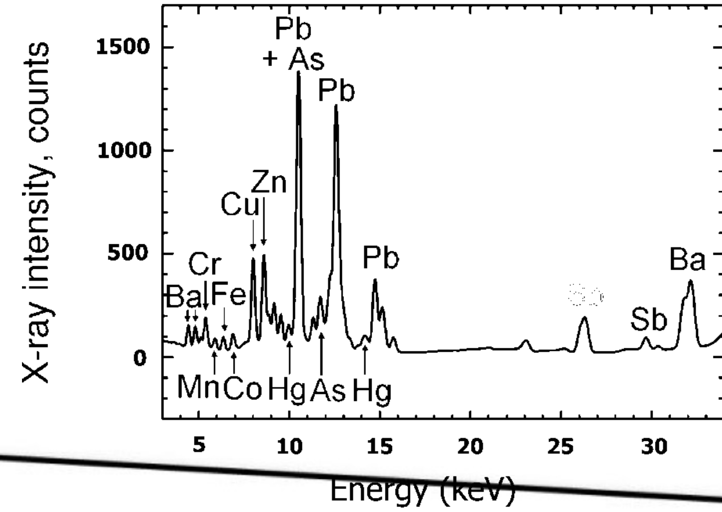
First succes Using synchrotron radiation XRF

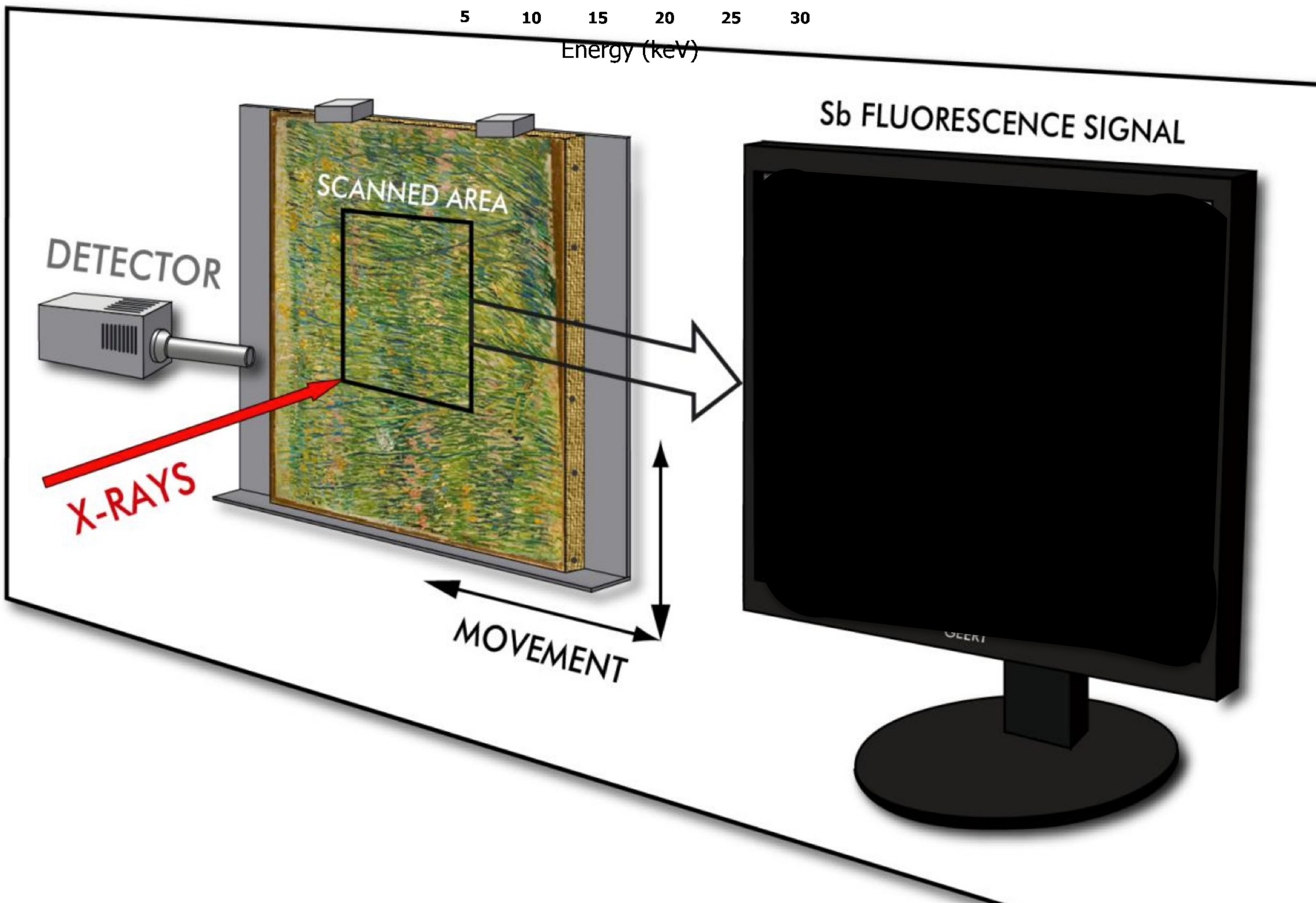
Applied to “Patch of Grass” V. Van Gogh (1887)

μXRF beamline L at Doris-III (Hasylab)

0.5×0.5 mm^2 @ 38.5 keV

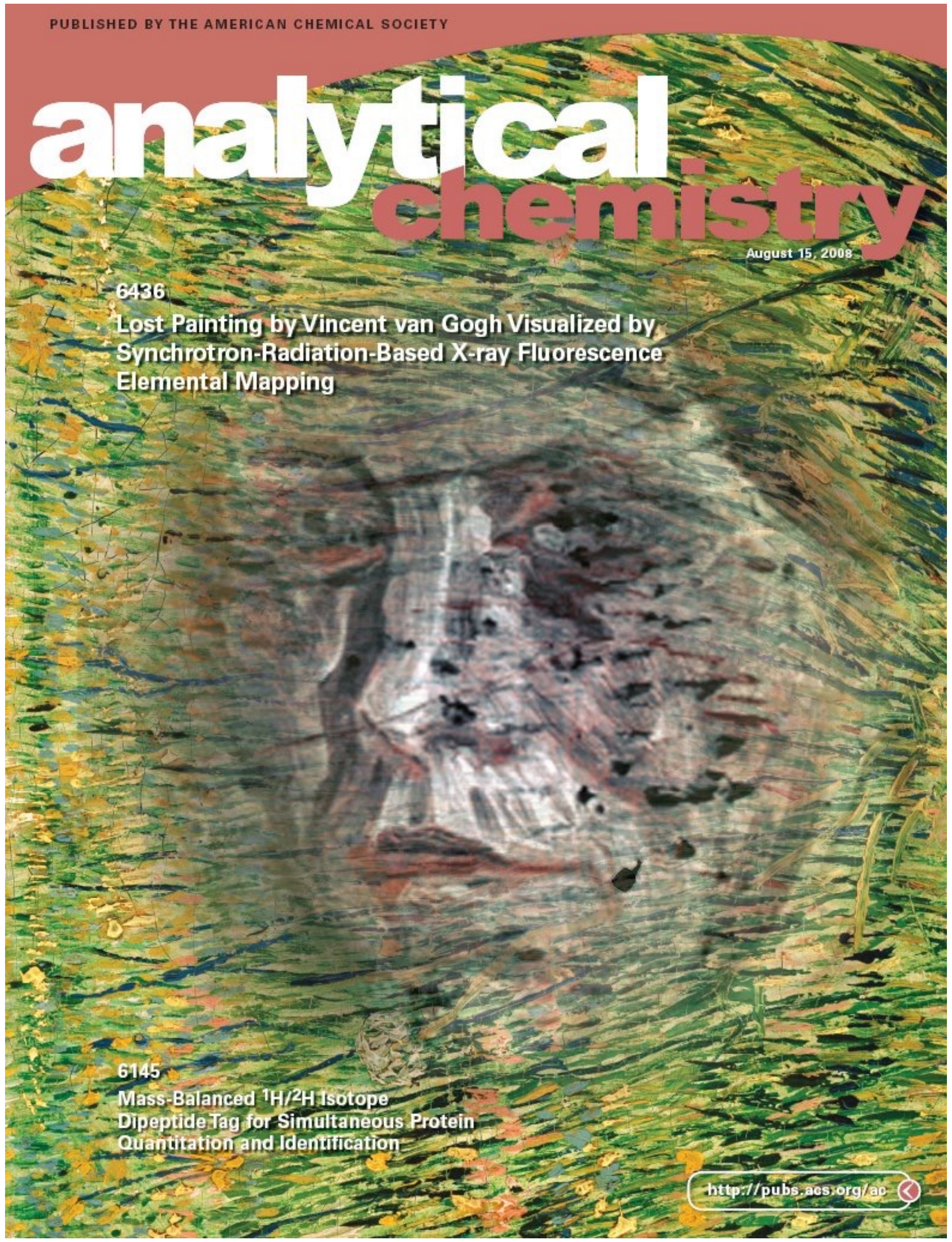
Scanned area 170×170 mm^2 @ 2s per pixel (64h)





Discovery of a hidden painting

Some “visibility”



naturenews

Published online 30 July 2008 | *Nature* **454**, 563 (2008) | doi:10.1038/454563a

News in Brief: Snapshot

The hidden van Gogh

X-ray reveals mystery portrait.

Philip Ball

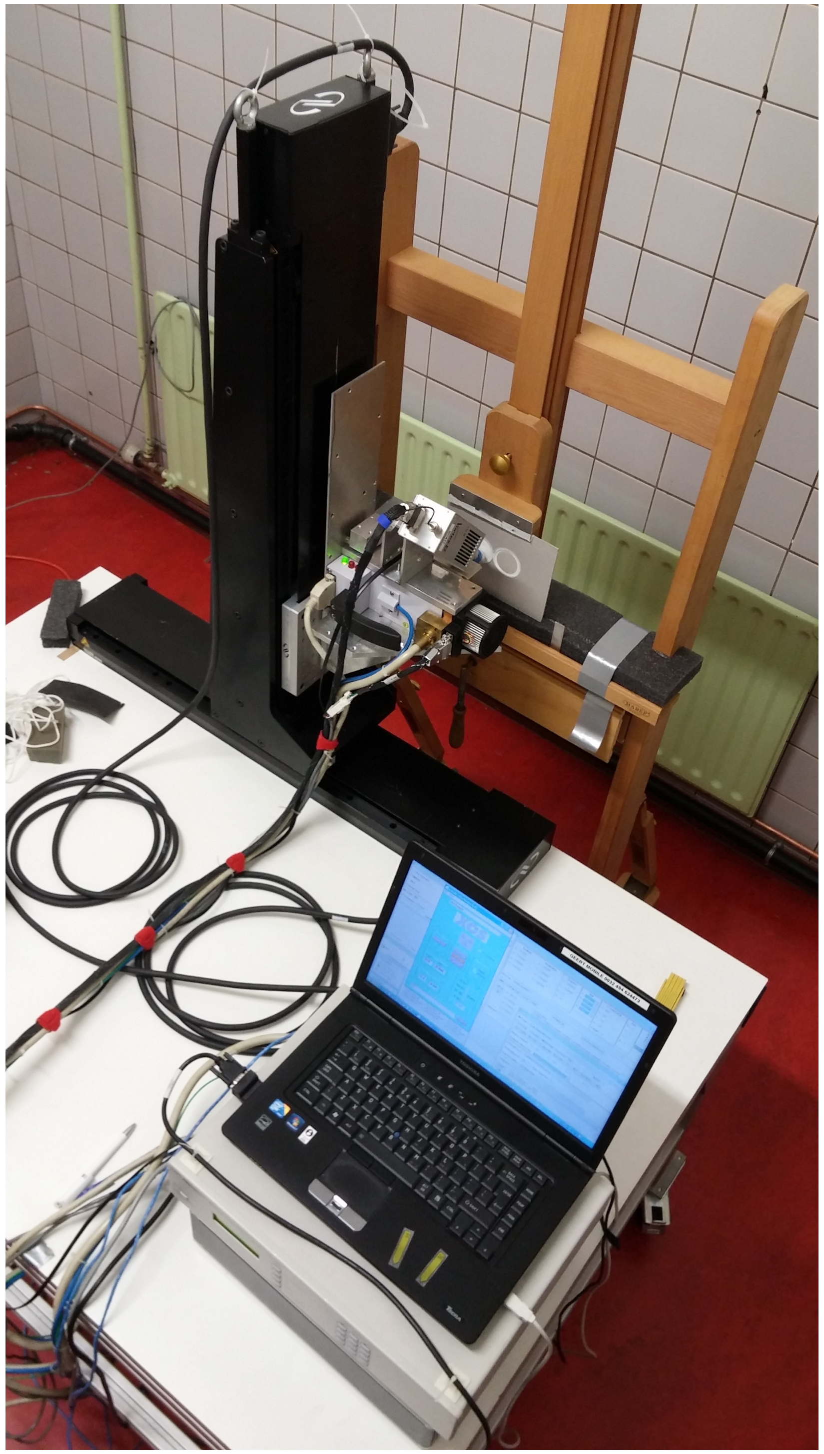
An unknown Vincent van Gogh painting of a woman's head has been revealed with X-ray technology. The painting is thought to have been made in 1884–85, during a period in which he painted several portraits of peasants in the

Method known as Scanning Macro XRF (MA-XRF)

Since then we scanned hundreds of paintings ⇒ ~ 13 TB data

More recent trend: X-ray fluorescence imaging using small x-ray tubes

Large size: moving the source and detector



e.g.
AXES MA-XRF scanner

Small size: moving the object

Commercial instruments

M6 JETSTREAM

Rh anode microfocus X-ray tube 50 kV 600 µA
30 mm^2 SDD, <145 eV @ Mn Ka
Spot size: 100 - 500 µm
Scanning area: 800 x 600 mm, 10 µm stepsize

2. The Space - Time Problem

Paintings are BIG!!!!

Research is needed to make the method work in practice!

Maybe some old recipes will do?

a) Measuring: a time problem

256 x 256 pixels @0.5mm = 12.8 cm x 12.8 cm painting

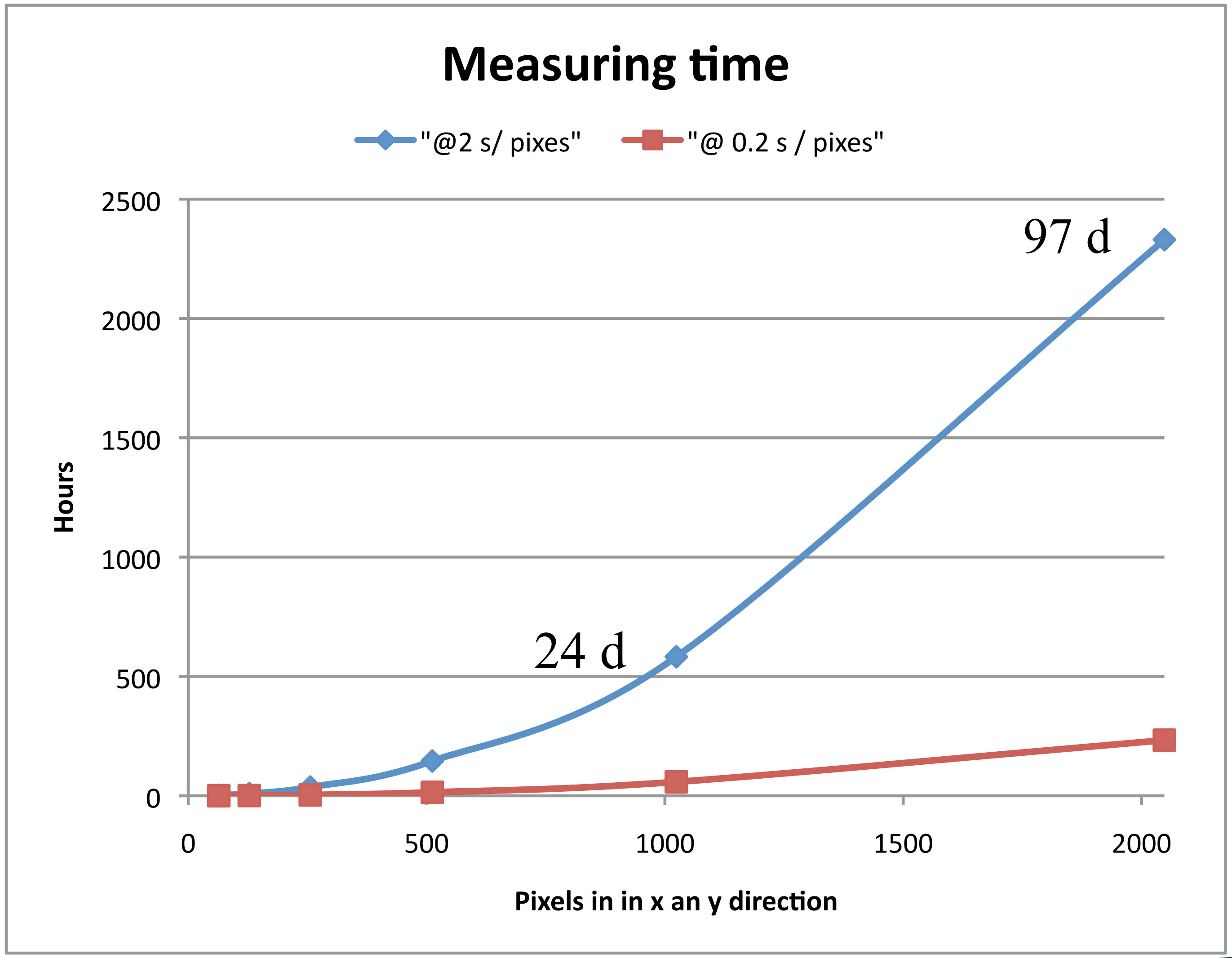
@ 2s/pixel = 36 h = 1.5 days

@ 0.2s/pixel = 3.6 h

1024 x 1024 pixels @0.5mm = 51.2 cm x 51.2 cm painting

@ 2s/pixel = 582 h = 24 days!!!

@ 0.2s/pixel = 58.2 h = 2.4 days



count rate 10^5 cps
@ 0.2s/pixel

20000 counts/spectrum
20 counts/channel

Posible solution

Technological
Pixelated detectors

b) Data store and retrieval: a space problem

1 channel in a spectrum = 4 bytes
1 spectrum = 1024 channel = 4 kB

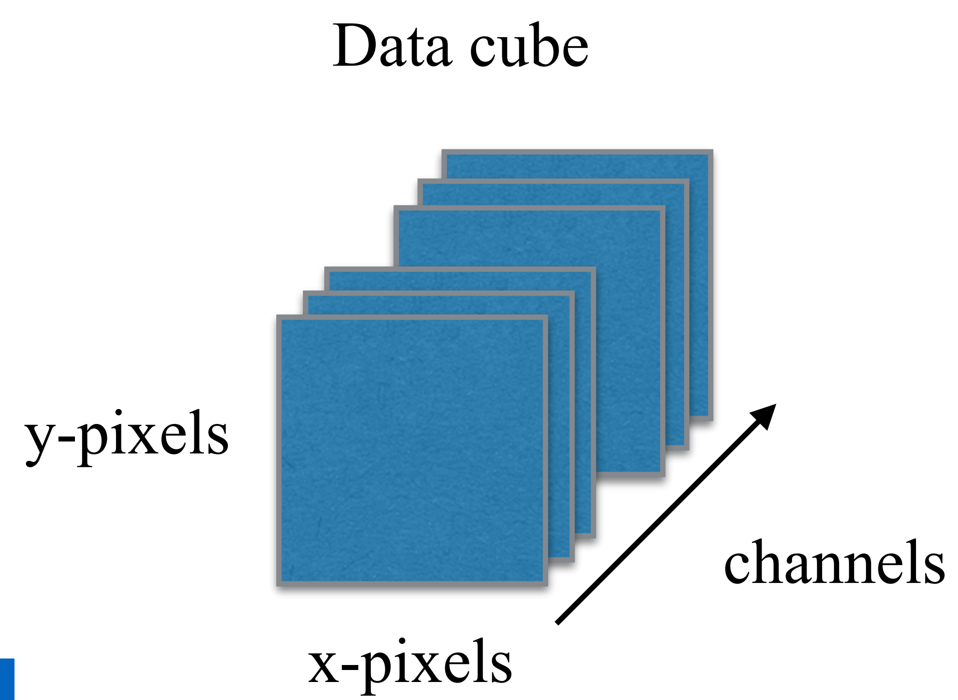


Image size	#spectra	MByte
256 x 256	65536	256
512 x 512	262144	1024 = 1GB
1024 x 1024	1048576	4096= **4 GB**

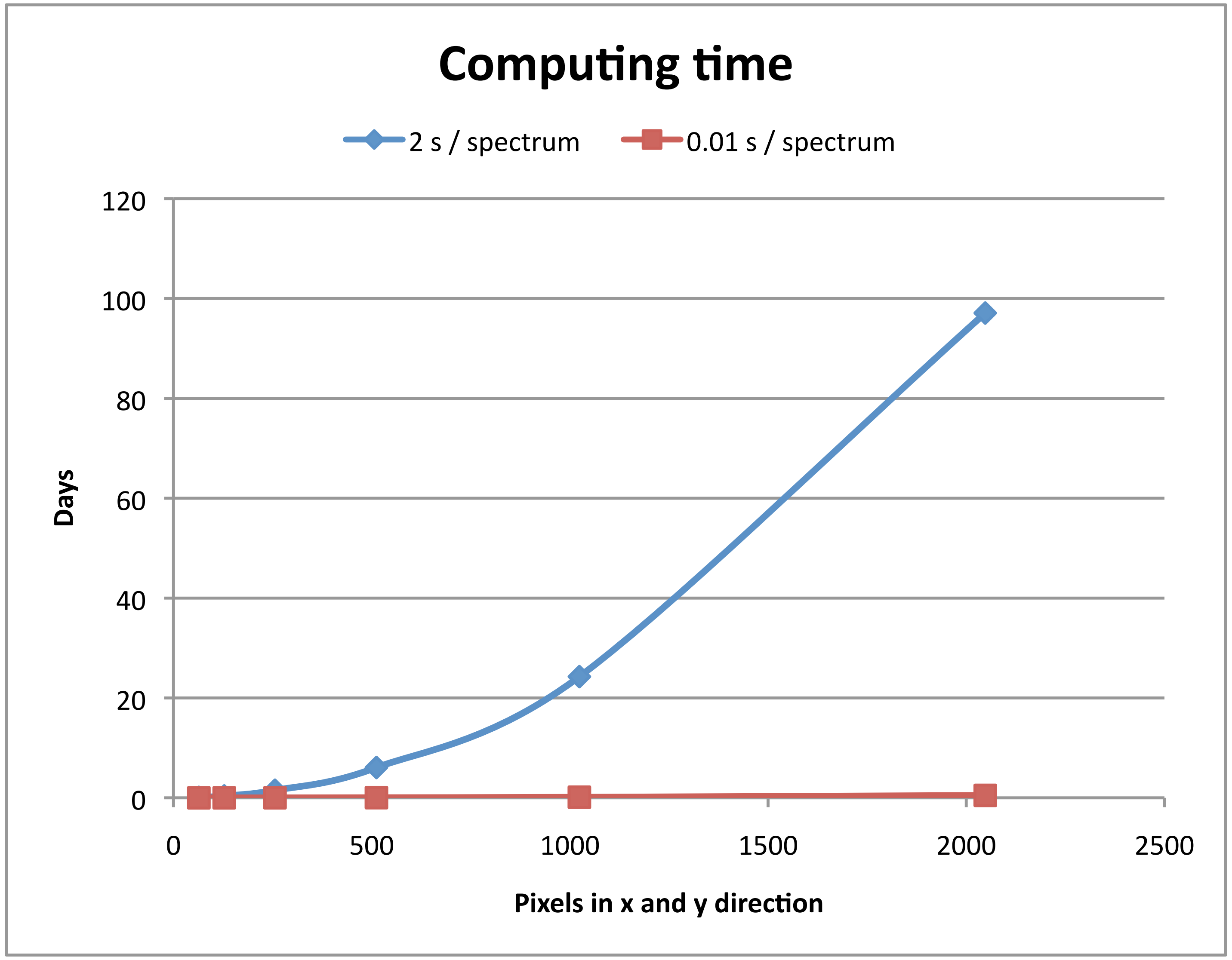
⇒ reduce the data size by a factor of 10 !!!

c) Analysis: another time problem

Analysis of 10^4 to 10^6 spectra ???

256 x 256 pixels @0.5mm = 12.8 cm x 12.8 cm painting @ 2s/spectrum = 36 h = 1.5 days

1024 x 1024 pixels @0.5mm = 51.2 cm x 51.2 cm painting @ 2s/spectrum = 582 h = 24 days!!!



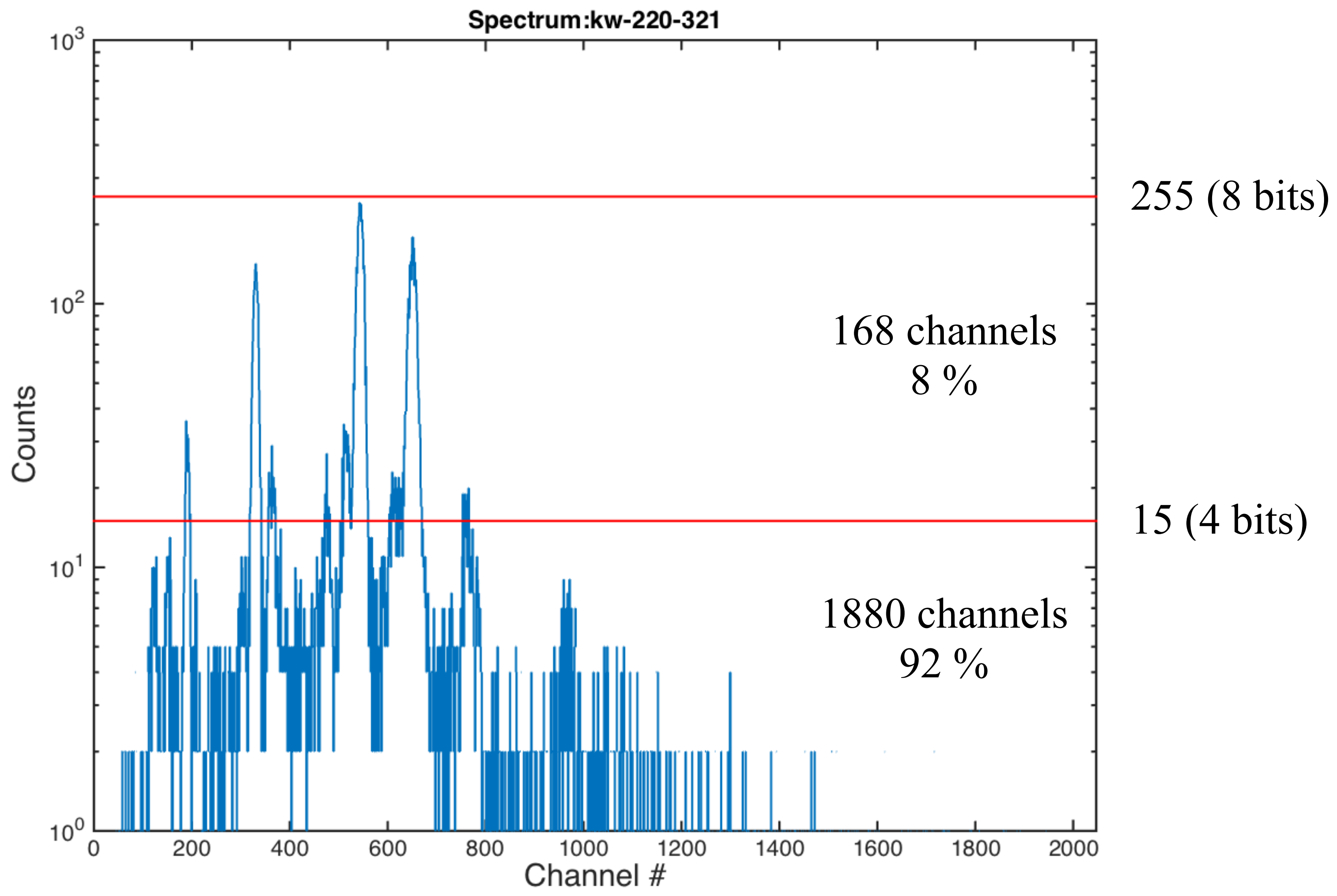
⇒ reduce the spectrum analysis time by a factor 100 !!!

3. Solving the Space Problem

Compression

Channel content stored in 4 bytes (values 0 — 4 294 967 295 counts)

Original: 2048 channel = 8192 bytes

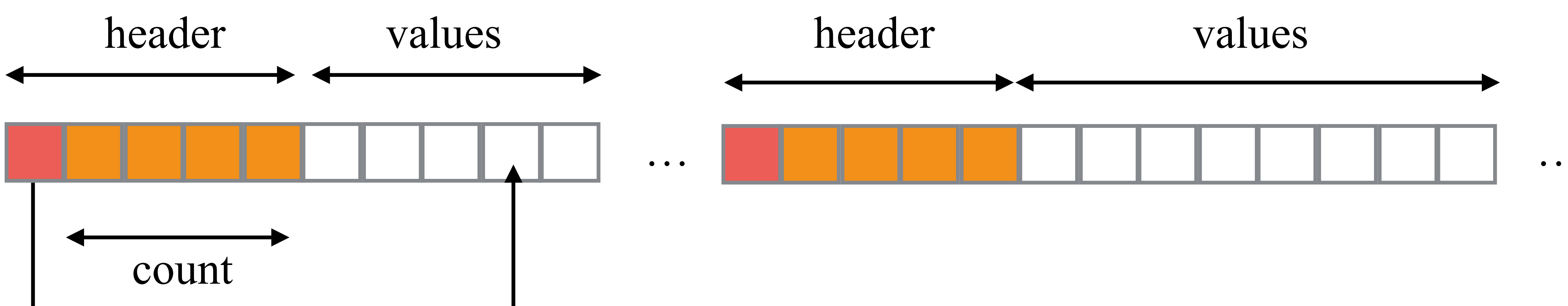


Real MA-XRF spectrum has many channels between 0 and 15 counts needs only 1/2 byte (nibble)

Lossless spectrum compression

Original: constant length 2048 × 4 bytes

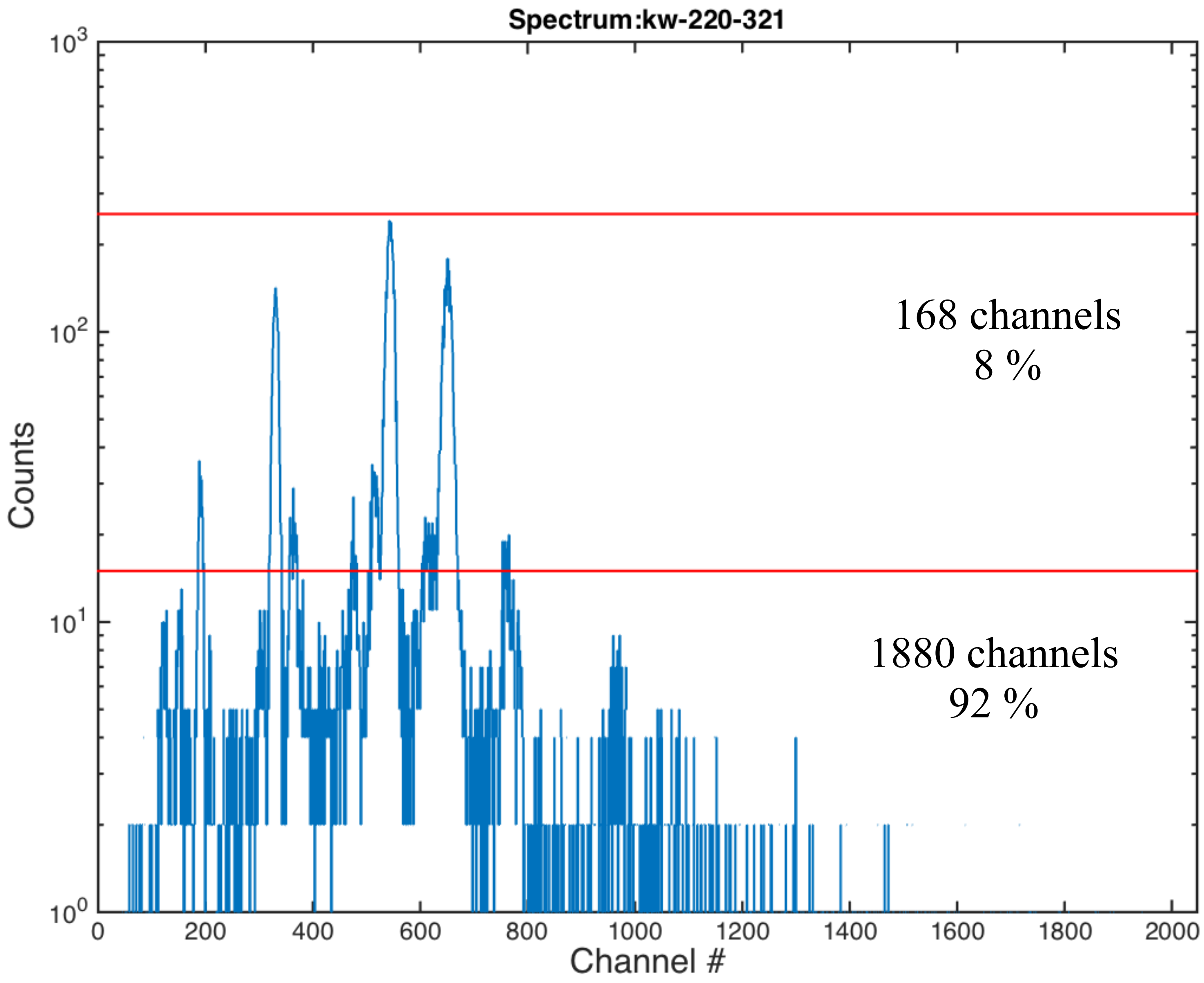
Variable length compression: use ½ byte when possible, 1 byte if needed or 1½, or 2…



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

Compression factor

Original:	8192 bytes
Compressed:	1005 bytes
Factor:	8.2



255 (8 bits)

168 channels
8 %

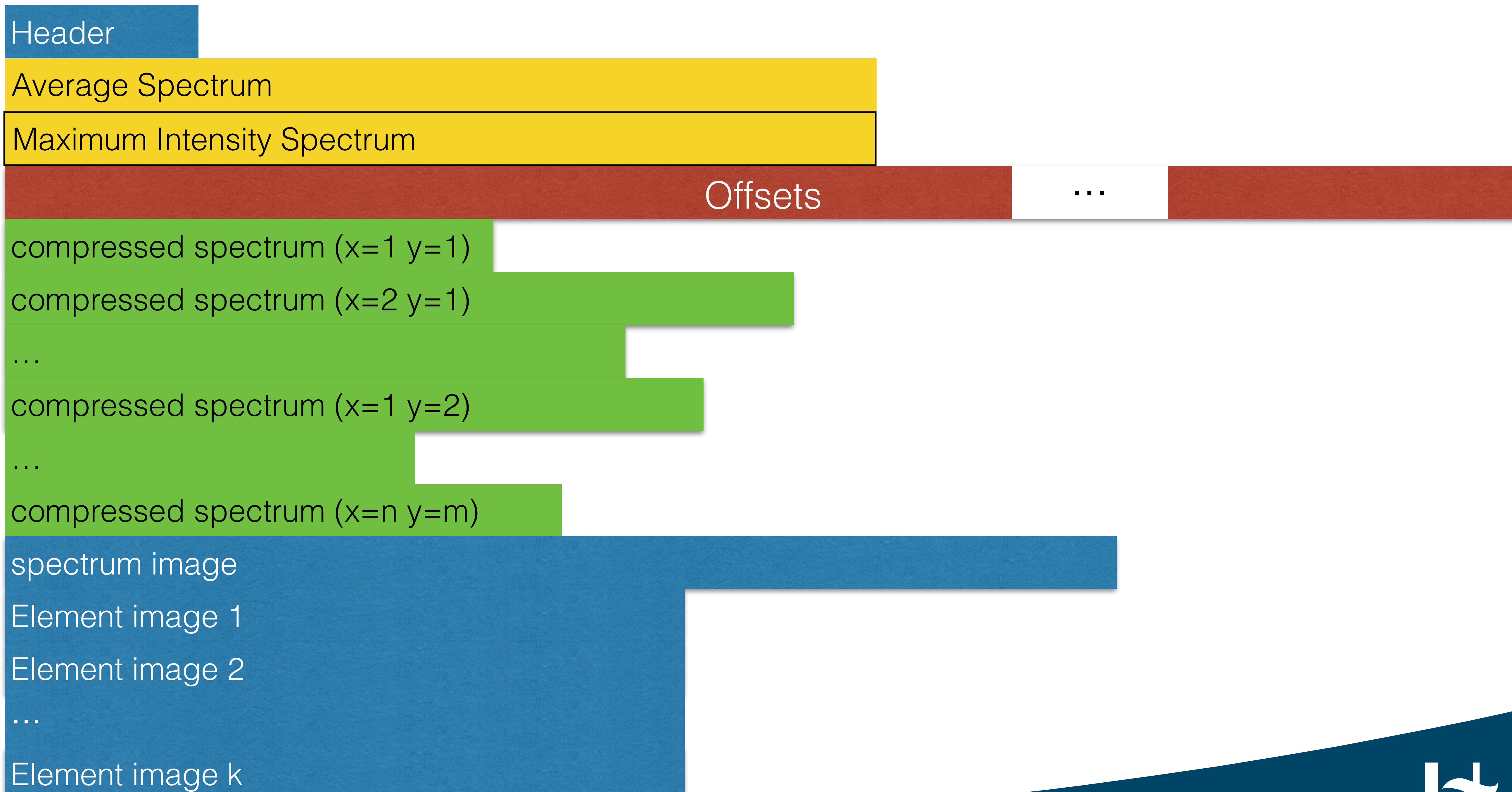
15 (4 bits)

1880 channels
92 %

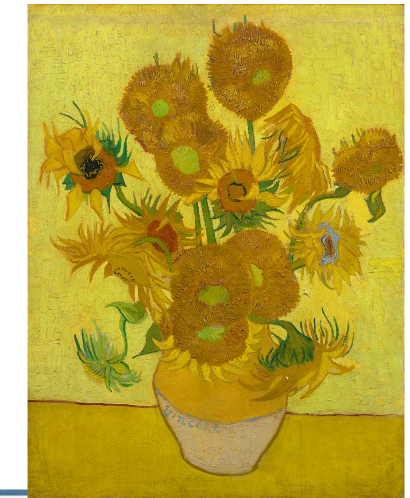
Efficiency of data compression

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

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

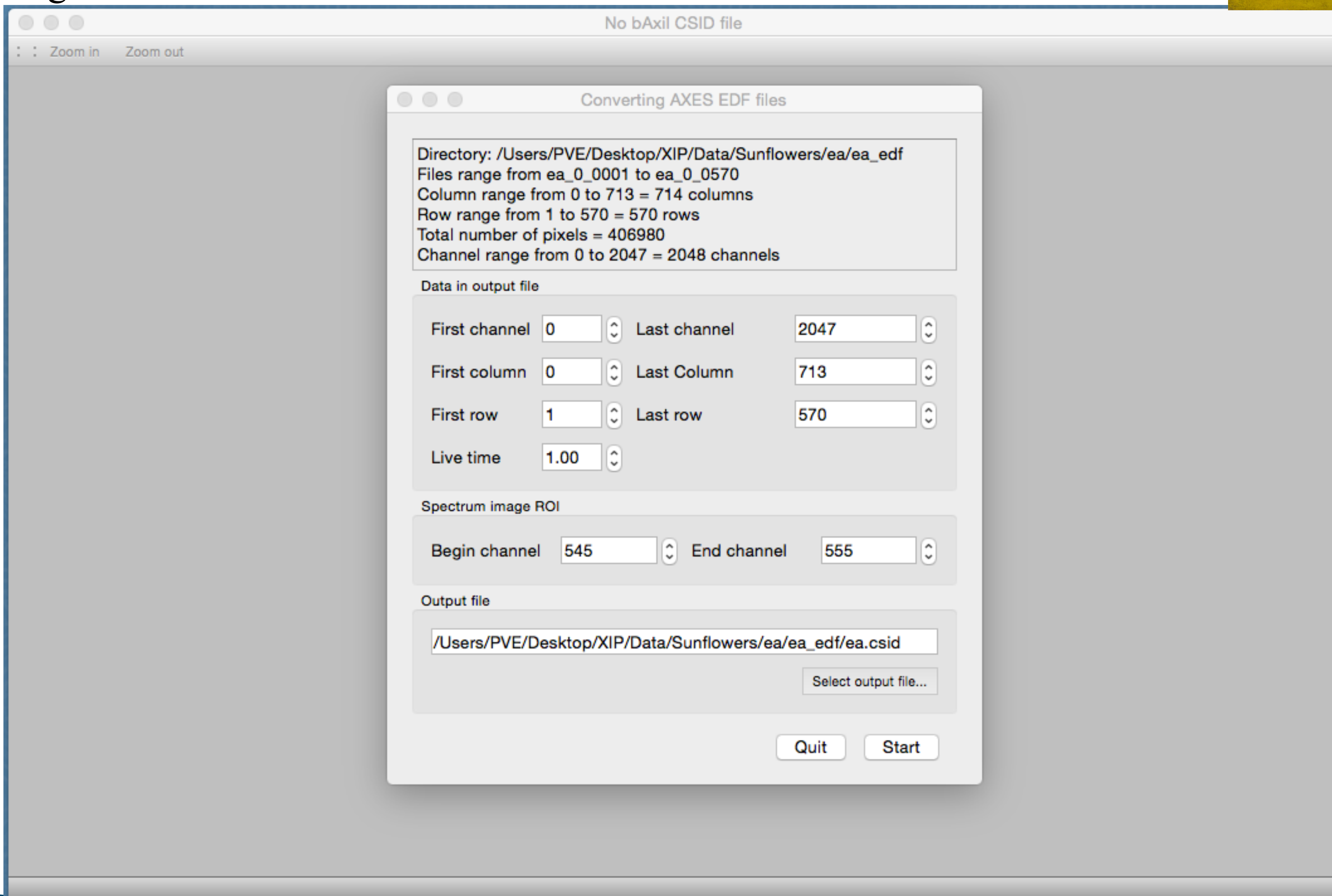


Example converting to compressed spectroscopic image data (csid)



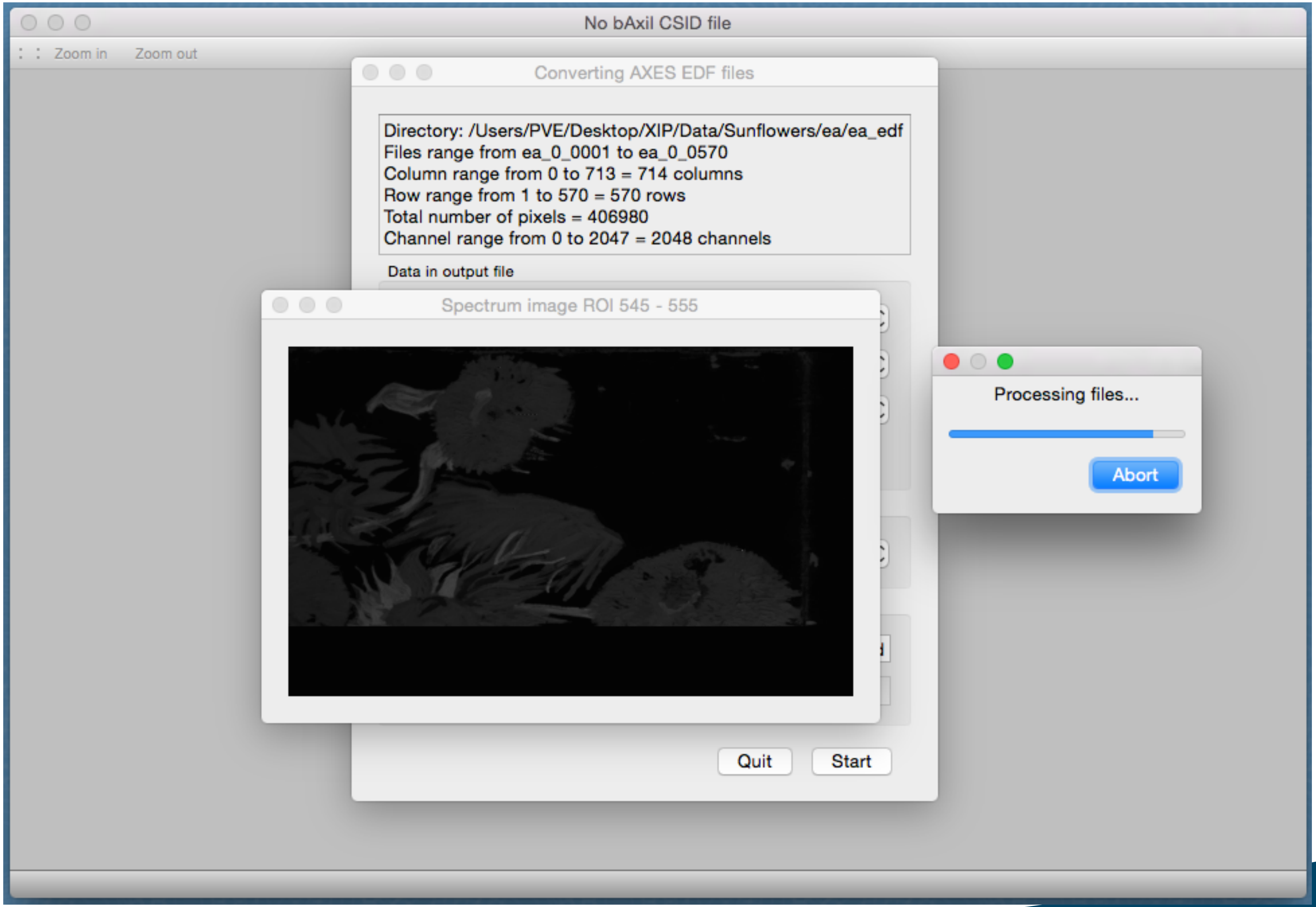
Original data: part of scan of Van Gogh's Sunflowers
570 edf files of 5.9 MB each = 3.34 GB (1.53 GB zipped)
each edf file: 714 spectra of 2047 channels (1 scan line)

Importing the 570 files

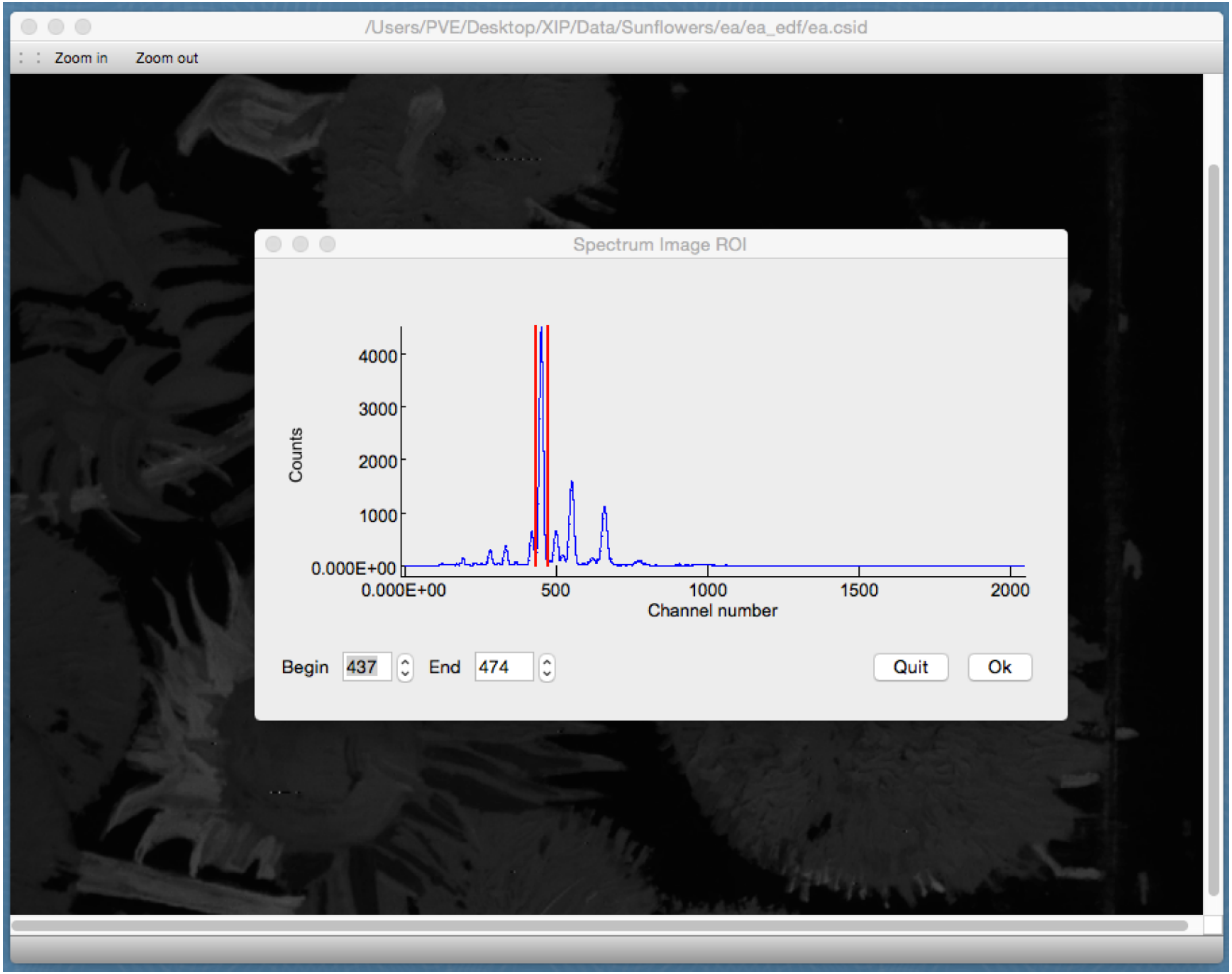


While importing

visualisation of the x-ray intensity between channel 545 and 555 (Pb-Lα) in each pixels
ROI image

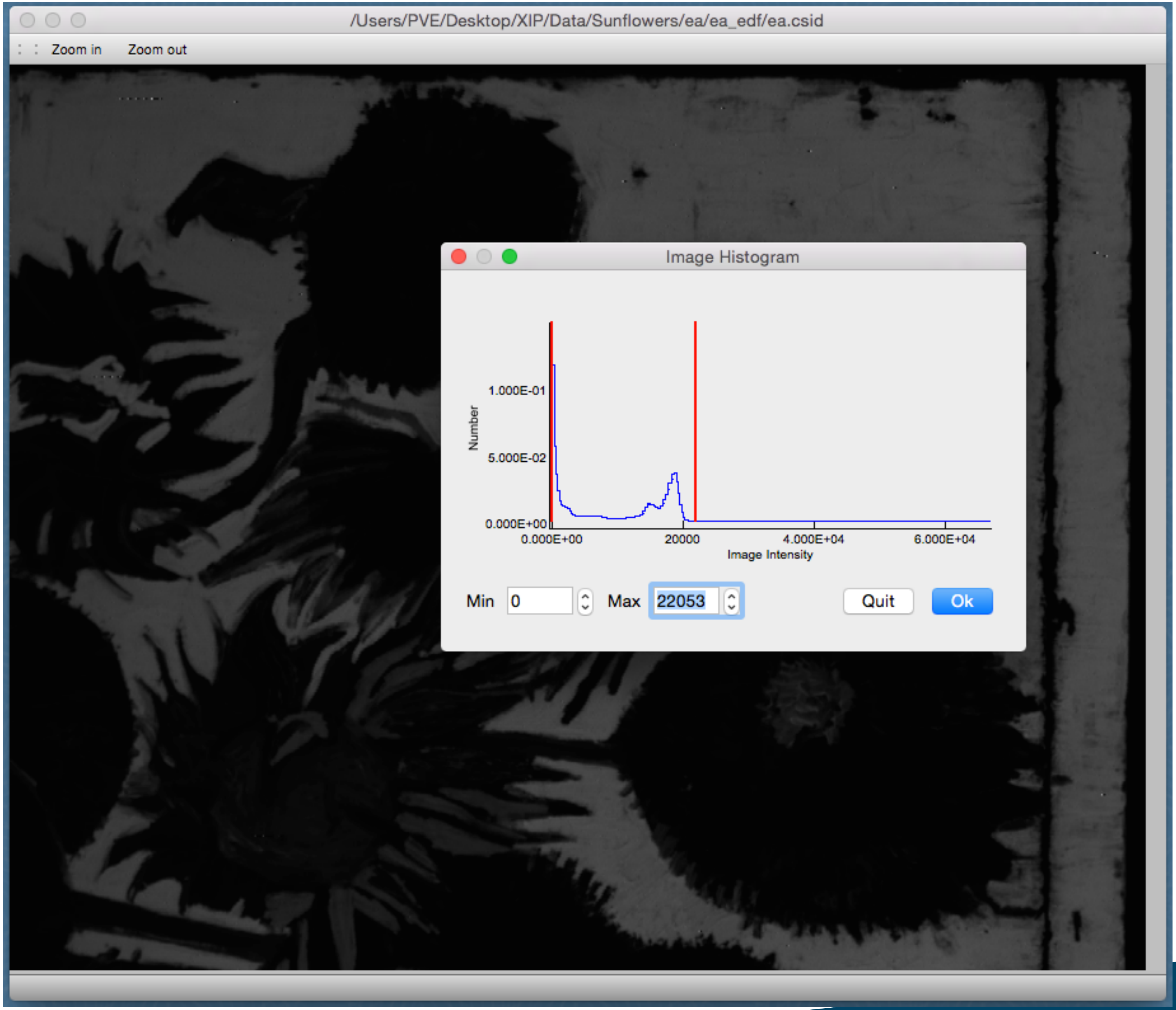


Adjusting the ROI image to channel 437 - 474 (Zn Kα)



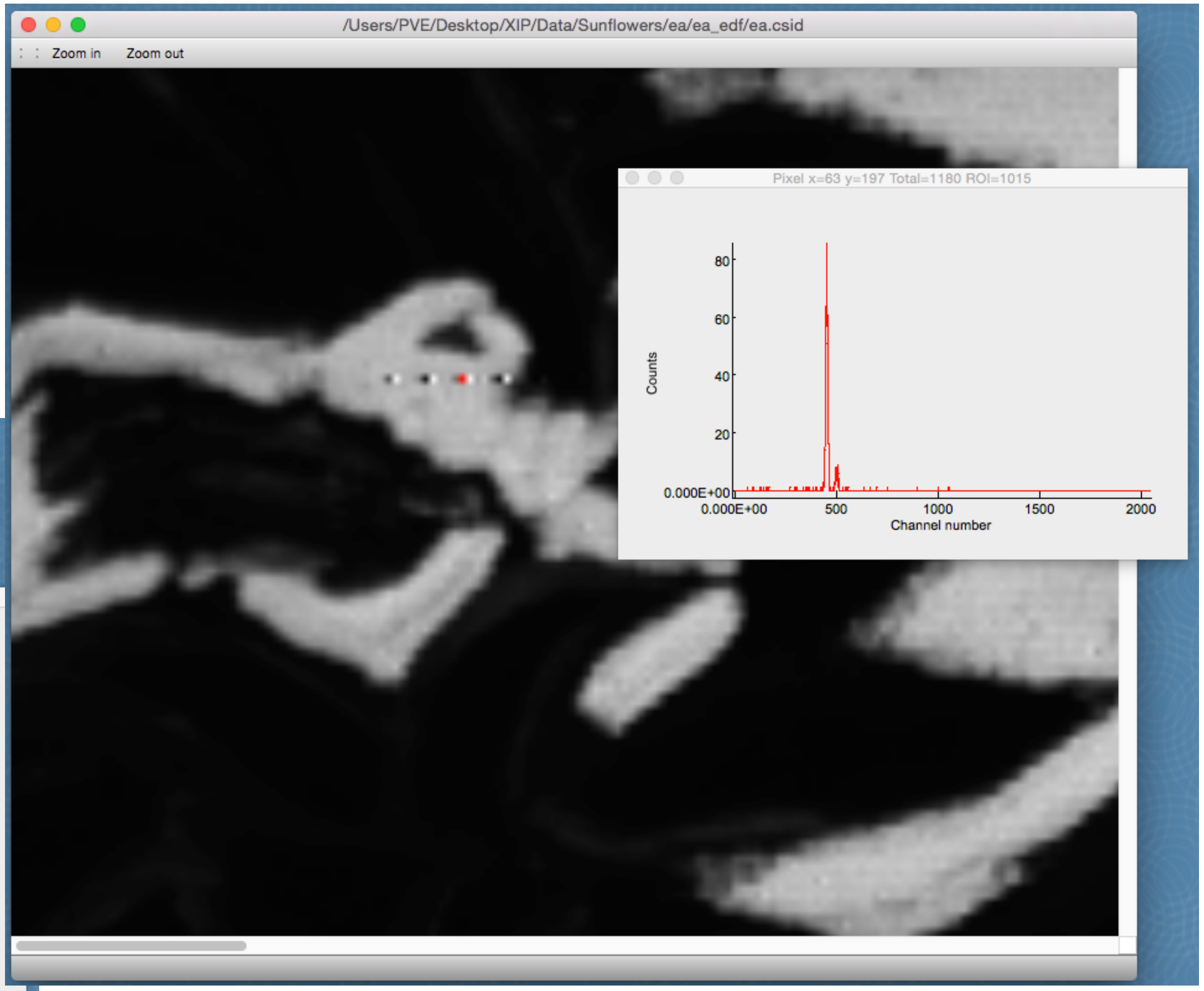
ROI image of channel 437 - 474 (Zn Kα)

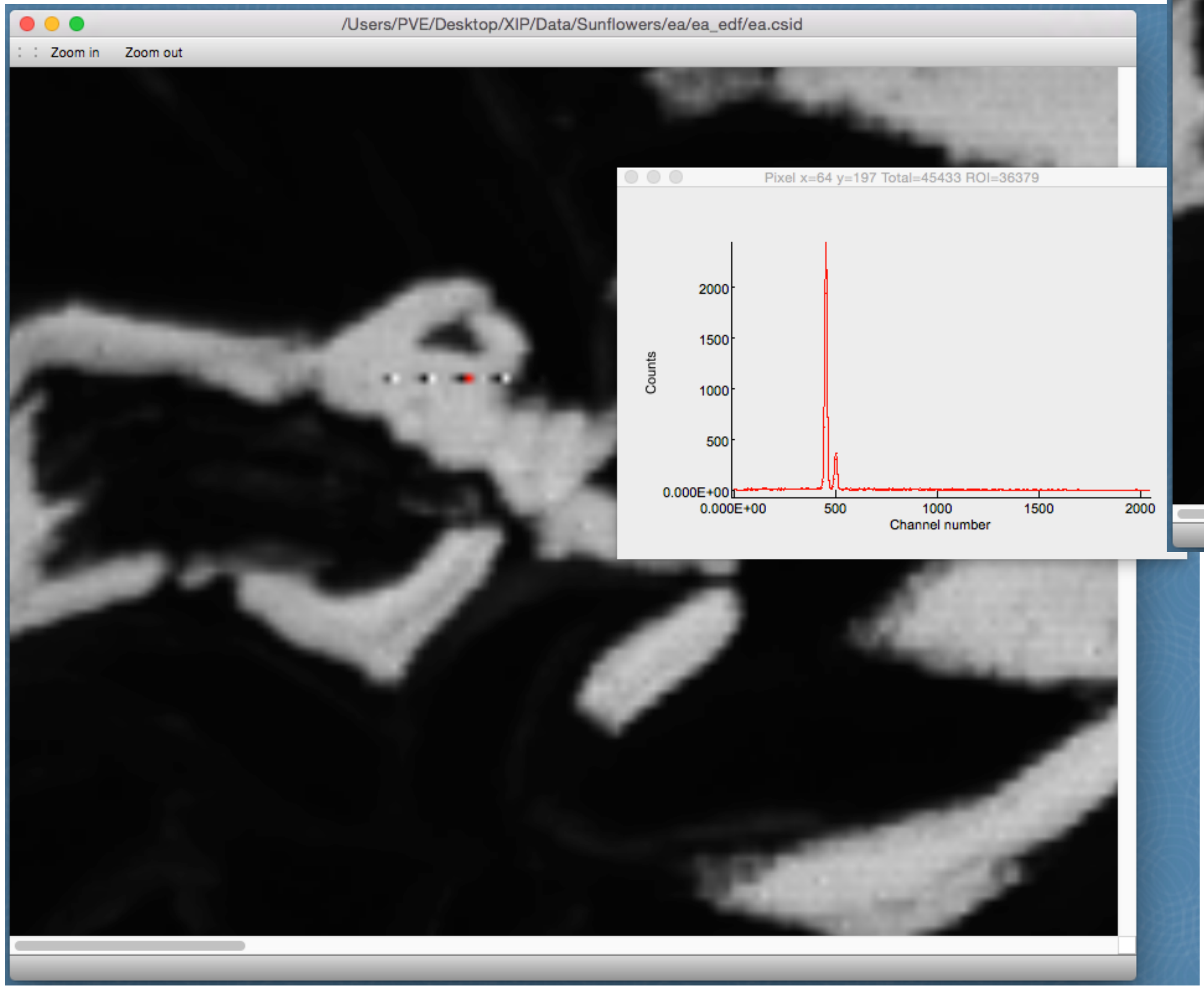
Adjusting the image histogram



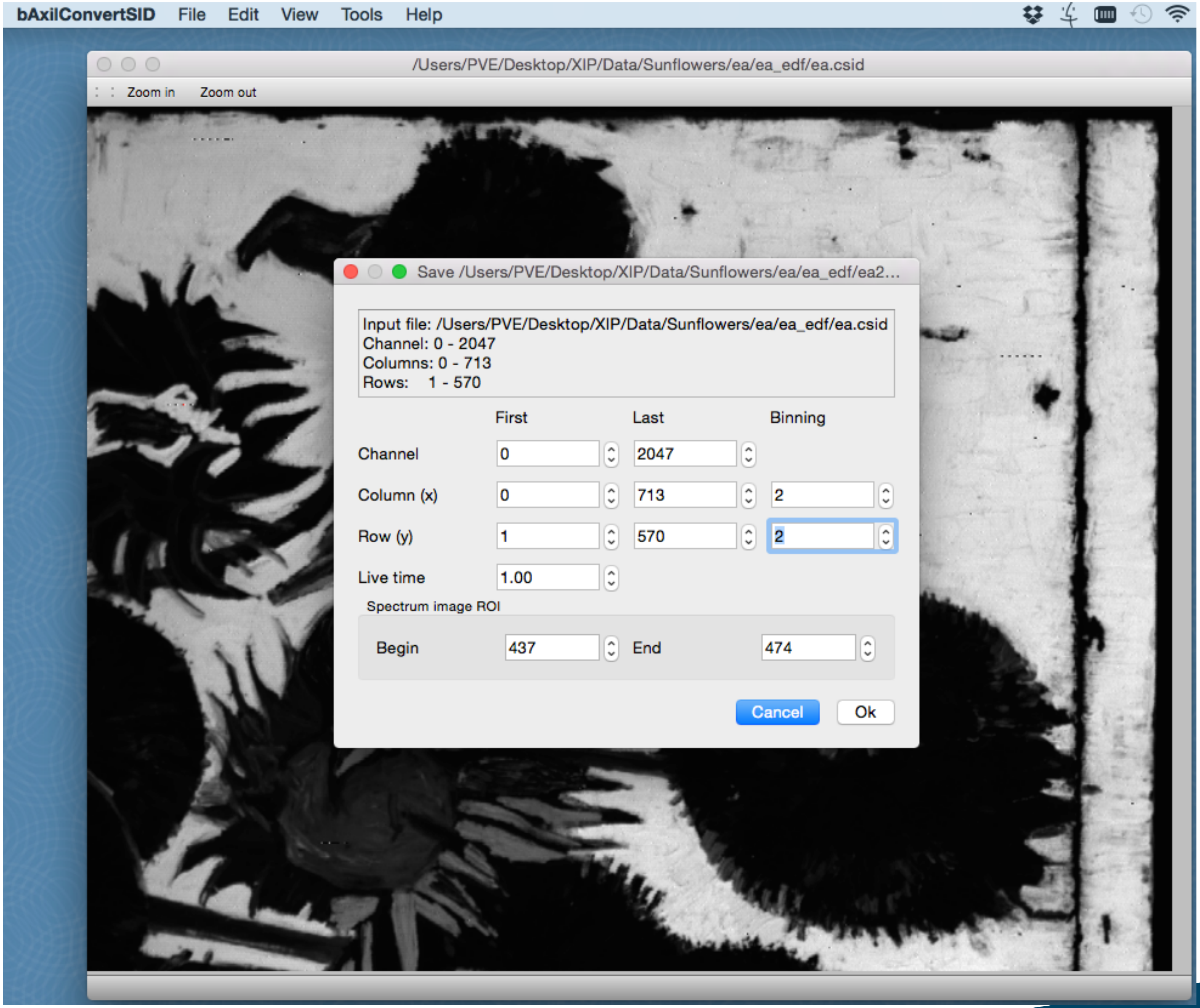
ROI image of channel 437 - 474 (Zn Kα) with adjusted histogram

Zooming the image and reviewing the spectrum at each pixel





Storing the data with optional reshaping and resampling (binning)



Original data: part of scan of Van Gogh’s Sunflowers

570 edf files of 5.9 MB each = 3.34 GB (1.53 GB zipped)
each edf file 714 spectra of 2047 channels (1 scan line)

MA-XRF scan 570 × 714 pixels = 406980 spectra

Compressed = 300.7 MB on disk

size reduction by factor of 11.1

Original data formats: ascii text format, EDF (ESRF) format…

Convert to CSID: Compressed Spectroscopic Image Data

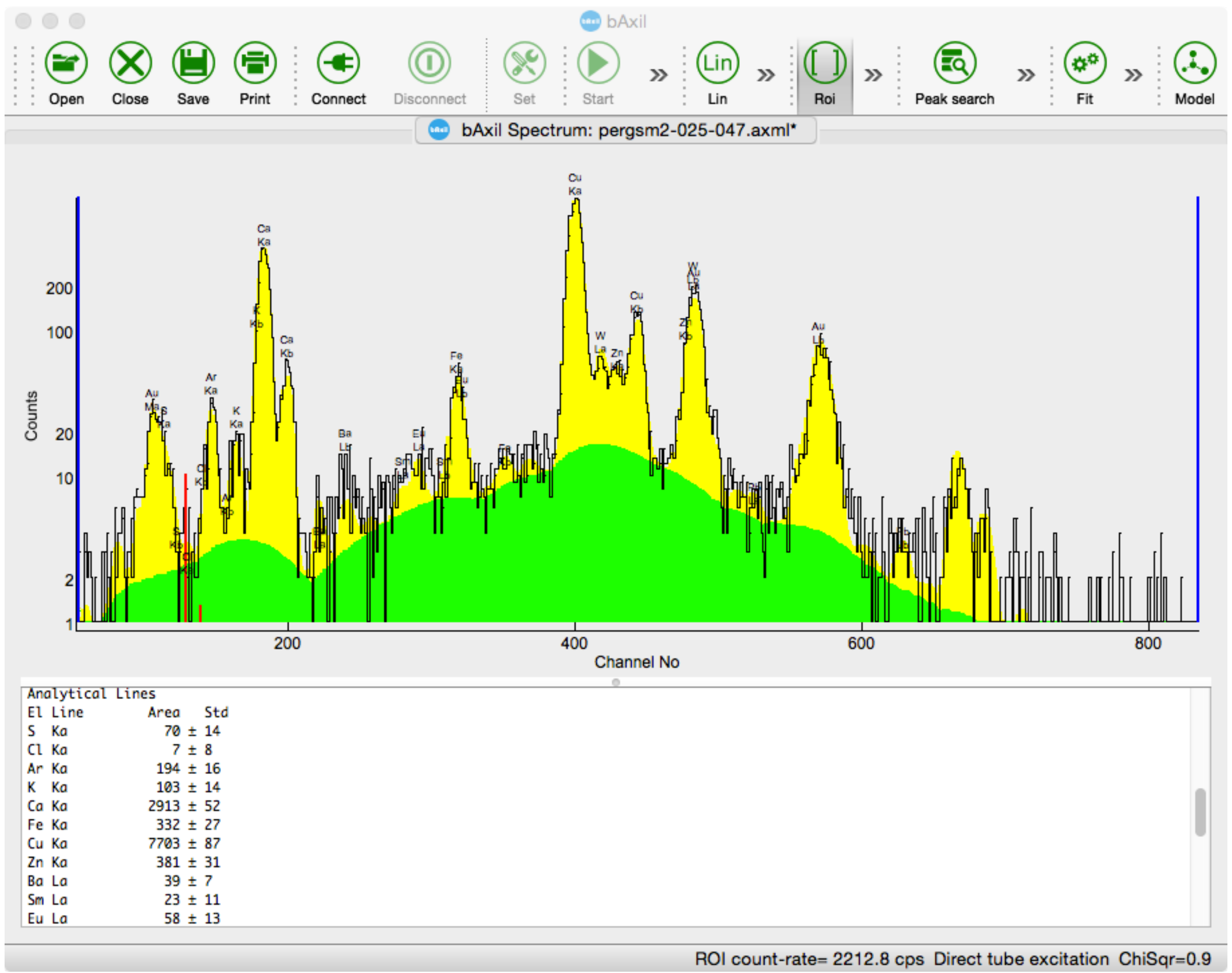
Data set	Type	# cols (x)	# rows (y)	# pixels	# ch	Space on disk MB	CSID space on disk	Compression
PE1-1-fluo	ascii	200	200	40,000	2048	178.5	28.5	6.3
PE2-fluo	ascii	202	201	40,602	2048	181.2	29.7	6.1
PE3-fluo	ascii	268	195	52,260	2048	233.5	37.8	6.2
PE4-fluo	ascii	202	201	40,602	2048	171.2	24.1	7.1
wk	edf	635	584	370,840	2048	3,026.8	237.0	12.8
nk	edf	1020	1199	1,222,980	2048	10,071.4	711.9	14.1
pp	edf	1060	1040	1,102,400	4096	36,130.0	759.5	47.5

4. Solving the Time Problem

Data analysis in XRF

Determining which elements are present and the intensity of the x-ray lines

Interactive building of a model
Fitting the model to the measured spectra



i.e. using non-linear least squares fitting software such as:

- Axil, WinAxil, bAxil
- PyMCA
- …

net element intensities:

- corrected for the background and interferences

Spectrum evaluation

Analytical important information in the spectrum

net peak area of the characteristic lines

= corrected for continuum and peak overlap

Established technique

Least squares fitting of a model to the spectrum

model = analytical function containing intensity of characteristic lines

model: $y(i, a_1, \ldots a_m)$

Spectrum: y_i

$$\chi^2 = \sum_{i=n_1}^{n_2} \frac{1}{\sigma_i^2} \left[y_i - y(i, a_1, \ldots, a_m) \right]^2$$

$$\frac{\partial \chi^2}{\partial a_j} = 0, \quad j = 1, \ldots, m$$

$$y(i) = y_B(i) + \sum_P y_P(i)$$

$$y(i) = y_{\mathrm{Cont}}(i) + \sum_{j\ \mathrm{Elements}} A_j \left[\sum_{k\ \mathrm{lines}} R_{jk} P(i, E_{jk}) \right]$$

Peak shape

Line ratio

Area (Linear parameter)

Continuum function

Peak profile: $$P(i,E_{jk}) = G(i,E_{jk}) + f_s(i,E_{jk}) + f_T(i,E_{jk})$$

Gaussian $$G(i,E_{jk}) = \frac{Gain}{\sigma_{jk}\sqrt{2\pi}} \exp\left[-\frac{(E_i - E_{jk})^2}{2\sigma_{jk}^2} \right]$$

Energy calibration $$E_i = \underline{Zero} + \underline{Gain} \times i$$

(Nonlinear parameters)

Resolution calibration $$\sigma_{jk} = \left[\left(\frac{\underline{Noise}}{2\sqrt{2\ln 2}} \right)^2 + \varepsilon \times \underline{Fano} \times E_{jk} \right]^{1/2}$$

Step: $$S(i,E_{jk}) = \frac{Gain}{2E_{jk}} \operatorname{erfc}\left[\frac{E(i) - E_{jk}}{\sqrt{2}\sigma} \right]$$

Tail: $$T(i,E_{jk}) = \frac{Gain}{2\gamma\sigma \exp\left[-\frac{1}{2\gamma^2} \right]} \exp\left[\frac{E(i) - E_{jk}}{\gamma\sigma} \right] \operatorname{erfc}\left[\frac{E(i) - E_{jk}}{\sqrt{2}\sigma} + \frac{1}{\sqrt{2}\gamma} \right]$$

Because the fitting function contains non-linear parameters

- there is no direct solution to the minimisation problem
- requires a iterative procedures
- involving a lot of calculations

e.g. the Marquardt -Leverberg algorithm

$$\boldsymbol{\beta} = \boldsymbol{\alpha}\delta\boldsymbol{\alpha}$$

$$\chi^2 = \sum_i \frac{1}{\sigma_i^2}\left[y_i - y_0(x_i, a) - \sum_j \frac{\partial y_0(x_i, a)}{\partial a_j}\delta a_j\right]^2$$

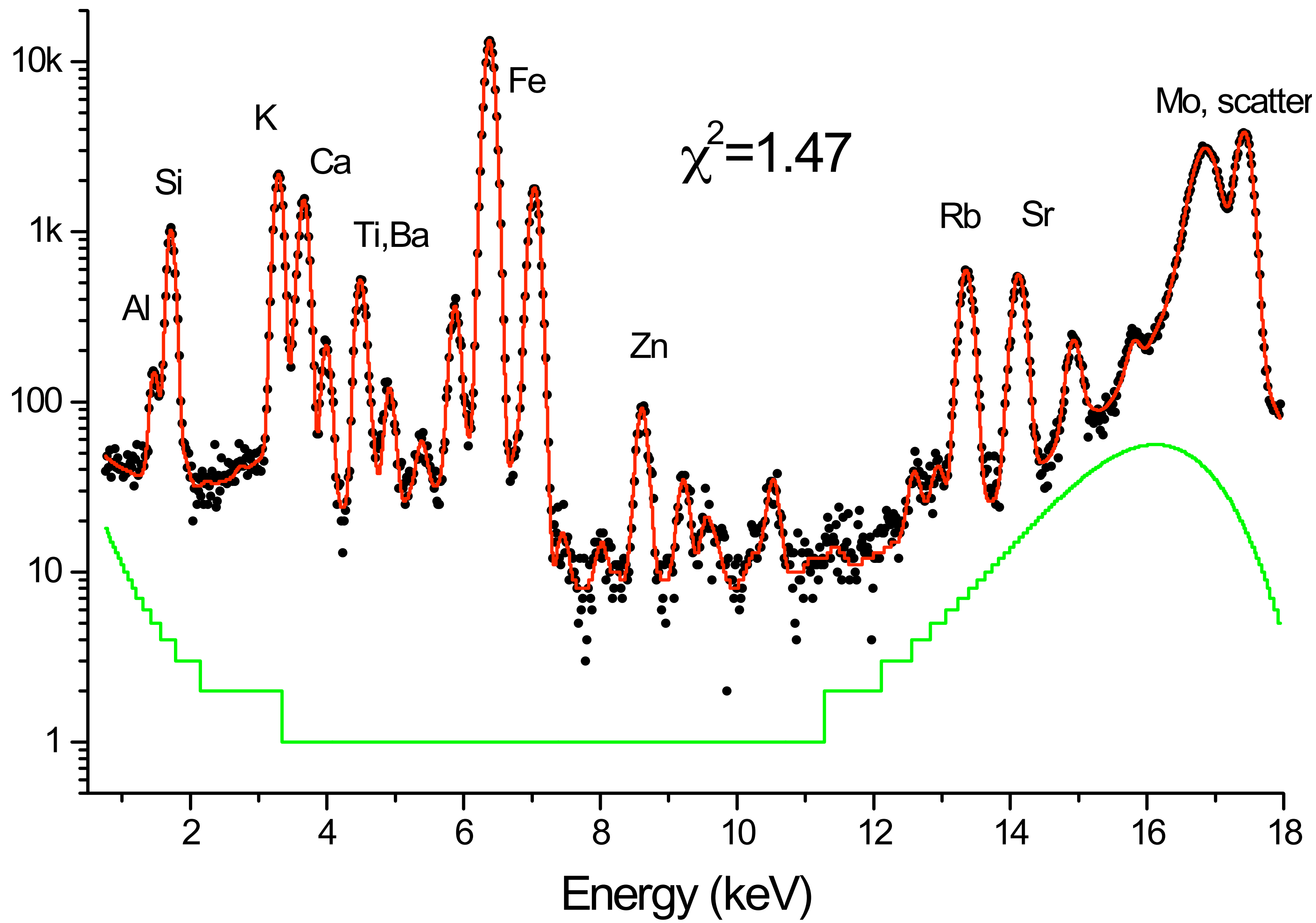
$$\beta_j = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}[y_i - y_0(x_i)]\frac{\partial y_0(x_i)}{\partial a_j}$$

$$\alpha_{jk} = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\frac{\partial y_0(x_i)}{\partial a_j}\frac{\partial y_0(x_i)}{\partial a_k}$$

$$\alpha'_{jk} = \begin{cases} a_{jk}(1+\lambda), & j = k \\ \alpha_{jk}, & j \neq k \end{cases}$$

$$\delta a_j = \sum_{k=1}^{m} \alpha_{jk}^{-1}\beta_k$$

Works very good



Known as

AXIL **A**nalysis of **X**-ray spectra by **I**terative **L**east squares

Axil → WinAxil → bAxil

But is rather slow a few seconds!!!

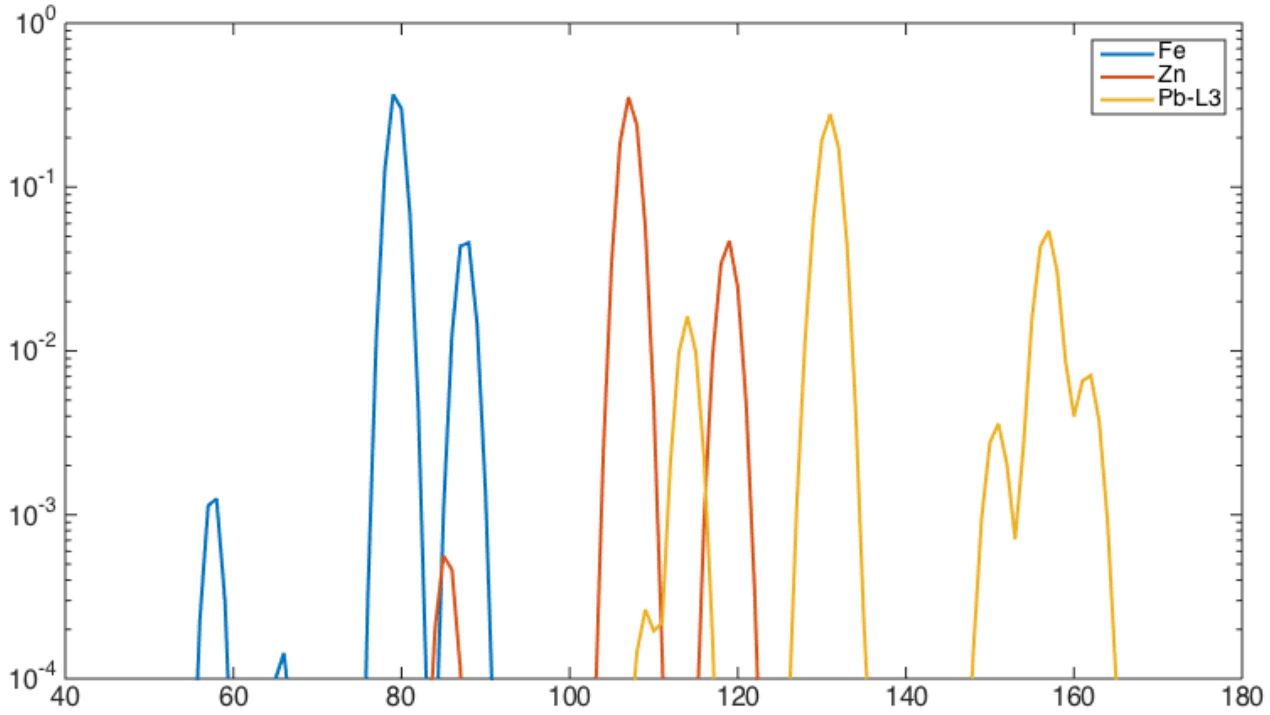
Solving the time problem: "hybrid fitting"

spectrum ≈ linear combination of "element" profiles P_j

$$y_i = y_b(i) + a_1 P_1(i) + a_2 P_2(i) + \ldots + a_m P_m(i)$$

except for the continuum

Continuum stripping $$y_i - y_b(i) = a_1 P_1(i) + a_2 P_2(i) + \ldots + a_m P_m(i)$$



Element profiles obtained by non-linear least squares

$$P_j(i) = A_j \sum_k R_k G(i, E_{jk})$$

$$G(i, E_{jk}) = \frac{Gain}{\sigma_{jk}\sqrt{2\pi}} \exp\left[-\frac{\left(E_i - E_{jk}\right)^2}{2\sigma_{jk}^2} \right]$$

Fitting model $$y_i' = a_1 P_1(i) + a_2 P_2(i) + \ldots + a_m P_m(i) \qquad i = i_{first} \ldots i_{last}$$

General linear model in matrix notation $$\mathbf{y} = \mathbf{P}\alpha + \varepsilon \qquad \mathbf{P} = \mathbf{P}_{n\times m}$$

Least squares solution of $\boldsymbol{\alpha}$ $$\mathbf{a} = \left(\mathbf{P}^T\mathbf{P}\right)^{-1}\mathbf{P}^T\mathbf{y} \qquad \left(\mathbf{P}^T\mathbf{P}\right)^{-1}_{m\times m}$$

Practical

1. Calculate the profiles P_j for each element using non-linear least squares

2. Pre-calculate the matrix $$\mathbf{M} = \left(\mathbf{P}^T\mathbf{P}\right)^{-1}\mathbf{P}^T$$

3. For each spectrum **y**,
 strip the continuum
 obtain the element contributions a_j $$\mathbf{a} = \mathbf{M}\mathbf{y}$$

ONE matrix multiplication per spectrum!!!

Hybrid aspect

Do a non-linear fit at regular intervals and
recalculate the peak profiles and the **M** matrix

Accounts for possible calibration changes during the scan
(peak shift and broadening)

Retains all the features of the non-linear fitting procedure
building the model (which elements to include)
calibrate the spectrum

The price we pay for this

Unweighted least squares solution

No estimate of the uncertainties in the parameters

No account for sum peaks…

But: it goes FAST

5. Results

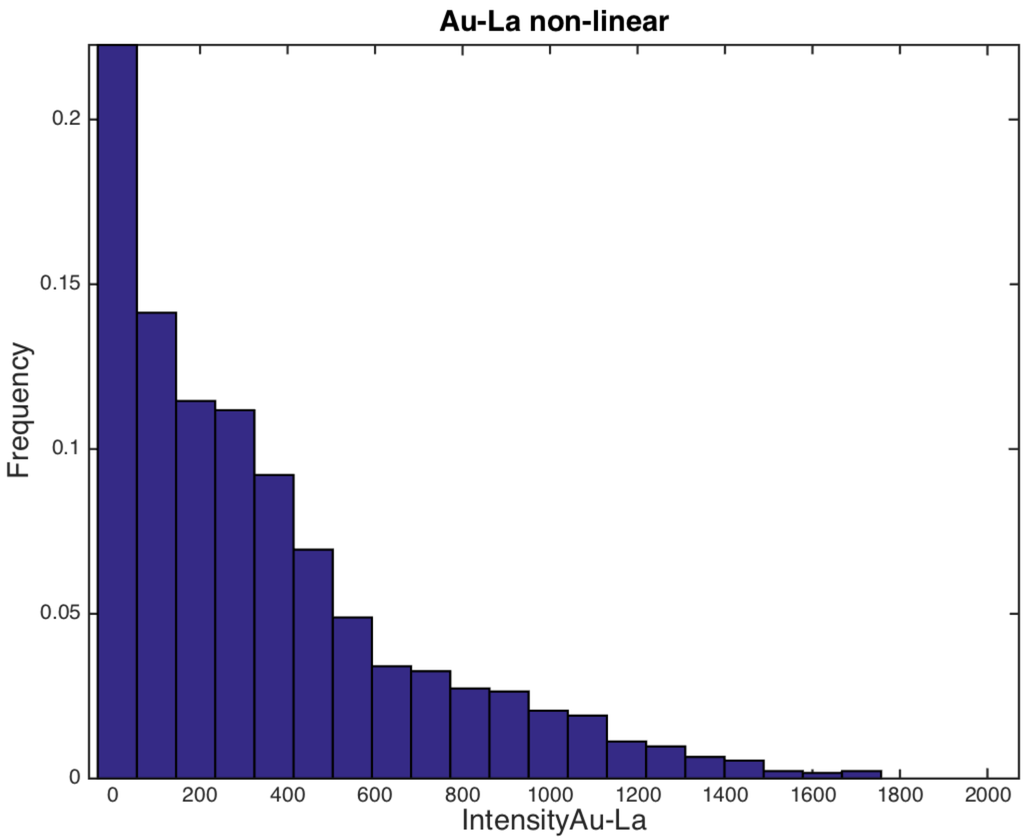
Validity of the hybrid fitting results: MA-XRF of an ancient parchment

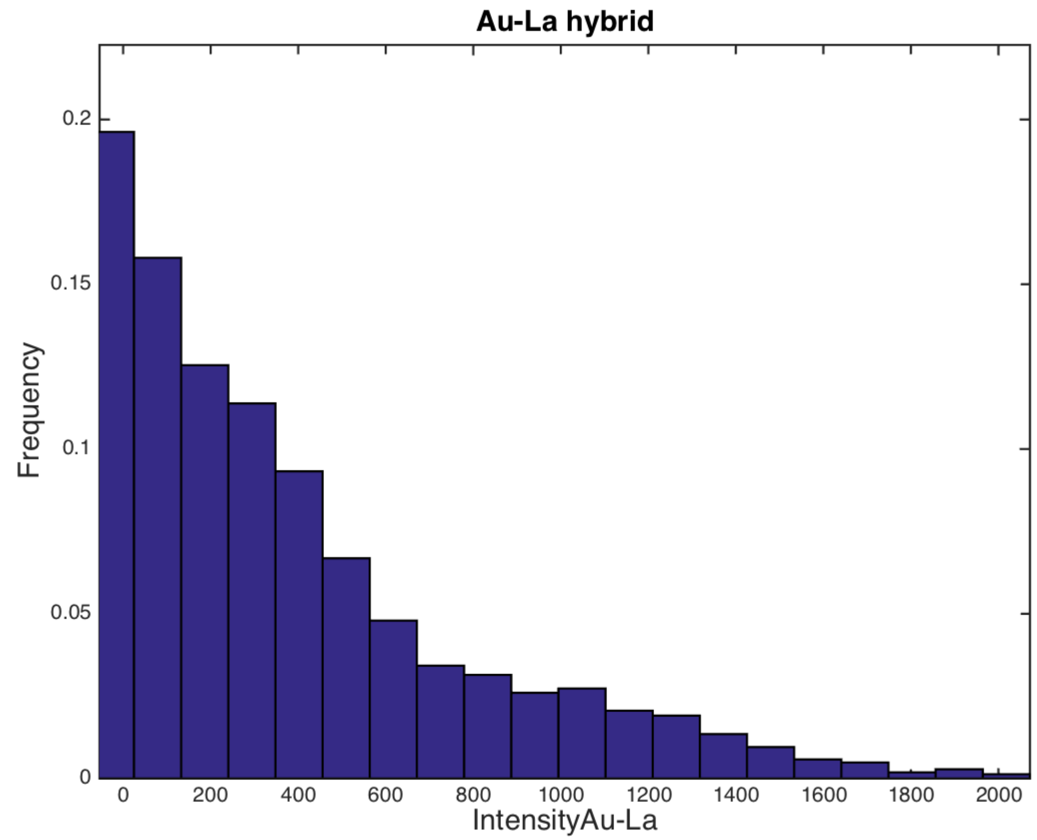
Comparison of results

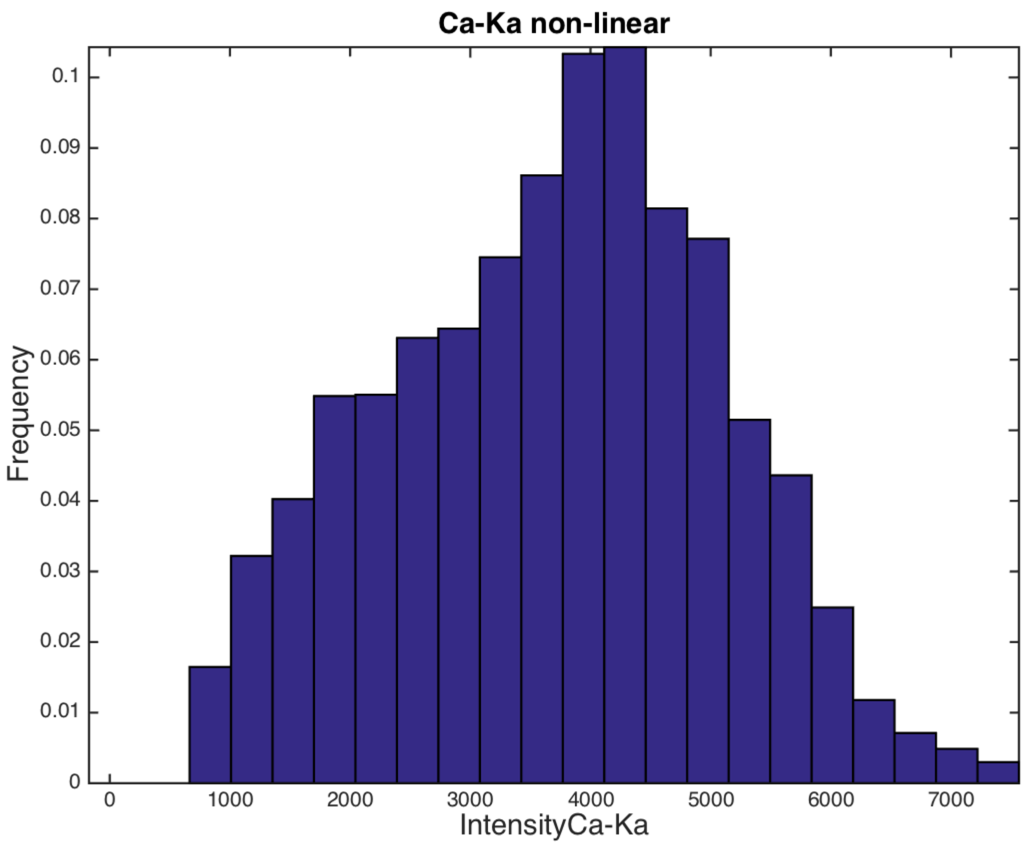
Parchment: 109 × 49 = 5341 spectra

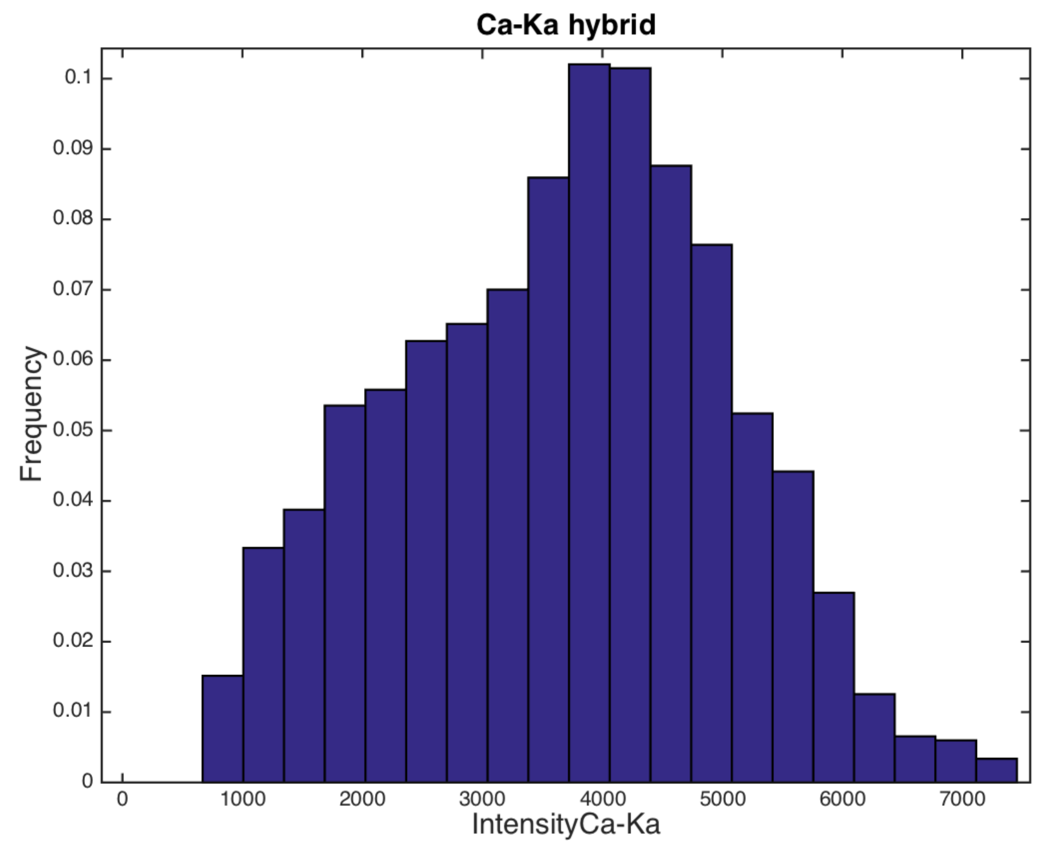
	Standard non-linear fitting	hybrid fitting
Au-Lα		
Ca-Kα		
Fe-Kα		
Cu-Kα		
	27:44 min 0.31 s/spectrum	**0:14 min** **0.0026 s/spectrum**

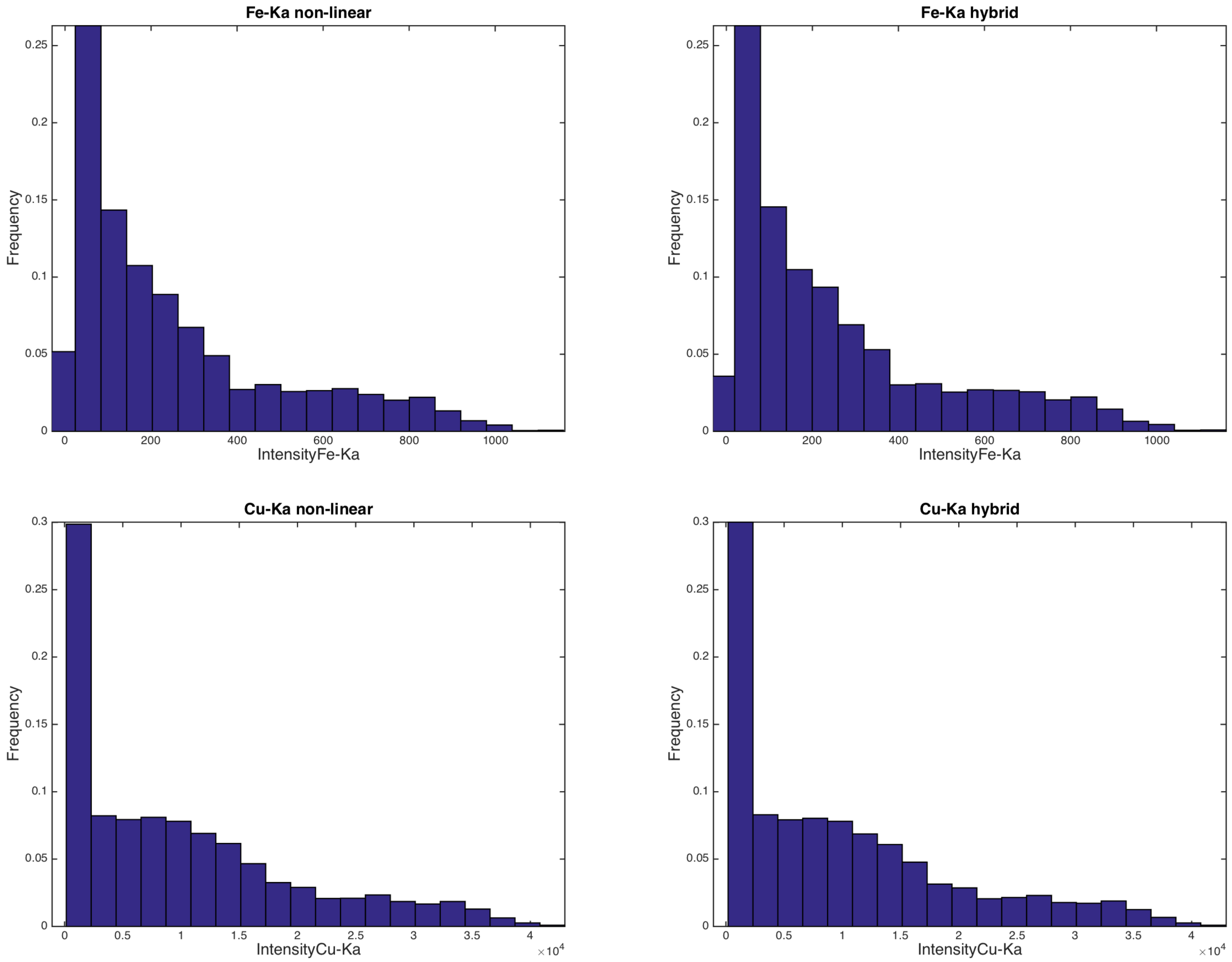
Comparison of results

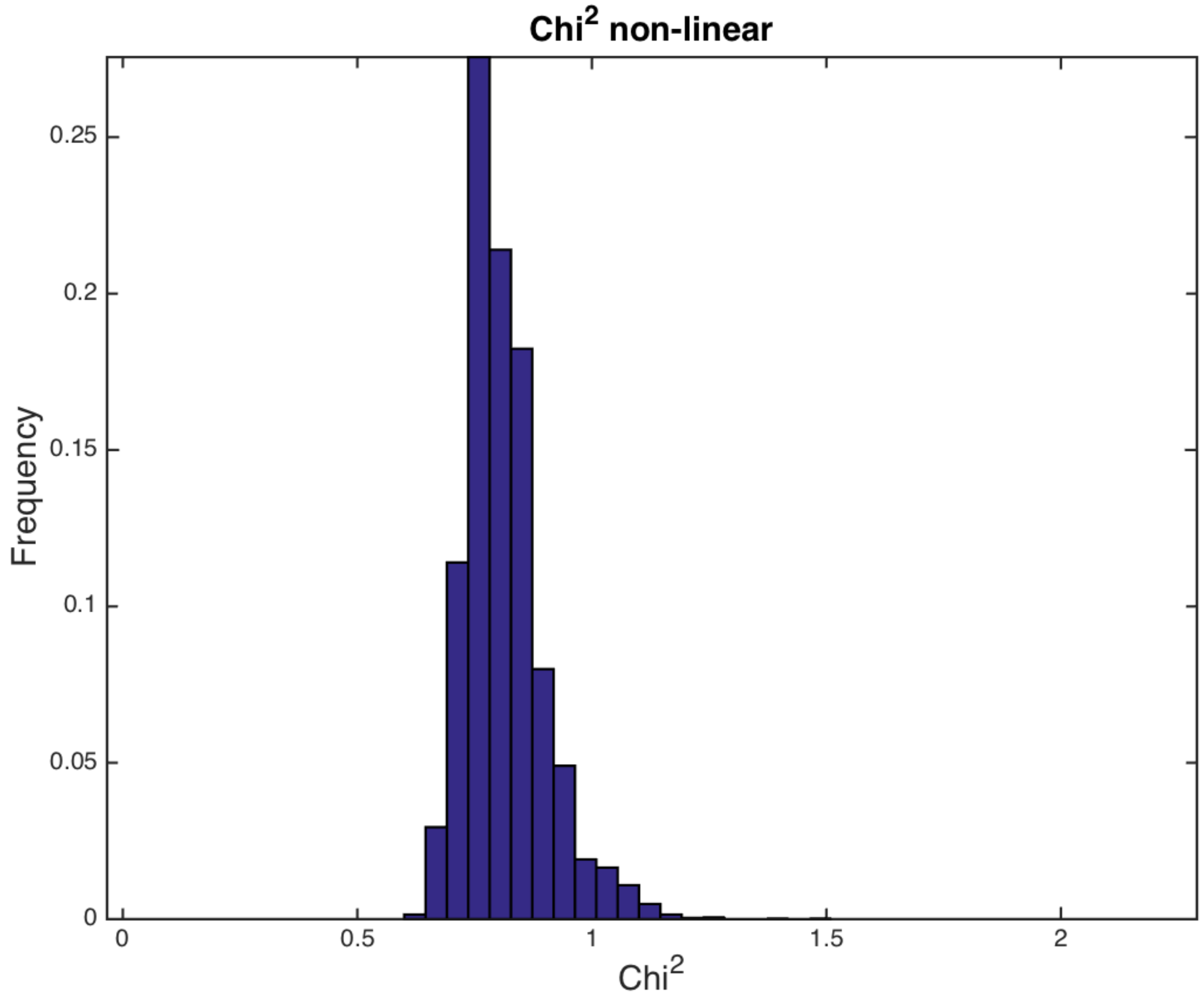




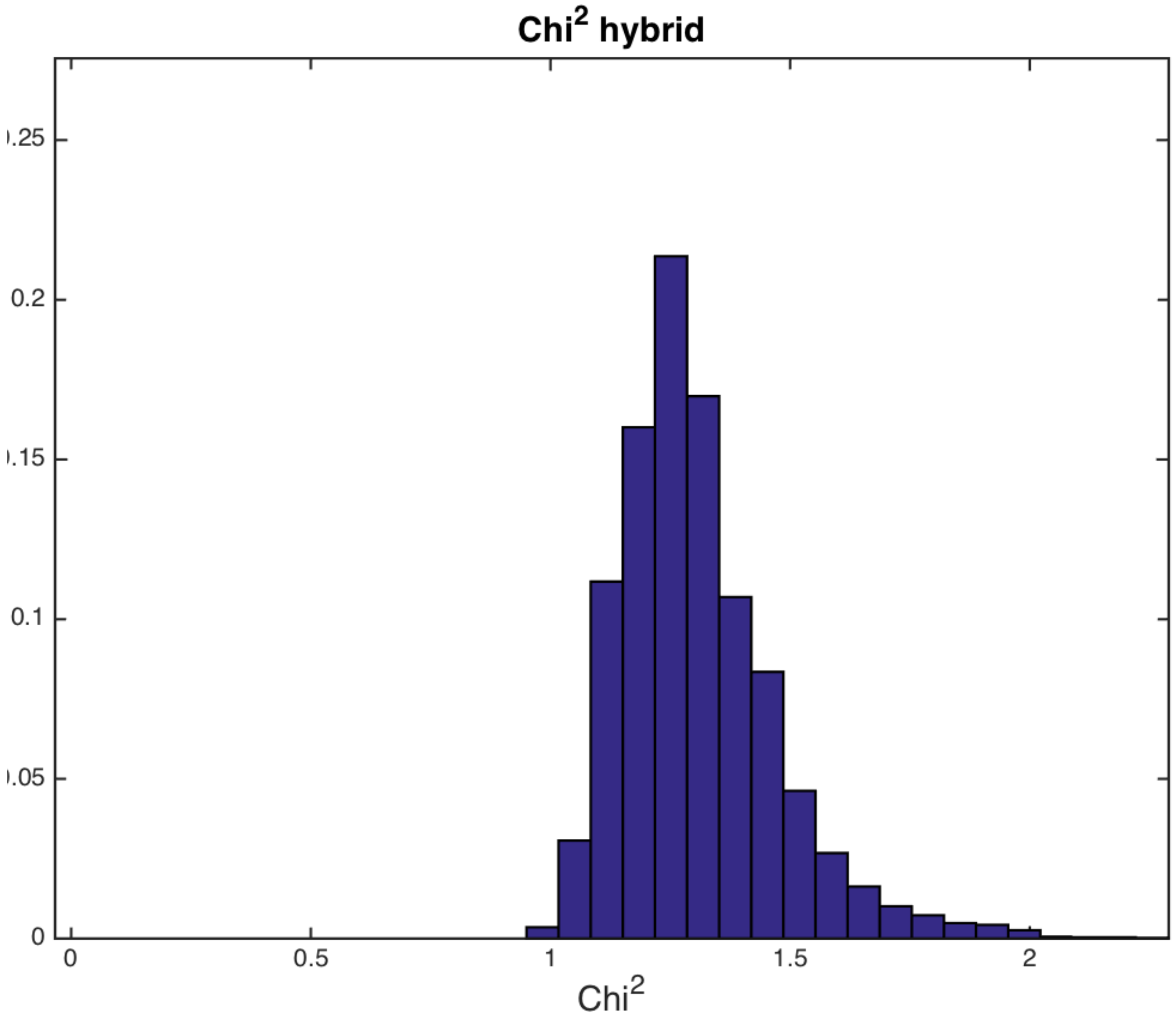





Fe-Ka non-linear
Frequency
IntensityFe-Ka
Fe-Ka hybrid
Frequency
IntensityFe-Ka
Cu-Ka non-linear
Frequency
IntensityCu-Ka
$\times 10^4$
Cu-Ka hybrid
Frequency
IntensityCu-Ka
$\times 10^4$



Chi^2 non-linear
Frequency
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1.5
2
Chi^2



Chi^2 hybrid
0.2
0.1
0
0
0.5
1
1.5
2
Chi^2

Efficiency of the hybrid fitting: speed

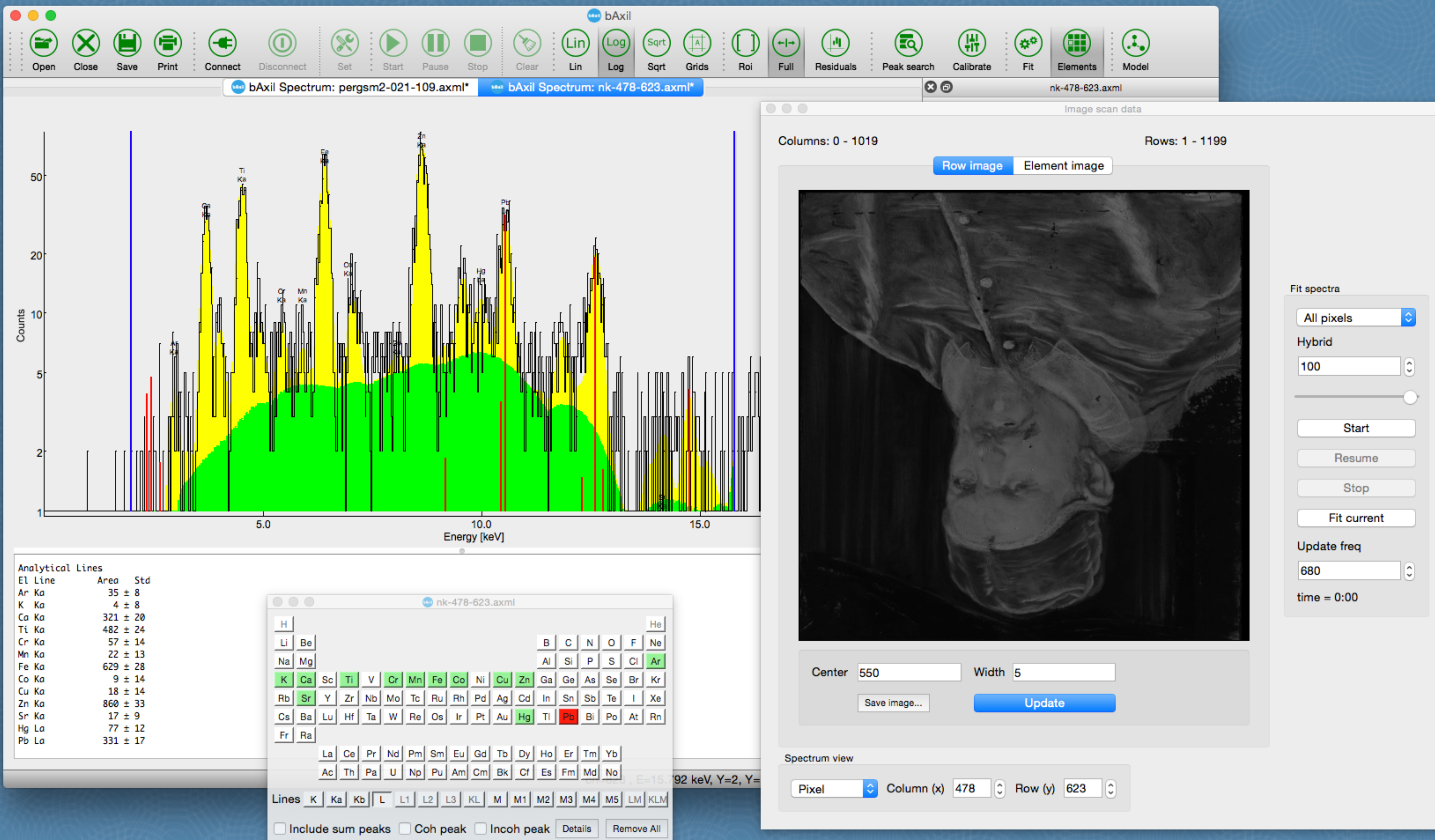
Computer:
MacBook Pro
2.53 GHz intel core i5
8 GB memory

Parchment:
5341 spectra
Non-linear least squares fit: 1164 seconds
Hybrid fitting: 14 seconds
Factor: ~120

MA-XRF scan 1020 × 1199 pixels = 1 222 980 spectra

Fitting model of 13 elements (profiles)
Fitted in 7.10 minutes or 0.35 msec / spectrum

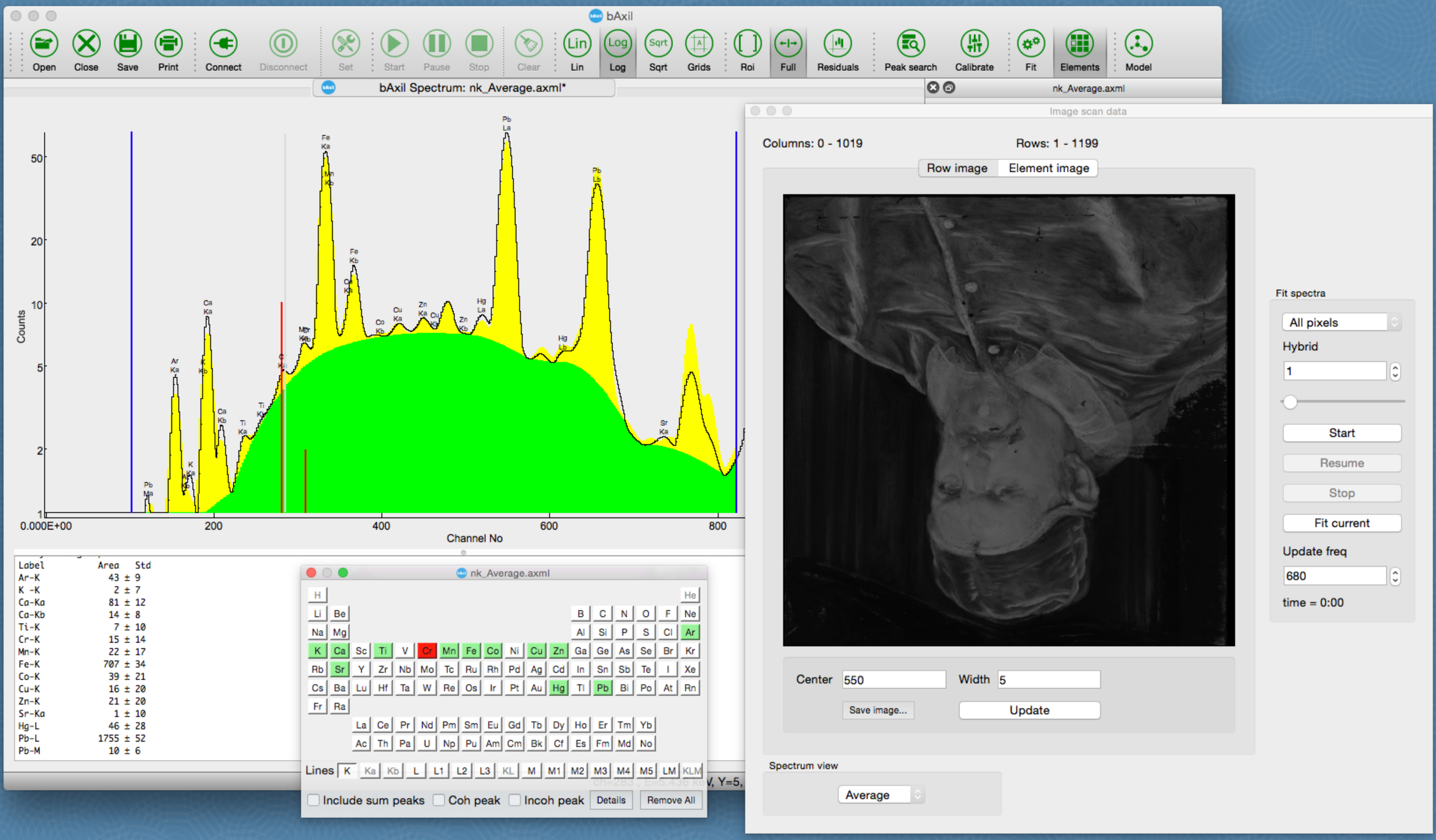
Interactive capabilities: MA-XRF scan 1020 × 1199 pixels = 1 222 980 spectra



interactive fitting of pixel 478-623

interactive fitting of pixel 241-918

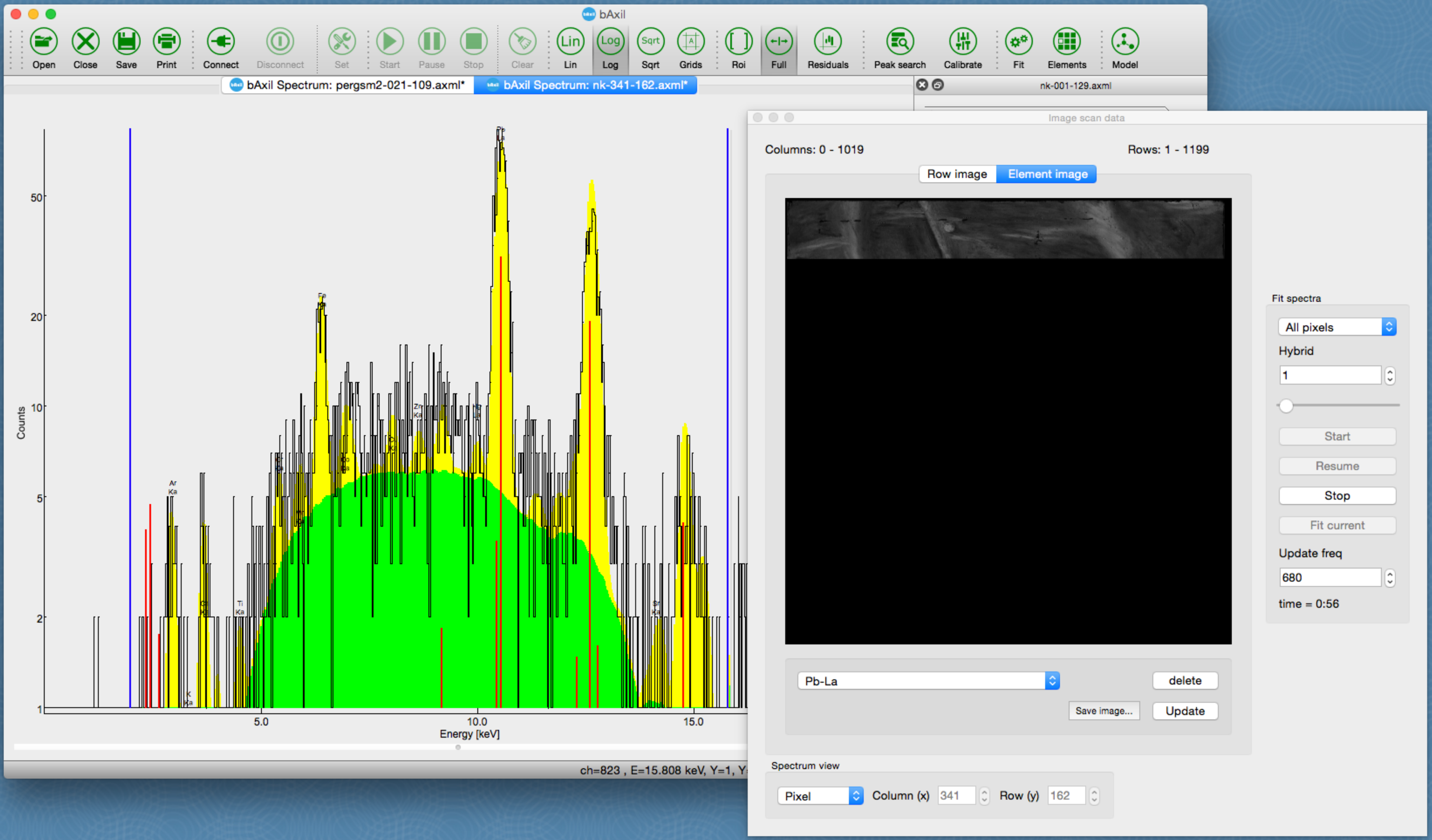
Investigating the maximum intensity spectrum



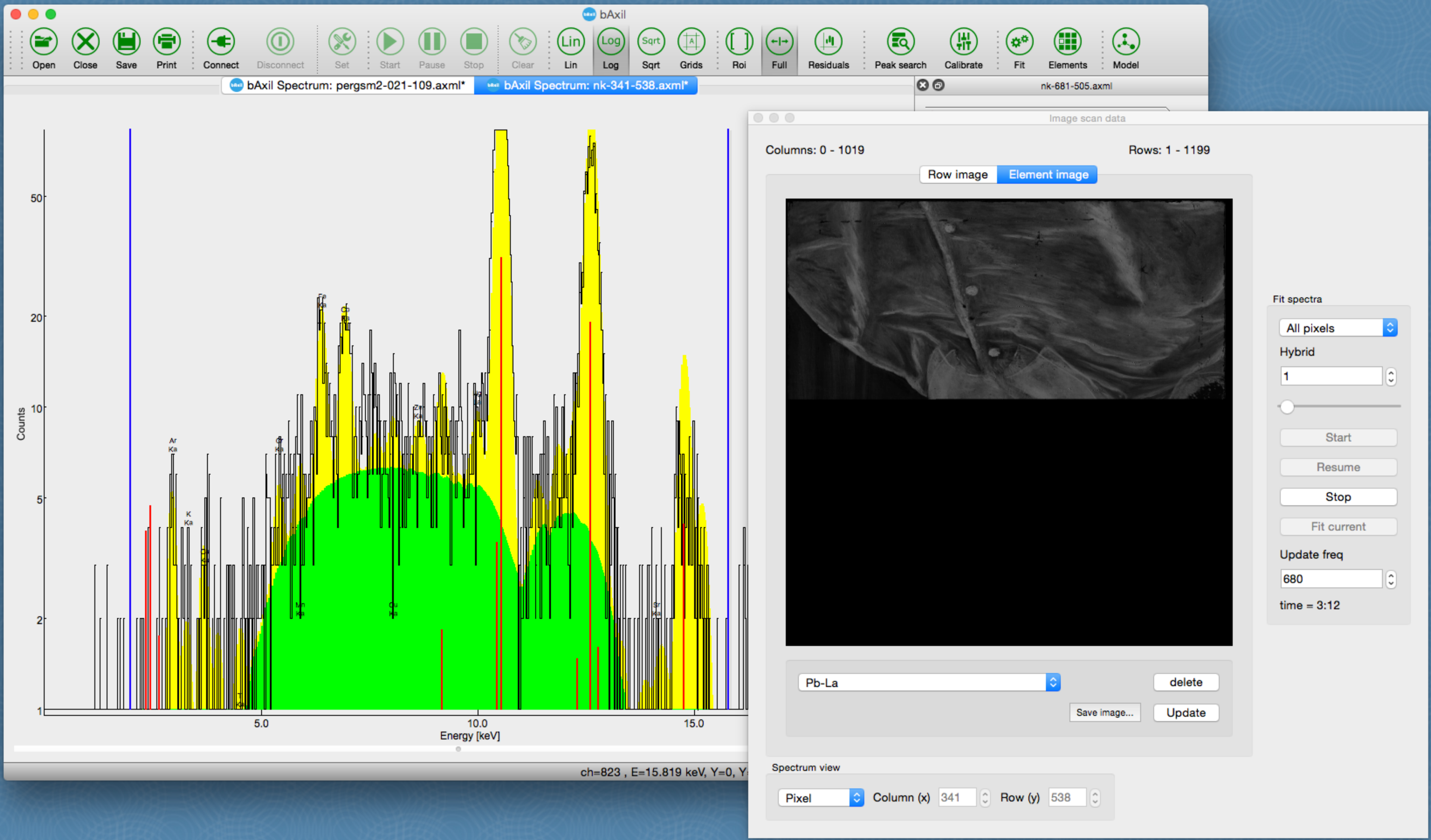
Investigating the average spectrum

Starting the hybrid fitting

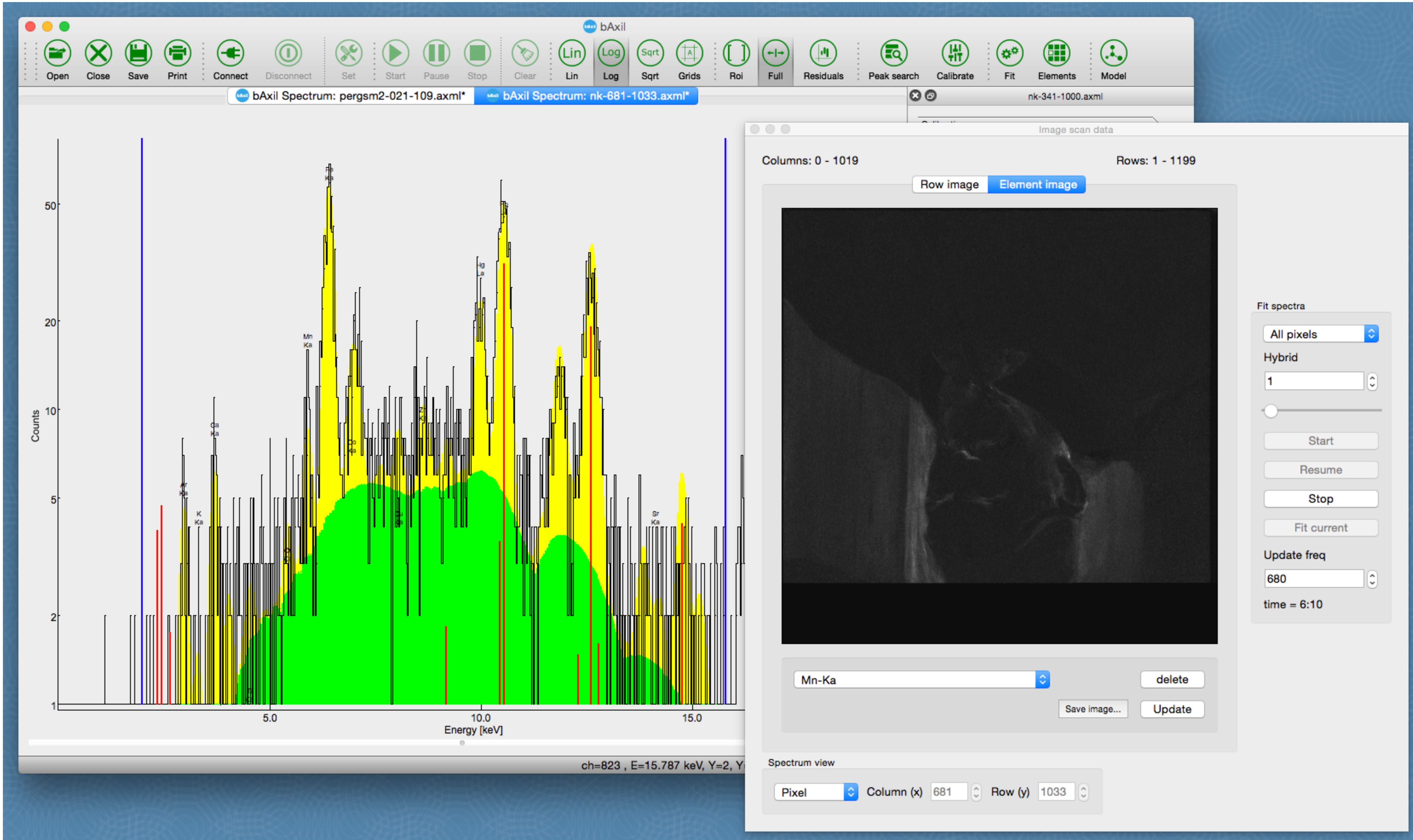
displaying the Pb-La image so far



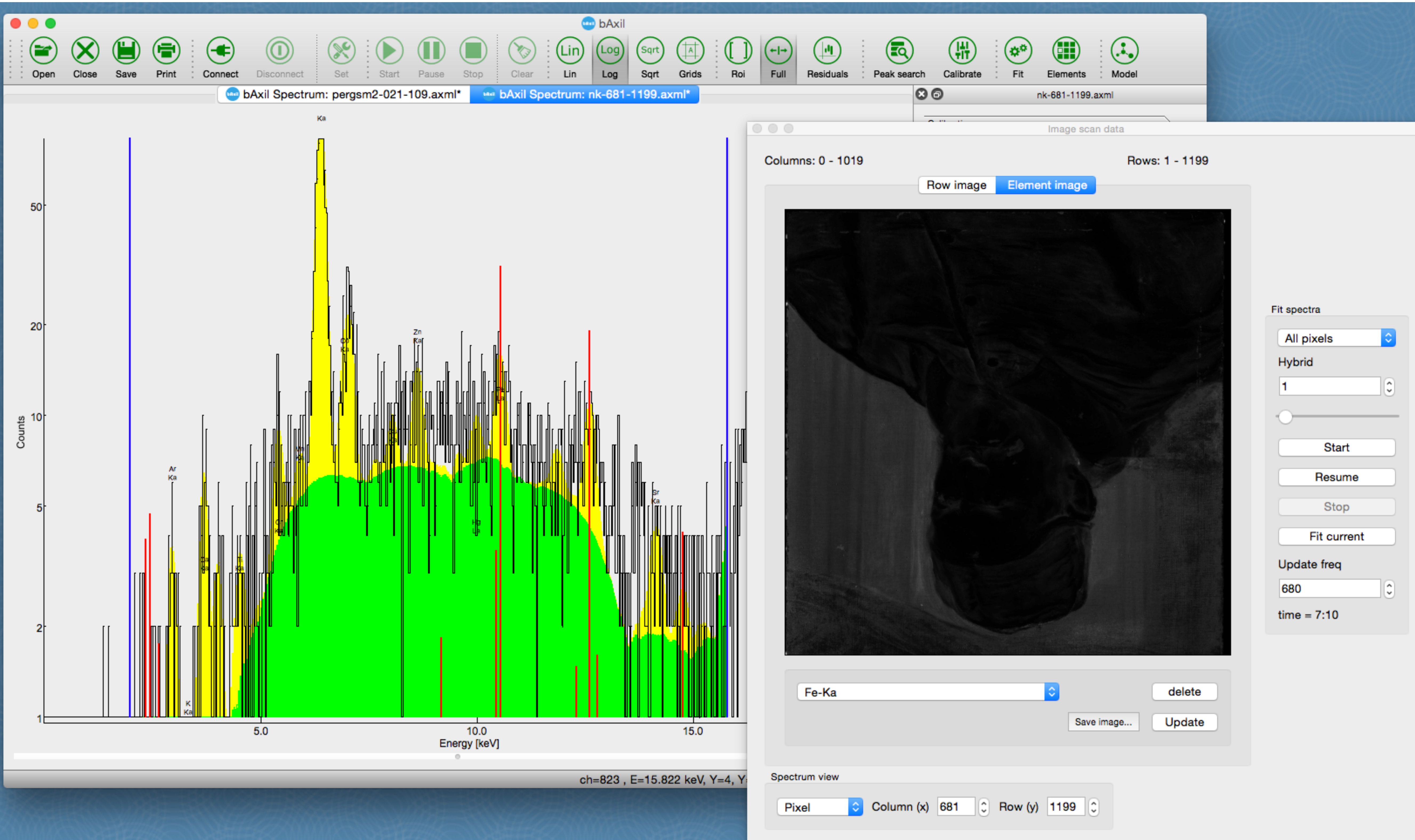
Results after 3 minutes



Display of the Mn-Ka image after 6:10 minutes



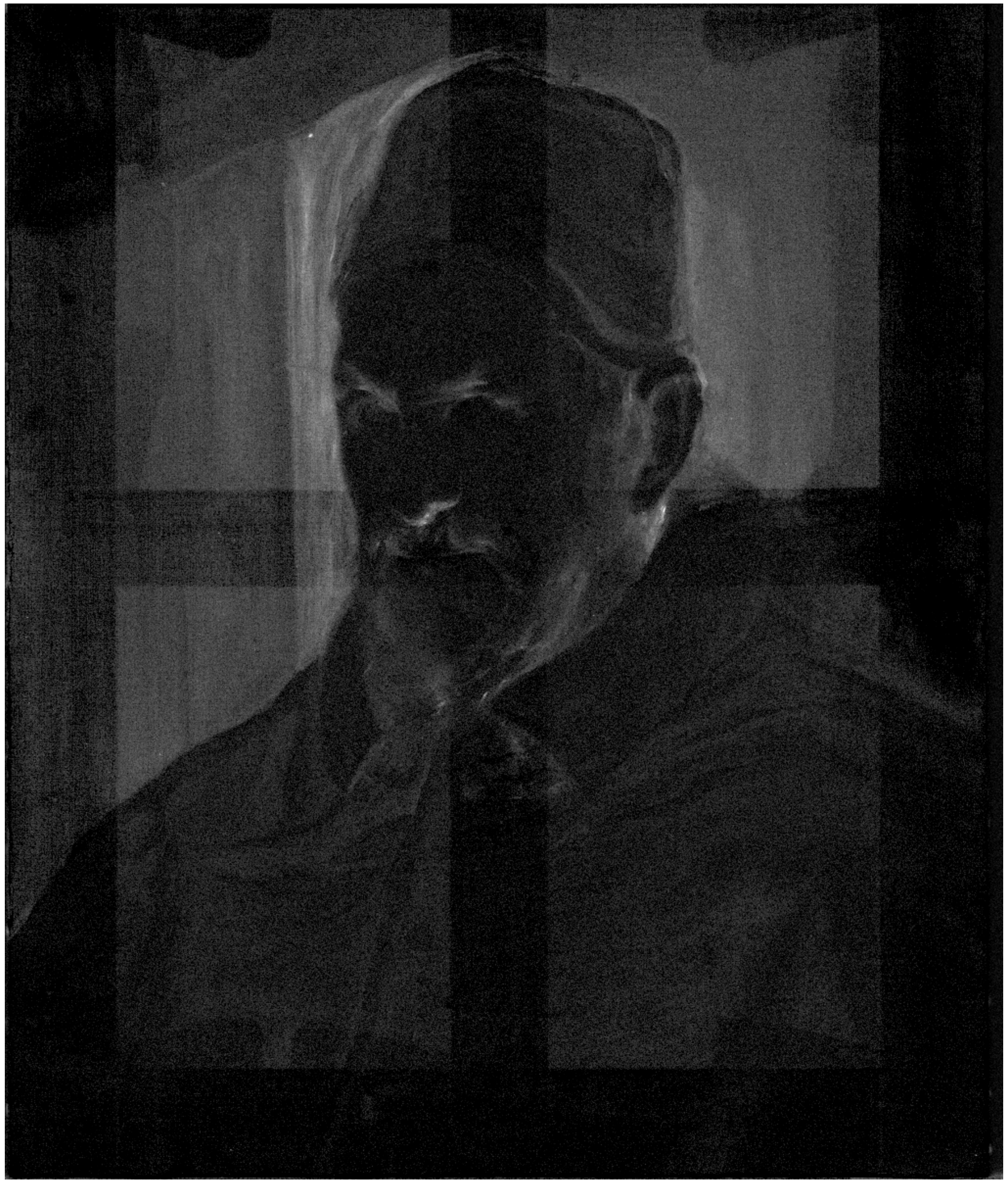
End of the analysis: Fe-Ka image



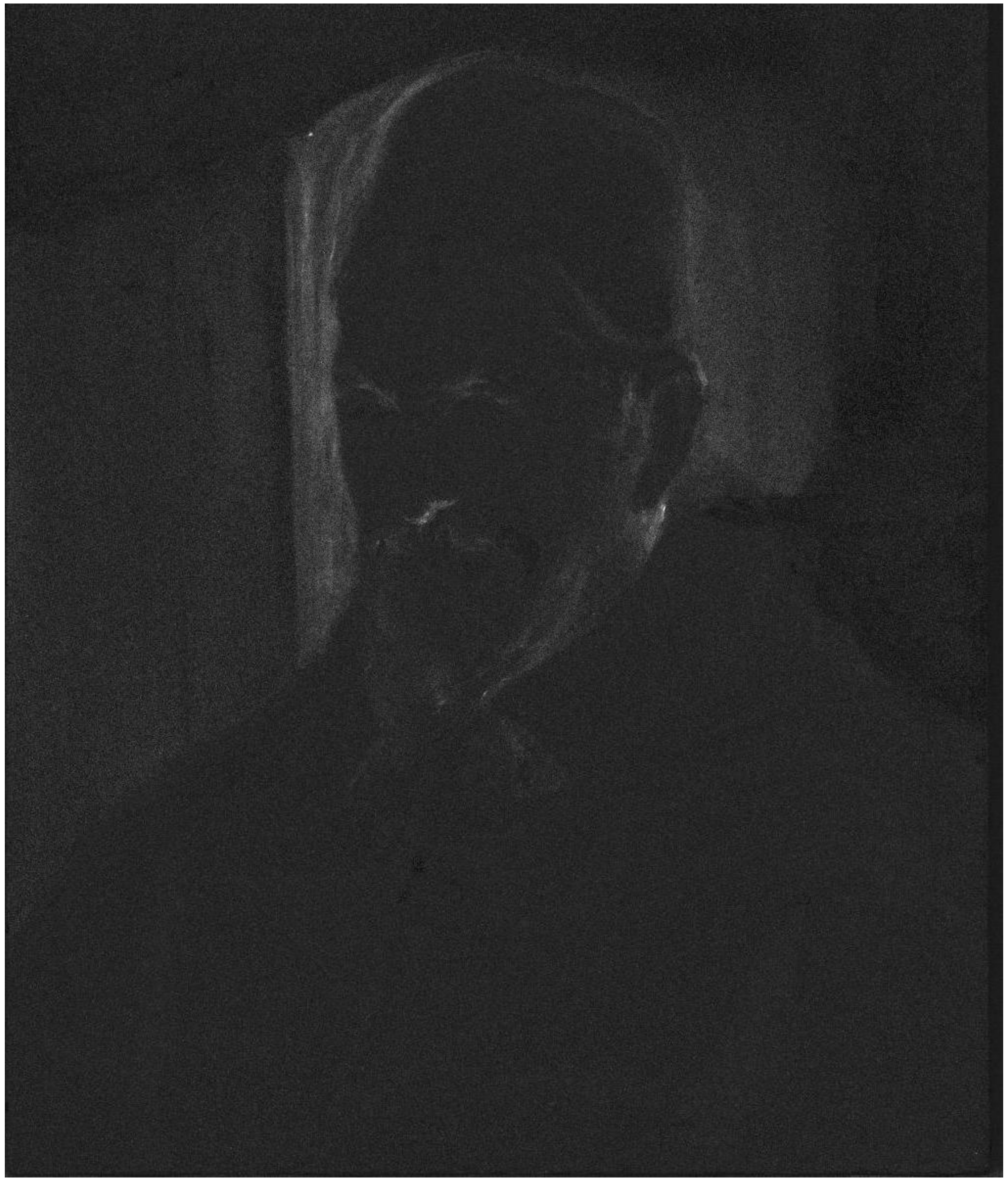
End of the analysis: the Pb-La image

Speed: 7:10 min = 430 s or 0.35 ms/spectrum

Mn image without background removal



Mn image with hybrid fitting



6. Applications

AXES
Antwerp X-ray Analysis, Electrochemistry & Speciation
University of Antwerp

5. Application

The Ghent Altarpiece Adoration of the Lamb Brothers Van Eyck 1432

Restoration 2012 - 2017

Original data (ESRF-EDF format)
1041 files of 34.7 MB each = 36.13 GB

1060 × 1040 pixels @ 0.2 s = 1,102,400 spectra = 2.5 d measuring

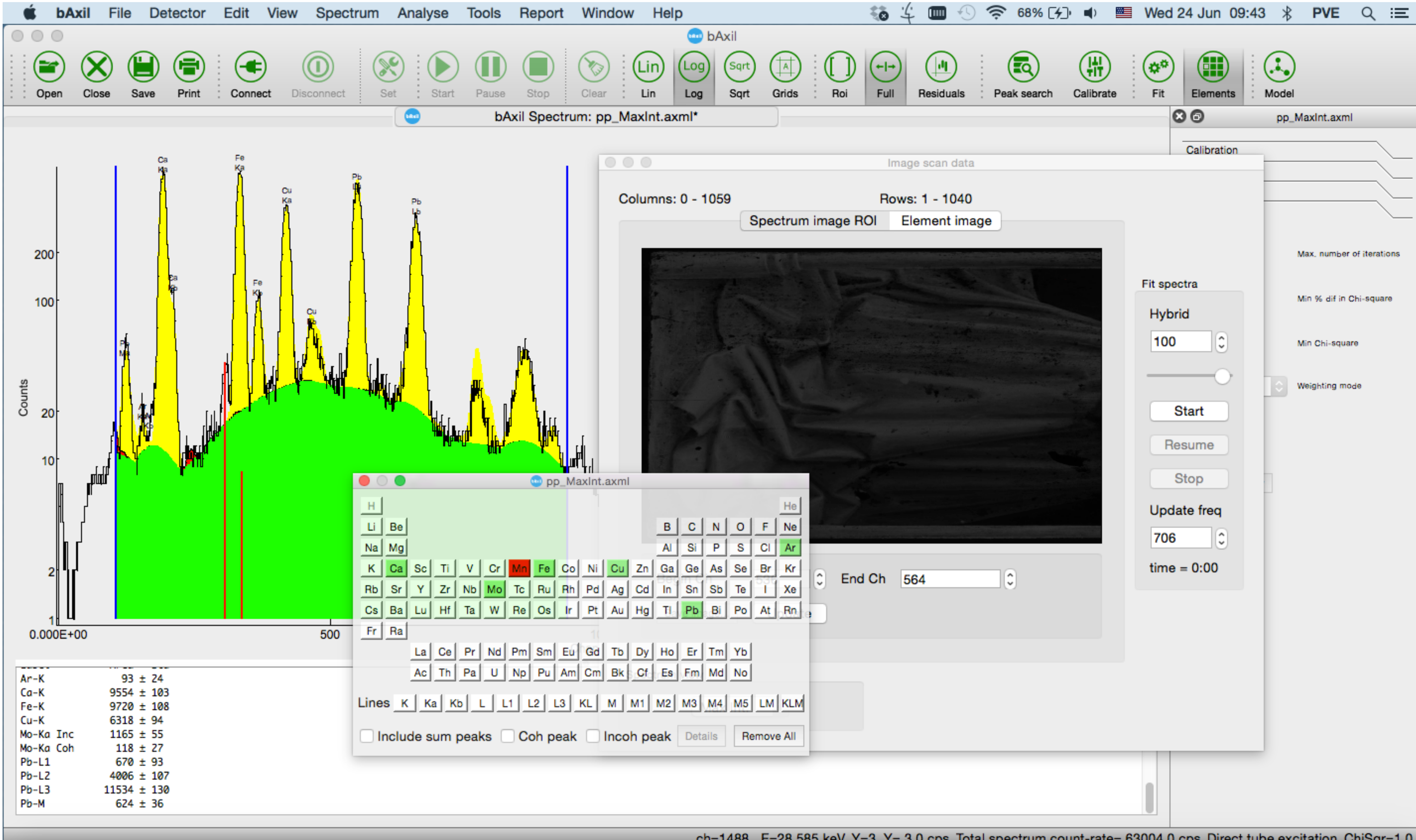
Compressed: 760 MB (factor 48!!!)

Multivariate analysis using the spectral data

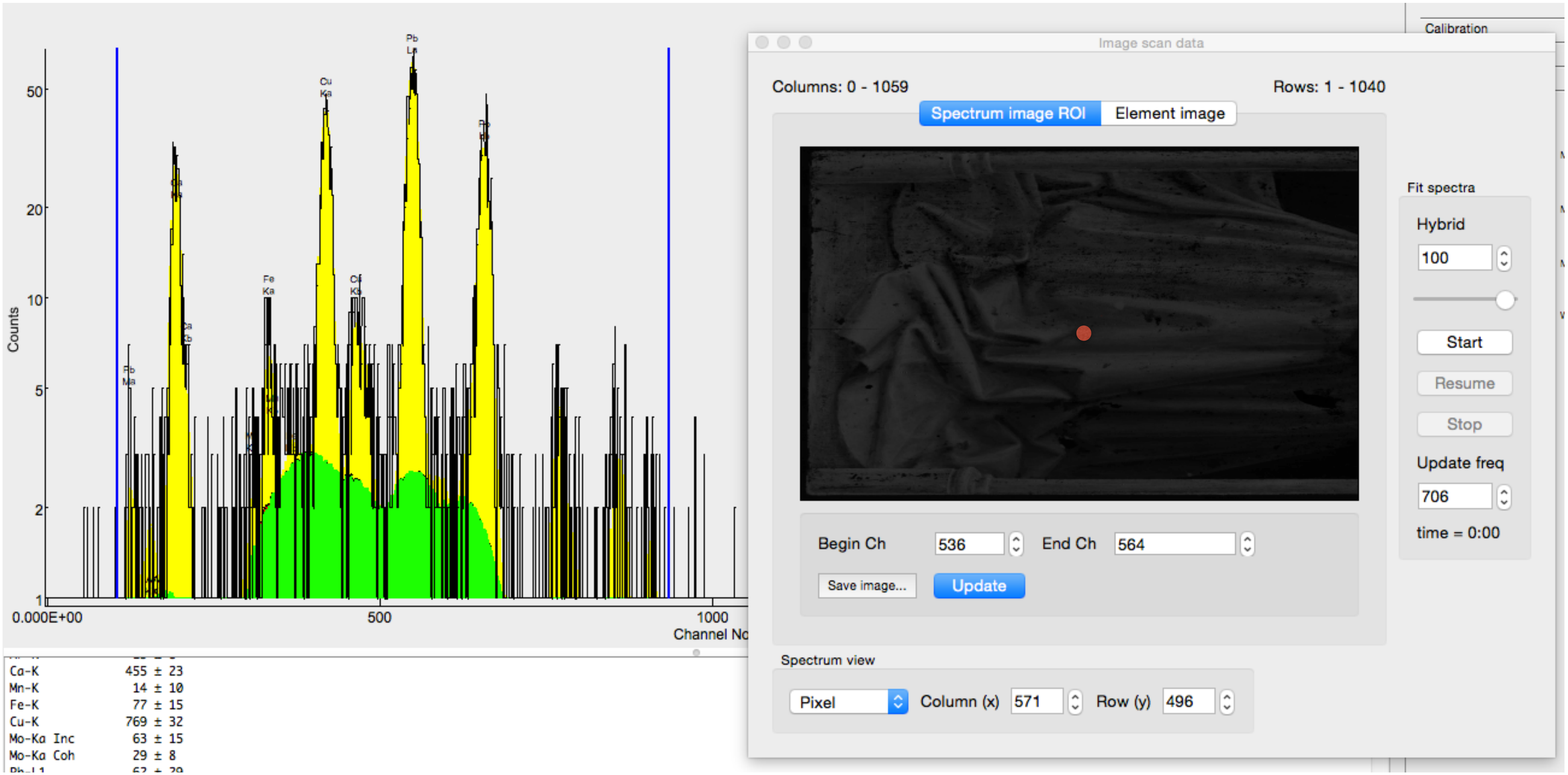
Wooden frame causes background in the spectra

Method implemented in software package bAxil

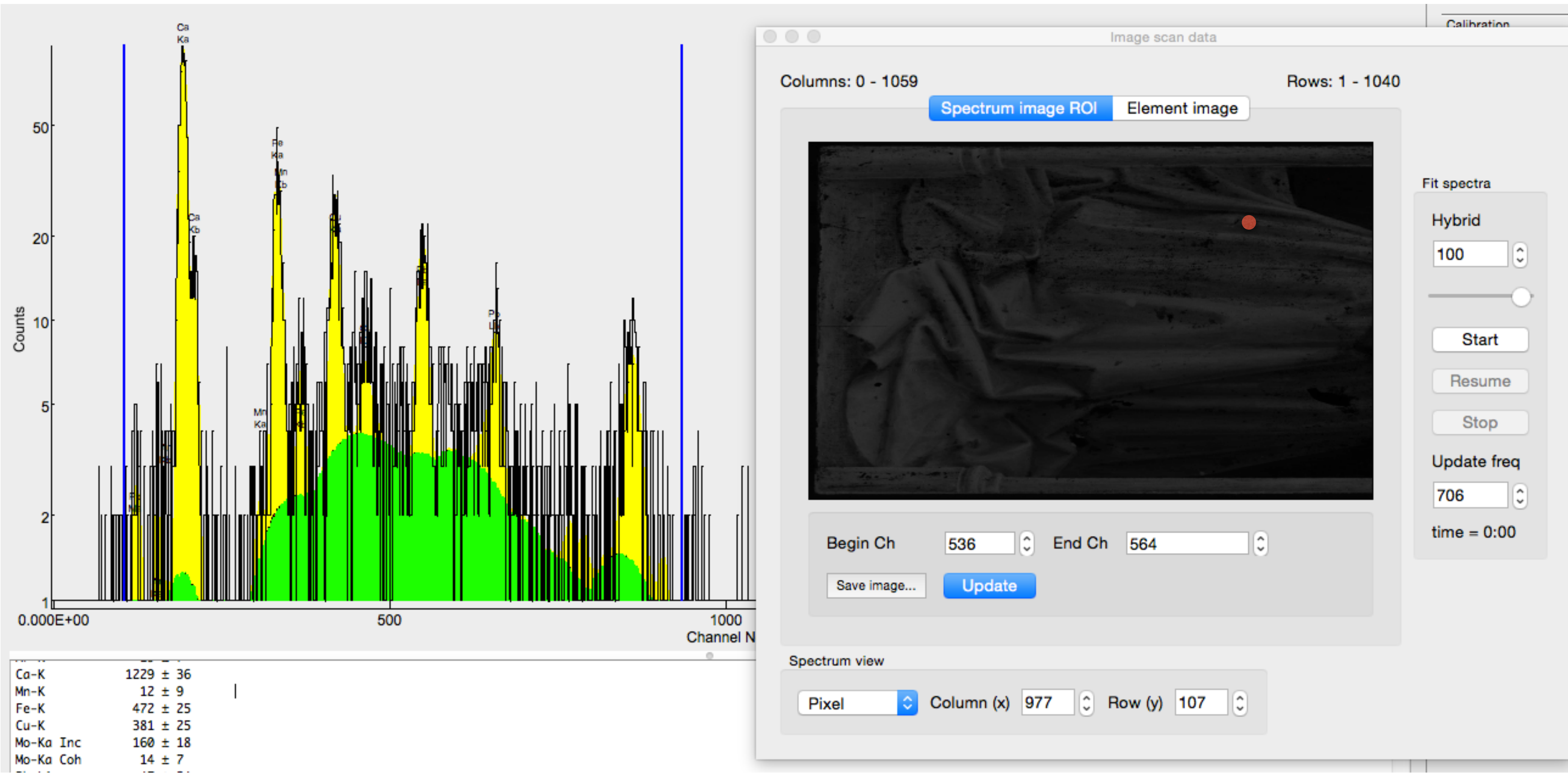
Building the fitting model based on the "maximum intensity spectrum"



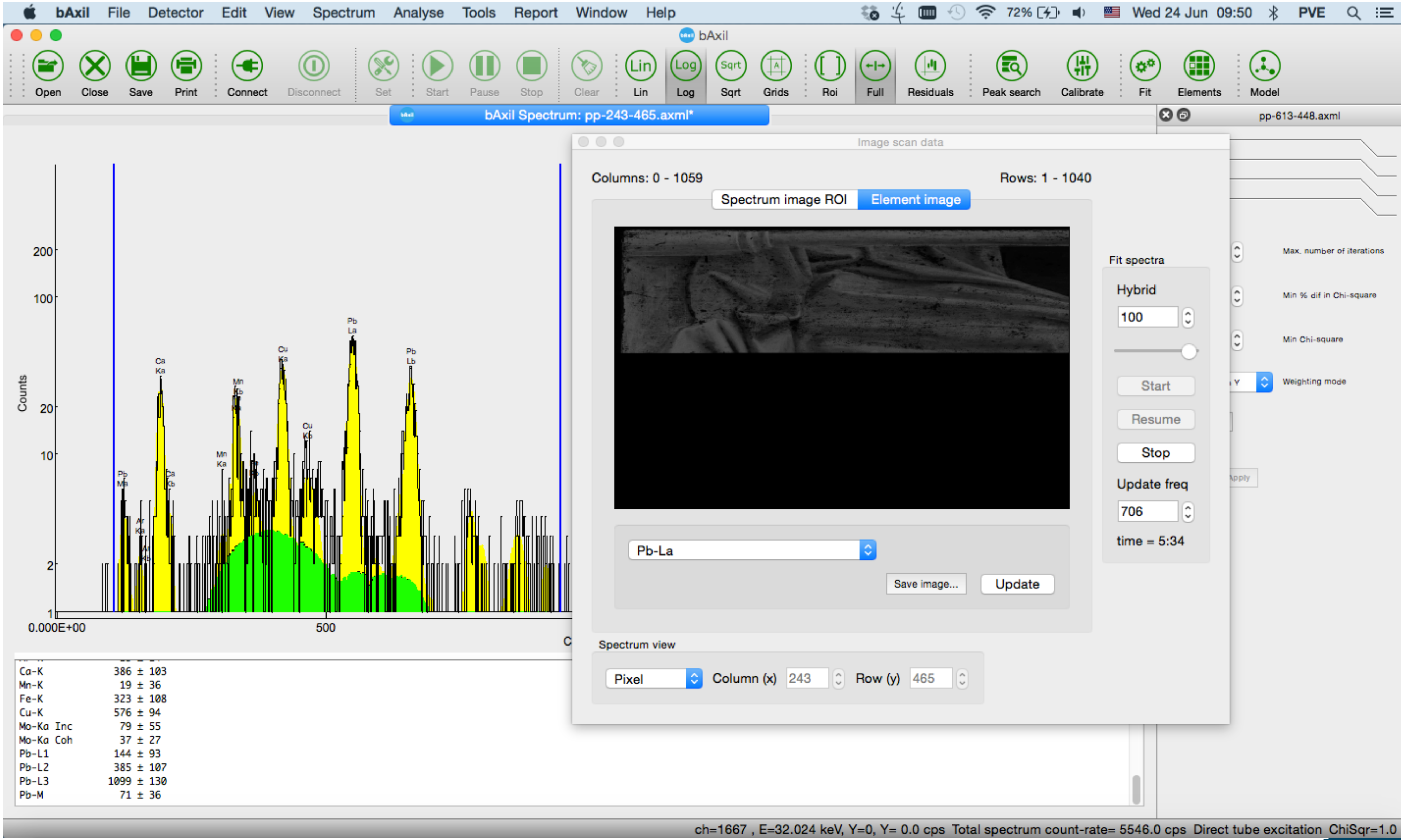
Looking at pixel 571,496



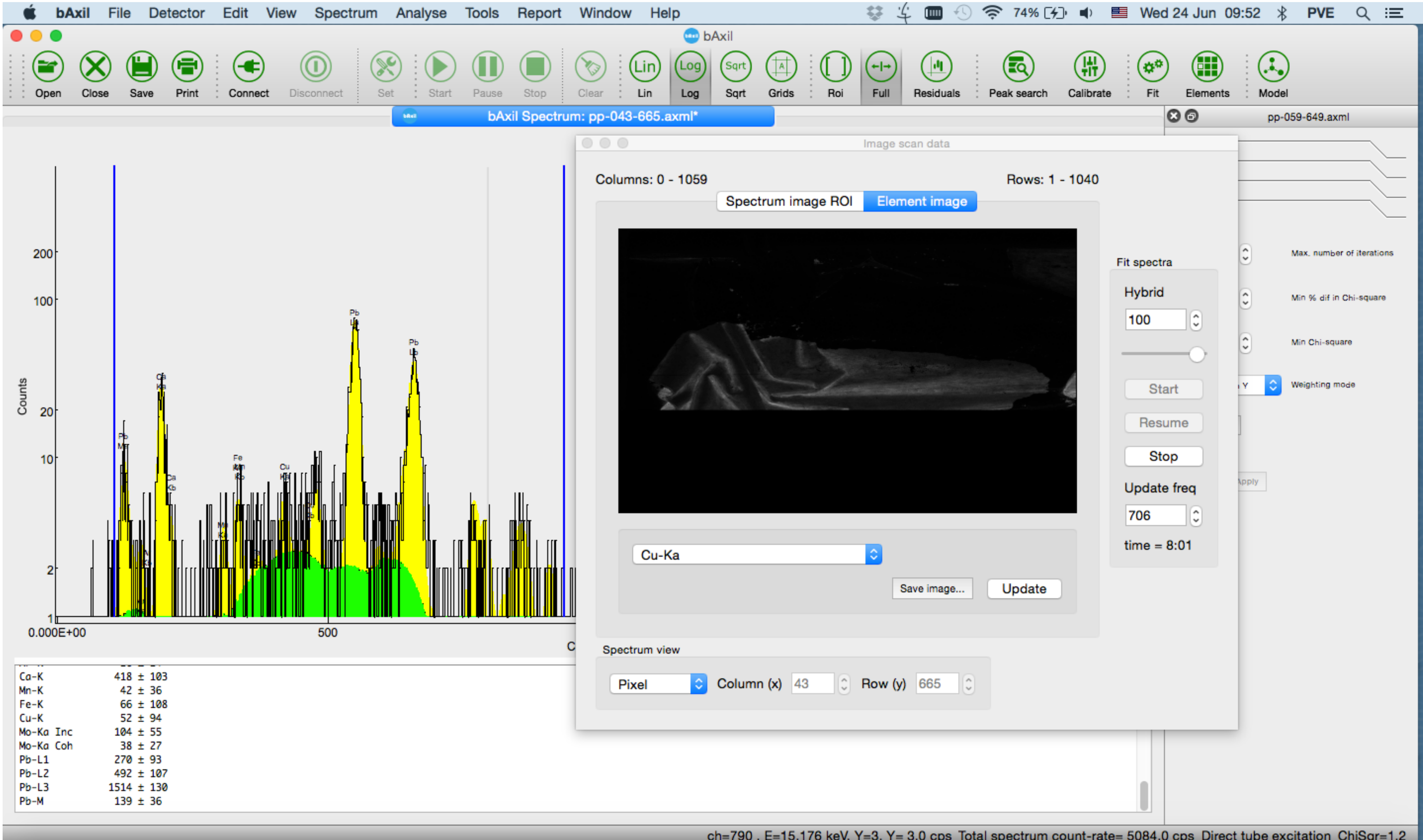
Looking at pixel 977,107



Cu-Ka image after fitting 465 rows of 1060 spectra (5.34 min)



Cu-Ka image after fitting 665 rows of 1060 spectra (8.01 min)



Analysed: hybrid fitting 6 elements in 12.25 min = 0.6 ms / spectrum

Pb La

Cu Ka

Cu overpainting
Restoration in 16th century

Ca Ka

Fe Ka

other defects and bad restorations

7. Conclusions:

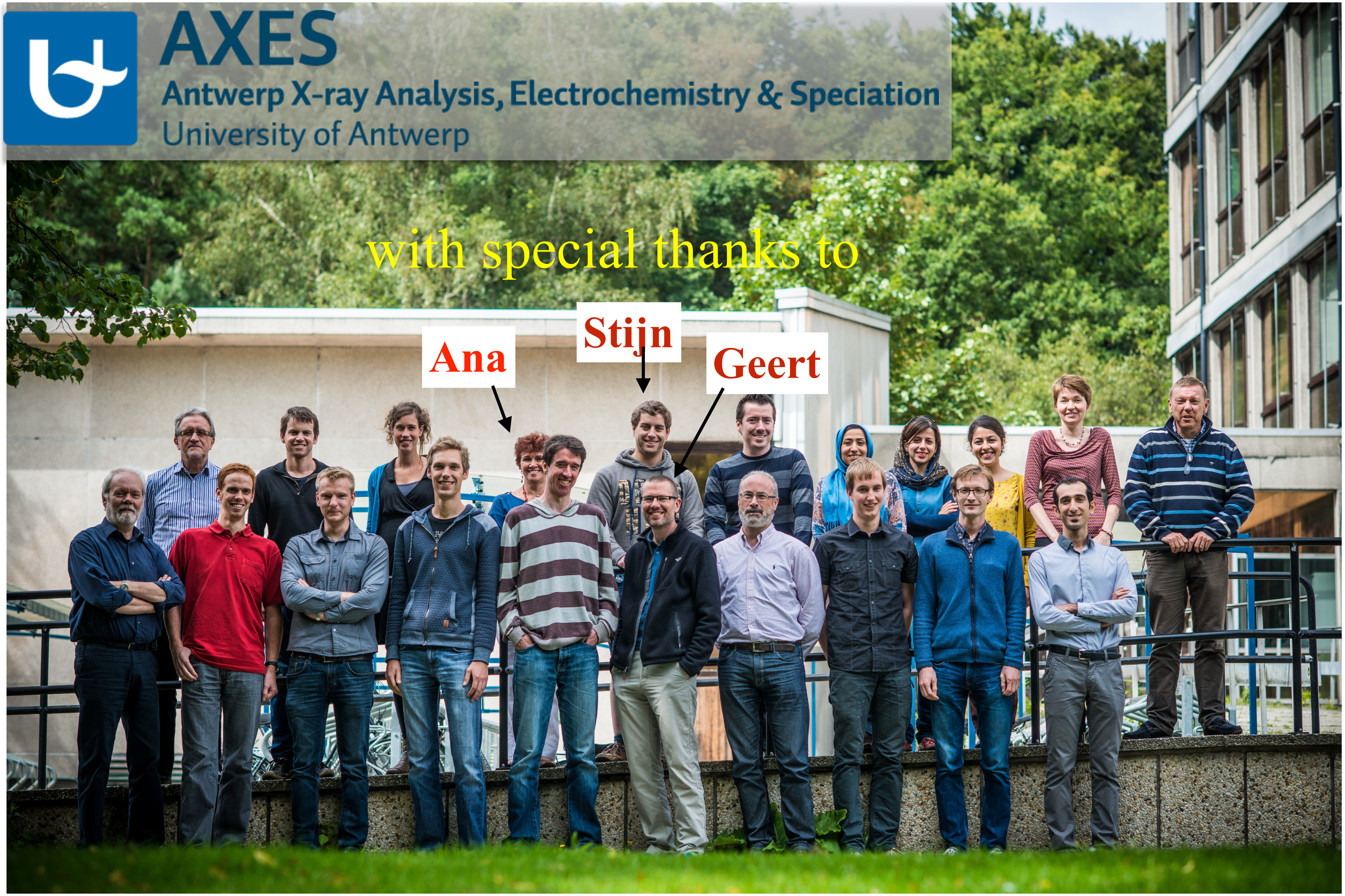
Reduced the data size on disk by a factor of ~ 10

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 tool for the processing of large MA-XRF datasets



Thanks for your attention