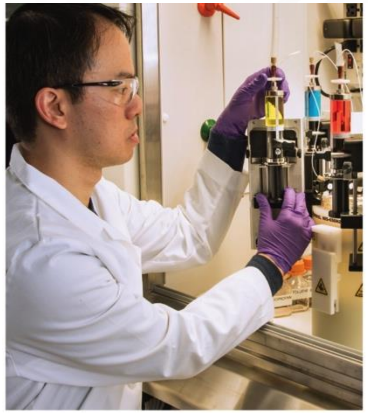
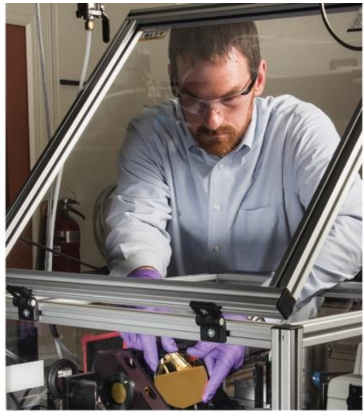




An Open Evaluation of Hyperspectral Unmixing Strategies for EDS Analysis

Joshua A. Taillon

August 8, 2018

M&M 2018 – Baltimore
A11.6.625 – Room 327 – 11:15 AM

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

MATERIAL MEASUREMENT LABORATORY

Disclaimer

Certain commercial equipment, instruments, materials, vendors, and software are identified in this talk for example purposes and to foster understanding. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Second Disclaimer

- **High-level introductory talk**
- **Transfer ideas from signal processing to microscopists**
- **Encouragement to try out new methods and tools**

Outline:

- **What is unmixing (phase mapping)?**
- **Vendor options vs. open solutions**
- **Demonstration of different unmixing algorithms**
- **How to implement on your data?**

What is hyperspectral unmixing?

- Start with some hyperspectral data…

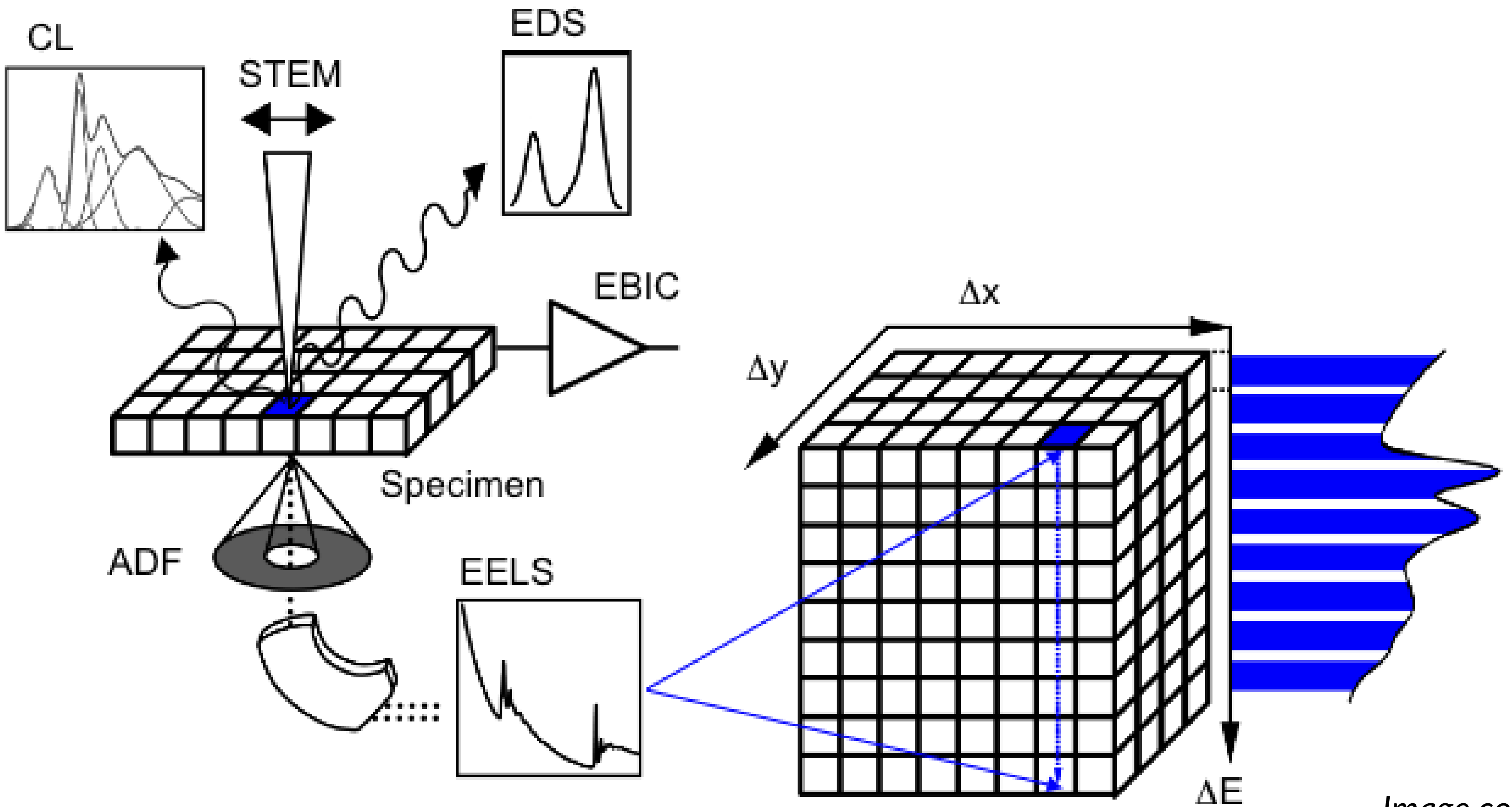
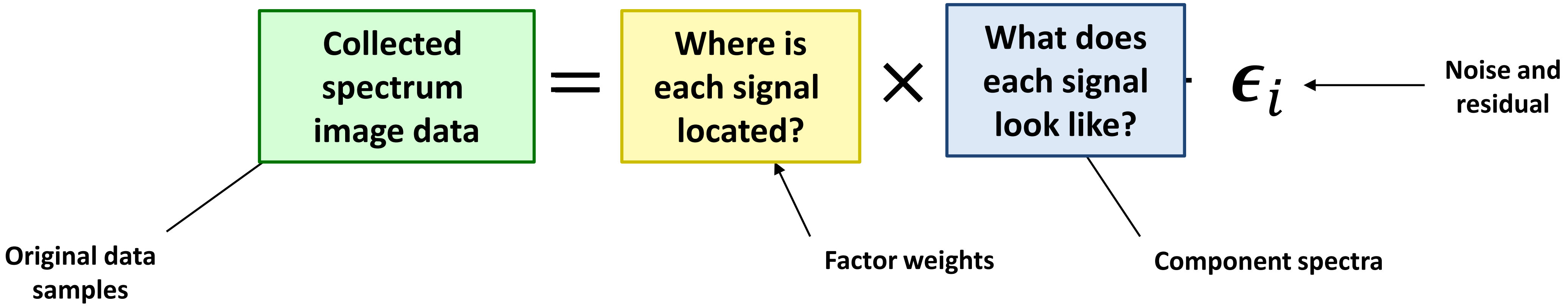
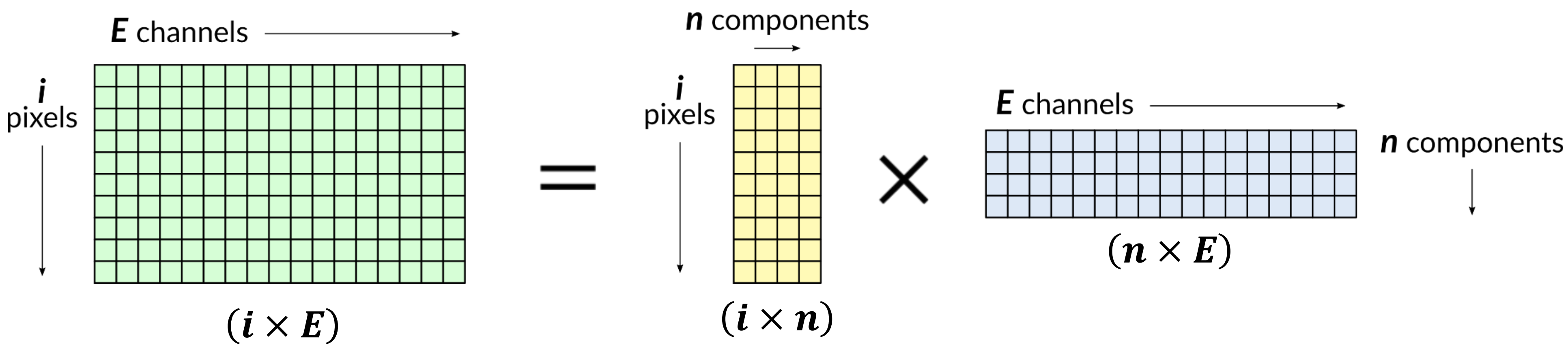


Image courtesy of Gatan, Inc.

What is hyperspectral unmixing?

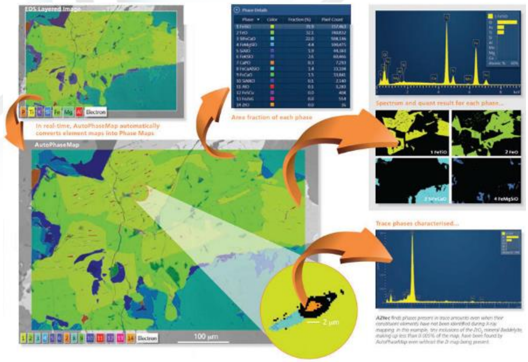


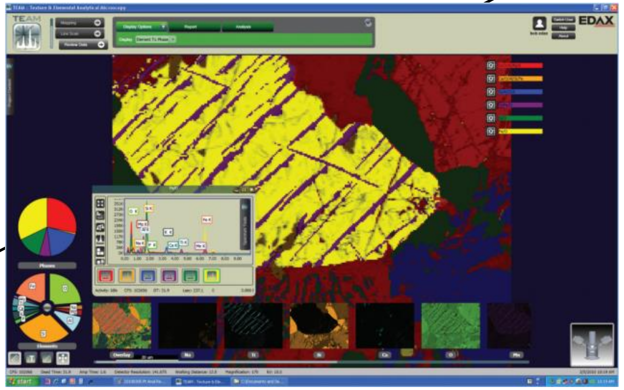


What do the vendors offer?

- **If you use a modern EDS software, you’ve done hyperspectral unmixing…**

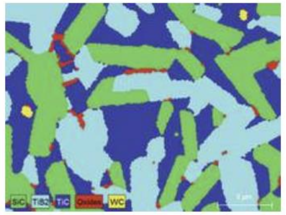
Vendor	Phase Mapping Tool
Oxford	AutoPhaseMap
EDAX	Smart Phase Mapping
Bruker	AutoPhase
Thermo Fisher	COMPASS





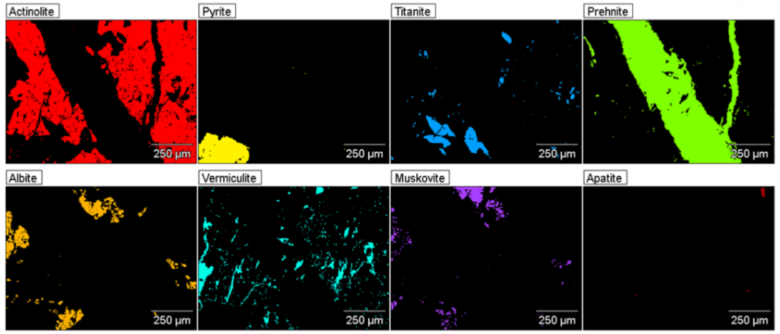
ESPRIT AutoPhase

The Automatic Phase Analysis Tool



- Fully automatic search for phases in maps
- Works with all types of ESPRIT maps
- Adjustable sensitivity and adjustable minimal phase area

Chemical phase map of a hard ceramic material



Strengths/challenges of *vendor* options

The Good

- Simple point-and-click operation
- Tight integration
  - Collection, visualization, reporting, etc.
- Can be run in real-time
- Integration with EBSD data
- Generally "just works"

The Not So Good

- Extremely "black box"
- Reproducibility (!)
  - Configurable options with little understanding of why
- What are the uncertainties?
- Tied to software ($)
- Choice of vendor should not change the scientific result

Strengths/challenges of *open* options

The Good

- No point-and-click
- (Can be) difficult to access raw data from the vendor software
- Generally only post-processing
- Learning curve can be substantial
- Can take a lot more fiddling

The Not So Good

- You know what's happening
- Reproducibility (!)
  - Anyone can recreate your analysis
- Uncertainty can be understood
- Usually free
- Results do not depend on vendor

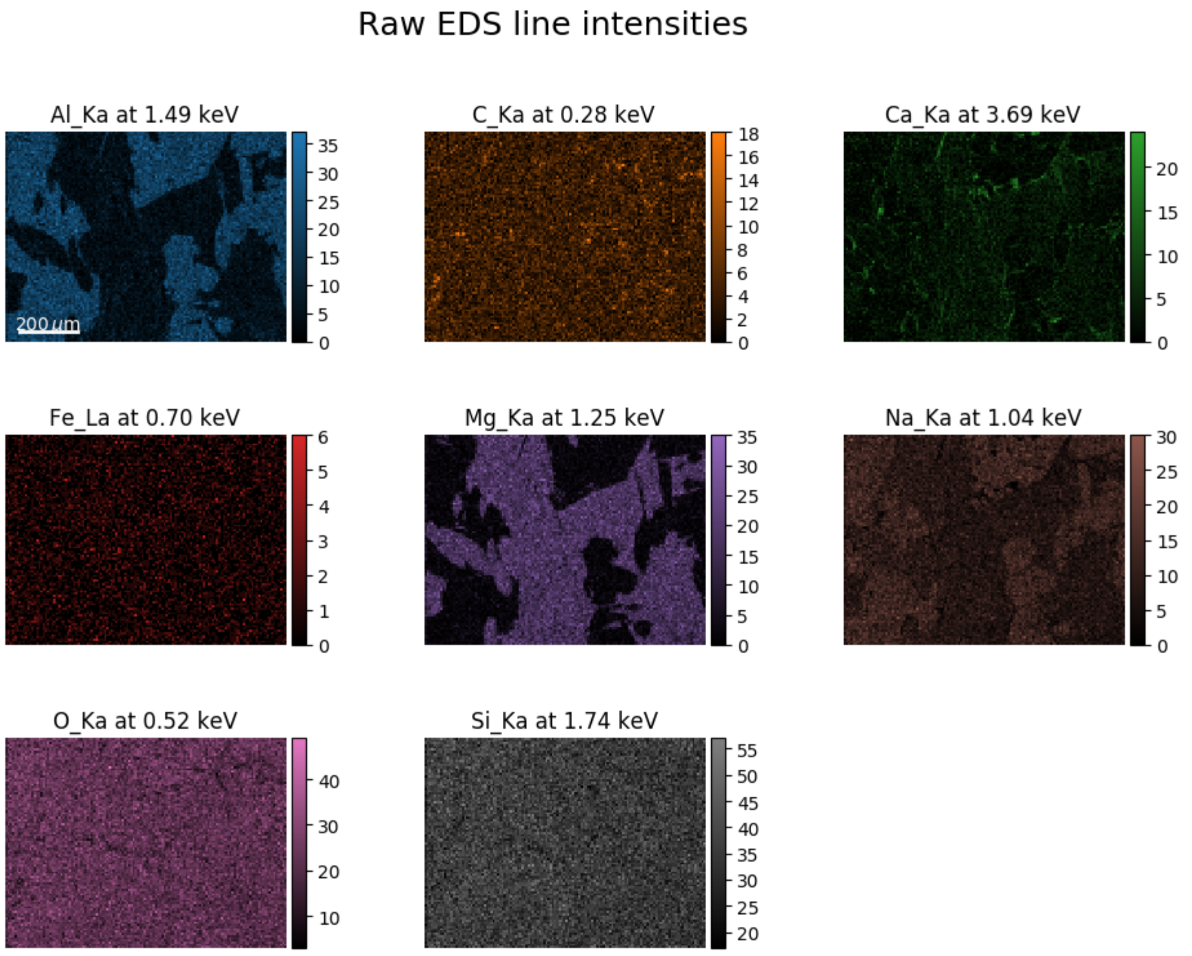
Offline "phase mapping"

- **Many algorithms exist to solve:** $\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y)} \times \boldsymbol{S}_E$
- **Assumptions implicit in each affect their suitability for EDS**

- **Primary methods:**
  - Principal component analysis (PCA) – finds non-physical spectra that describe the most variance in the datacube
  - Independent component analysis (ICA) – maximizes independence between spectral results
  - Multivariate curve resolution (MCR) and non-negative matrix factorization (NMF) – enforce positivity in spectral components and weights

Looking at some real data…

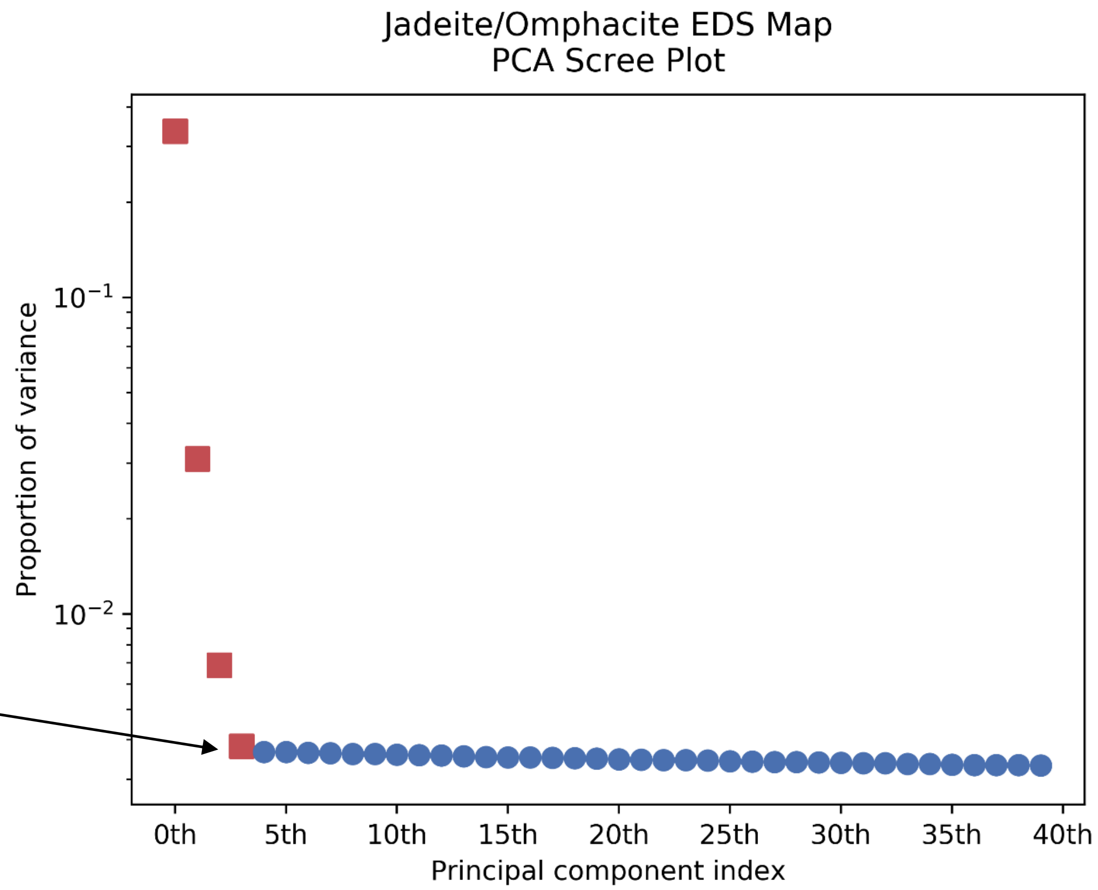
- **SEM-EDS mapping data**
- **Japan 1957 Research Specimen from Freer Gallery of Art**
    - Data from an ongoing research project at Smithsonian's Museum Conservation Institute
    - Courtesy of Thomas Lam and Edward P. Vicenzi
- **Map specifics:**
    - 30 keV primary beam
    - 512 x 384 pixels; 1564 spectral channels
    - Jadeite ($NaAlSi_2O_6$) and Omphacite ($(Ca,Na)(Mg,Fe^{2+},Al)Si_2O_6$)



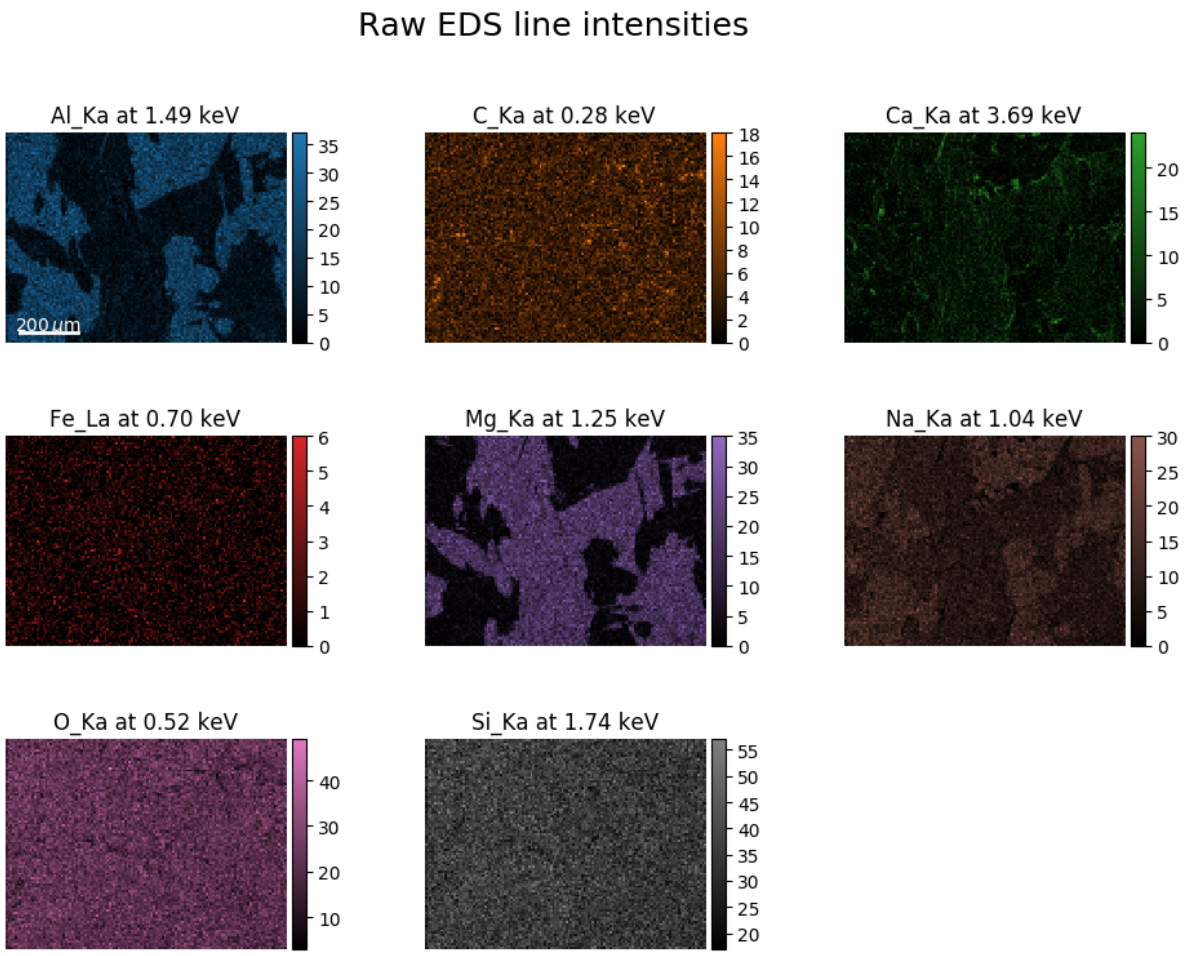
How many components are needed?

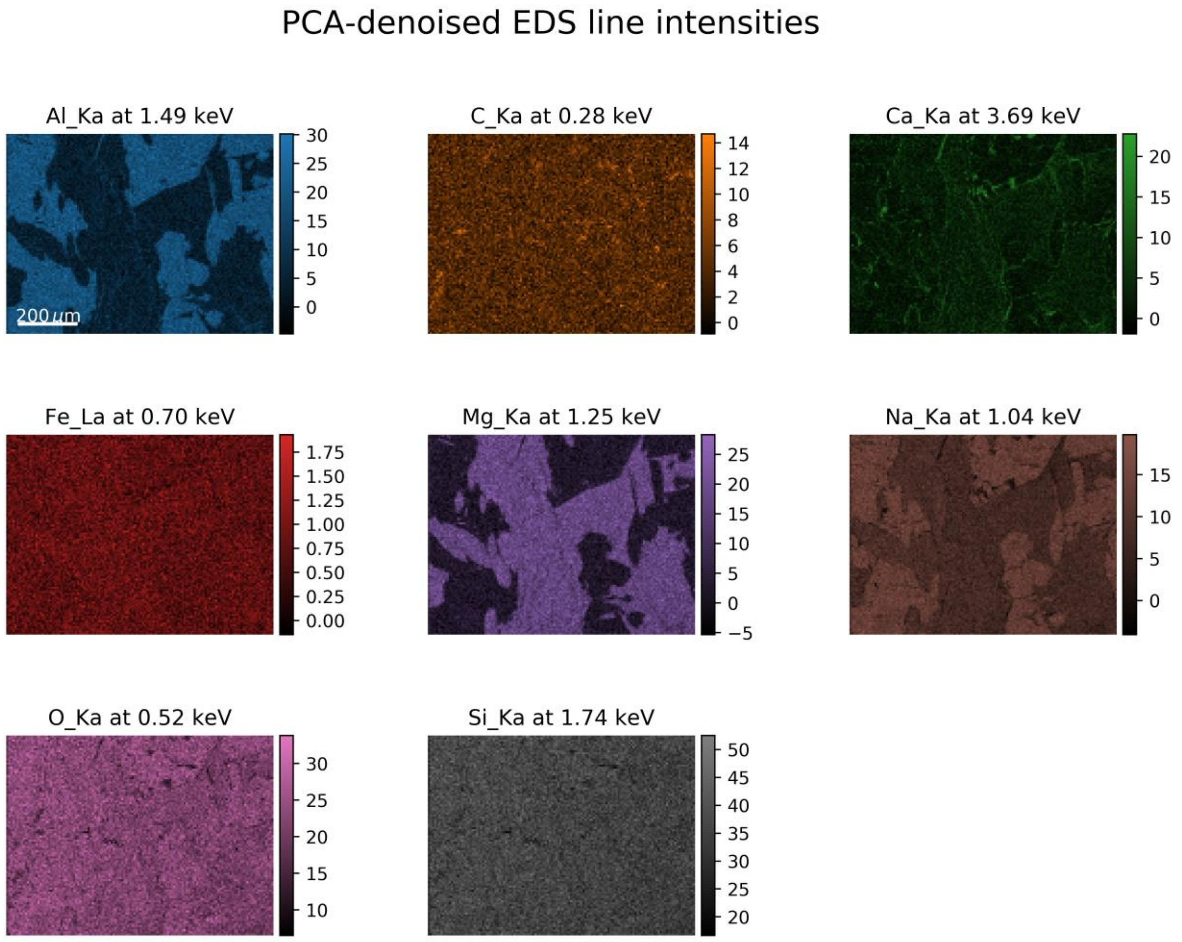
$$\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y),n} \times \mathbf{S}_{(n,E)}$$

- PCA helps determine the answer with a scree plot
- Order the components by decreasing amount of contained variance on logarithmic scale
- "Correct" number of components generally at the discontinuity in the scree plot

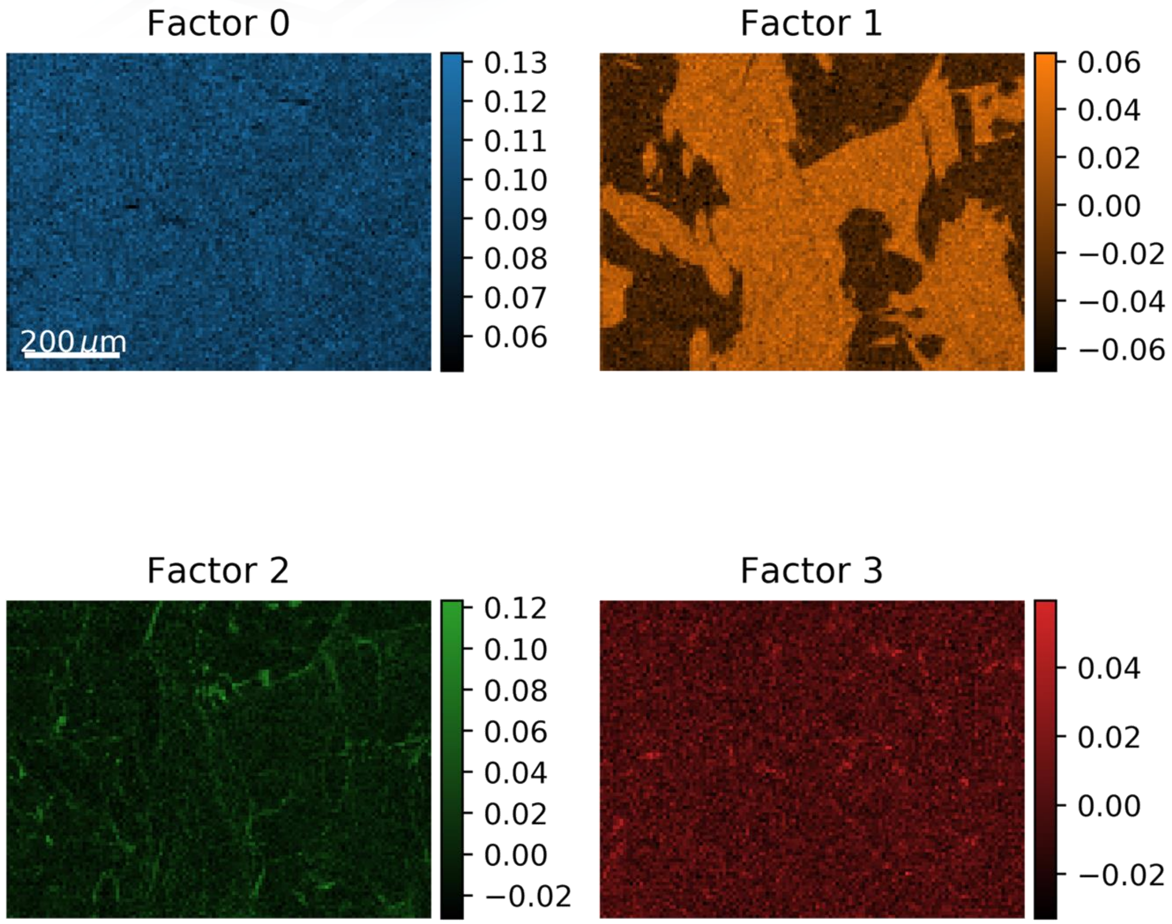


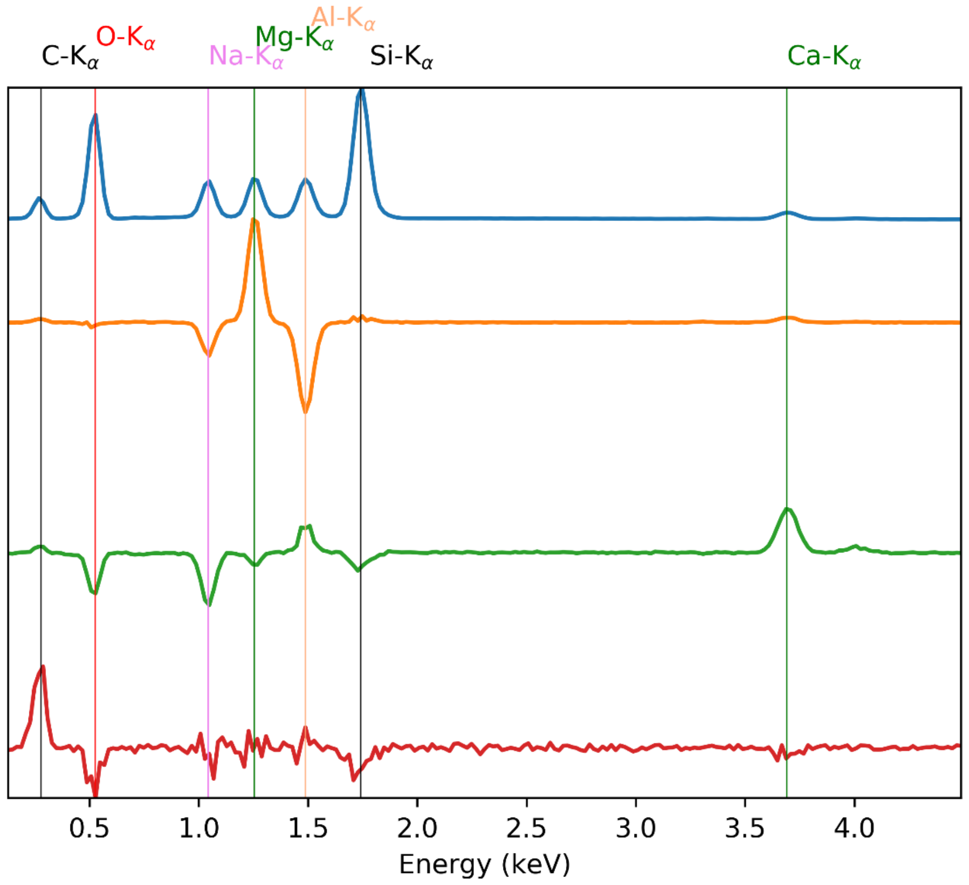
Denoising with PCA



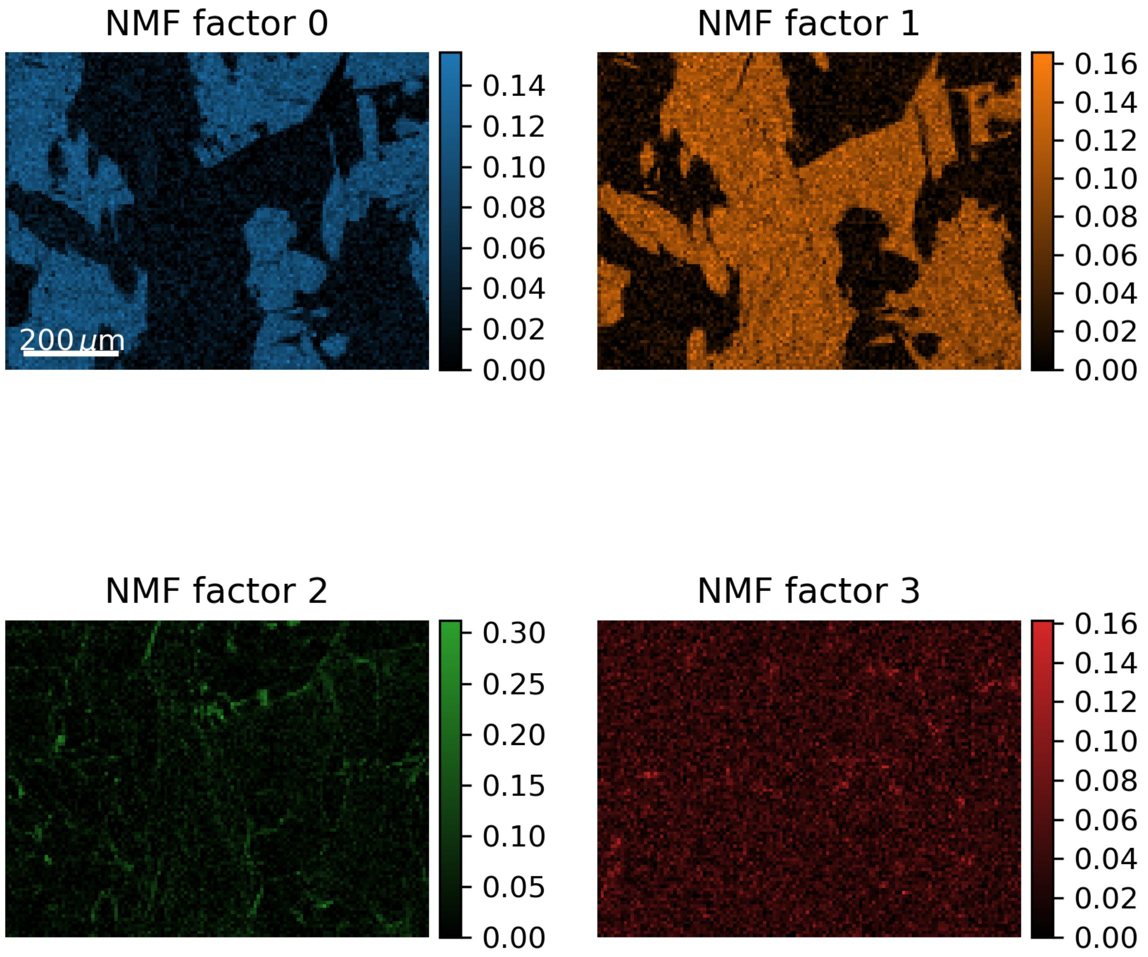


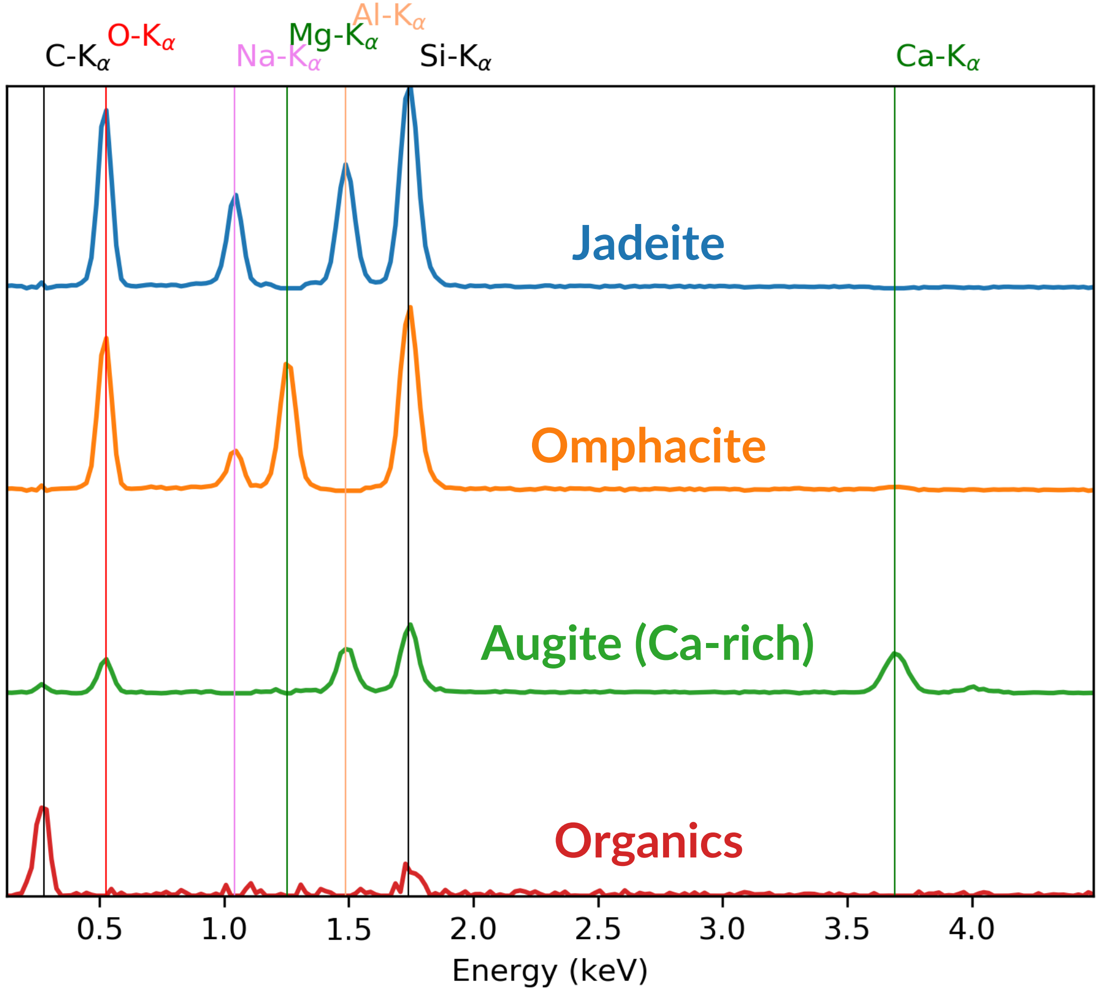
Principal component analysis (*n* = 4)





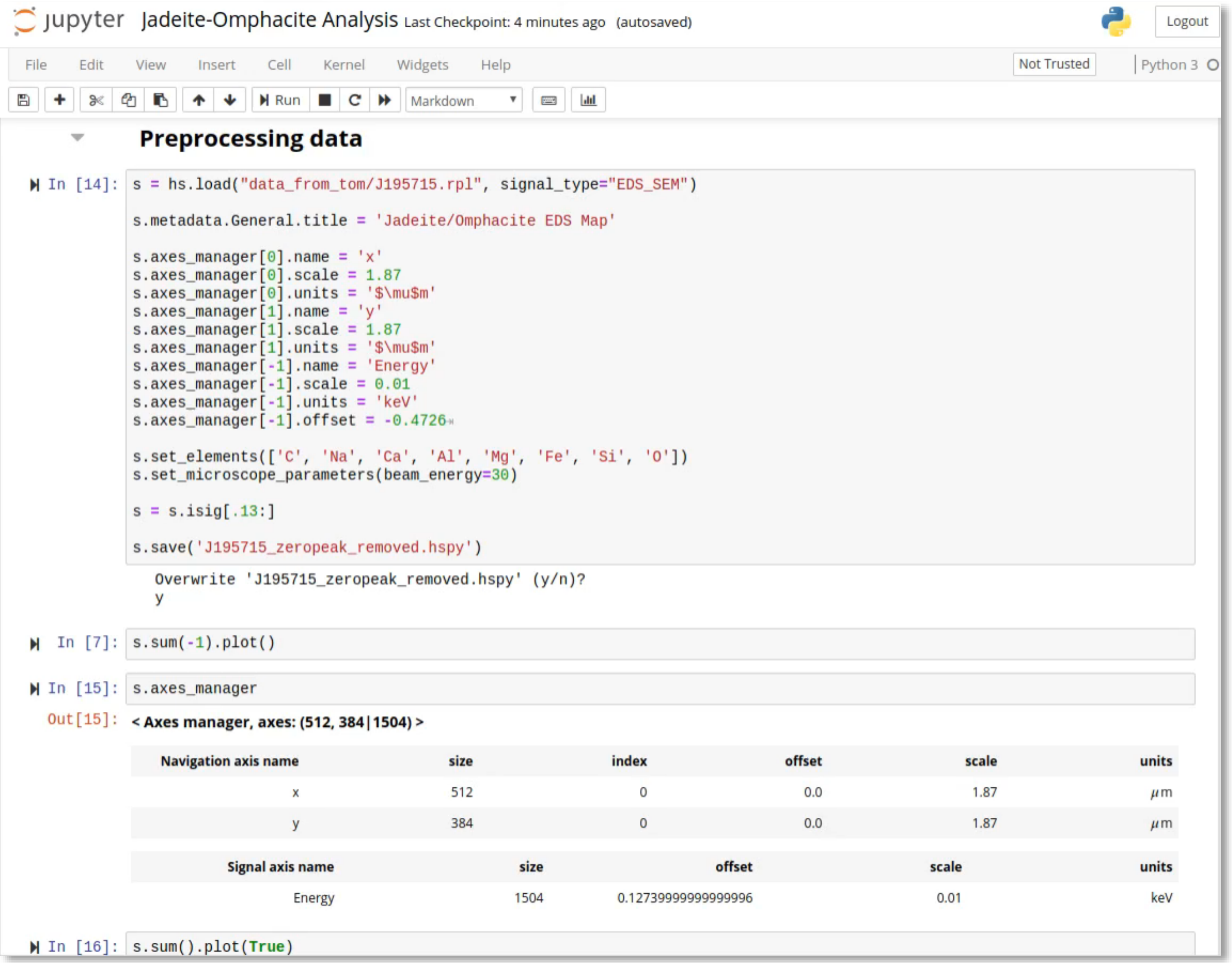
Non-negative Matrix Factorization (NMF)





How can you do this with your data?

- **HyperSpy + Jupyter notebooks!**
  - http://www.hyperspy.org
  - Open-source hyperspectral data processing toolkit based on Python
  - Jupyter notebooks provide interactive, and reproducible data analysis documents
  - Full access to entire scientific Python ecosystem (machine learning, statistics, larger-than-memory processing, etc.)



More reading for the interested:

- **Reviews and microscopy-specific information:**
  - P.M. Voyles, “Informatics and data science in materials microscopy,” *Curr. Opin. Solid State Mater. Sci.* 21 (2017) 141–158.
    - Section 3.1, especially
  - P. Potapov, “Why Principal Component Analysis of STEM spectrum-images results in “abstract”, uninterpretable loadings?,” *Ultramicroscopy*. 160 (2016) 197–212.
  - R. Kannan, *et al.*, “Deep data analysis via physically constrained linear unmixing: universal framework, domain examples, and a community-wide platform,” *Adv. Struct. Chem. Imaging*. 4 (2018) 6.
- **Example applications:**
  - D. Rossouw, *et al.*, “Blind source separation aided characterization of the γ′ strengthening phase in an advanced nickel-based superalloy by spectroscopic 4D electron microscopy,” *Acta Mater.* 107 (2016) 229–238.
  - G. Lucas, P. Burdet, M. Cantoni, C. Hébert, “Multivariate statistical analysis as a tool for the segmentation of 3D spectral data,” *Micron*. 52–53 (2013) 49–56.

Thank you!

Questions/comments?

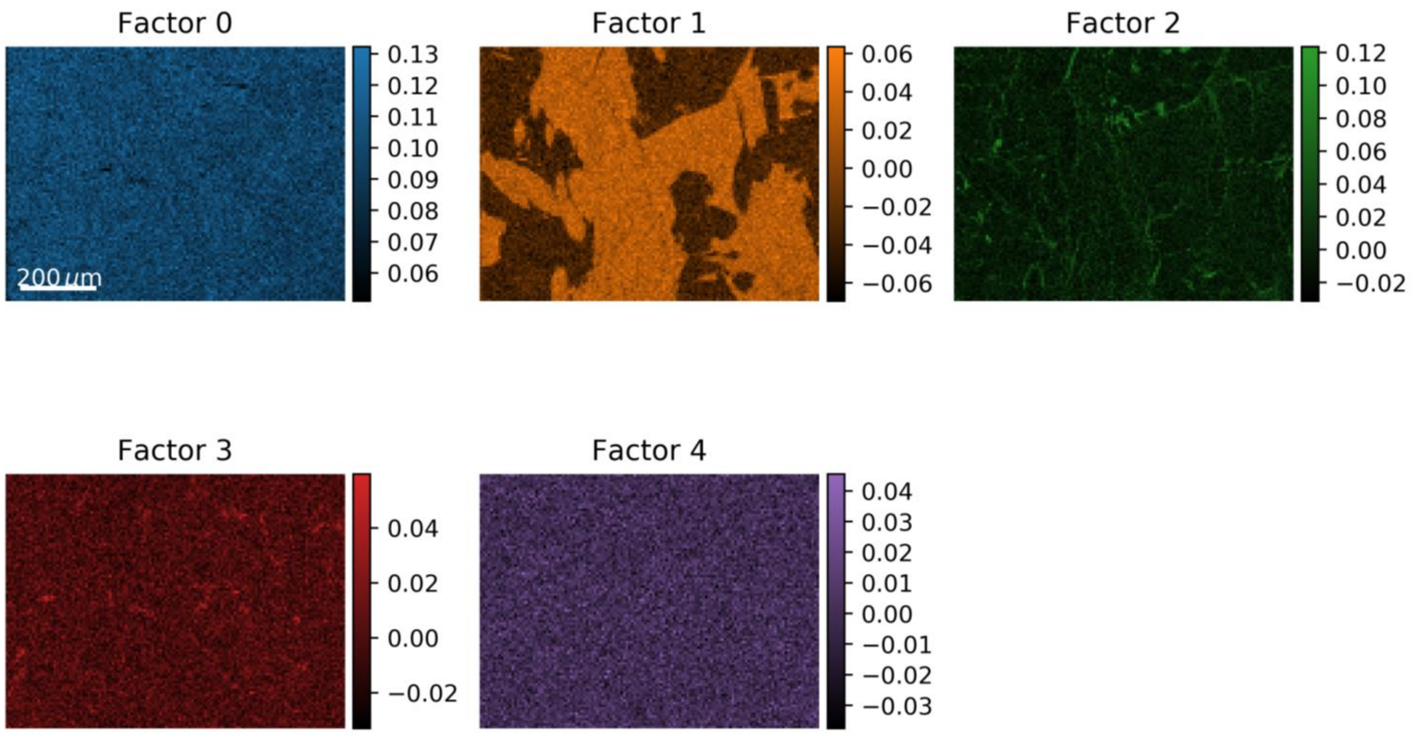
joshua.taillon@nist.gov
(301) 975-2913

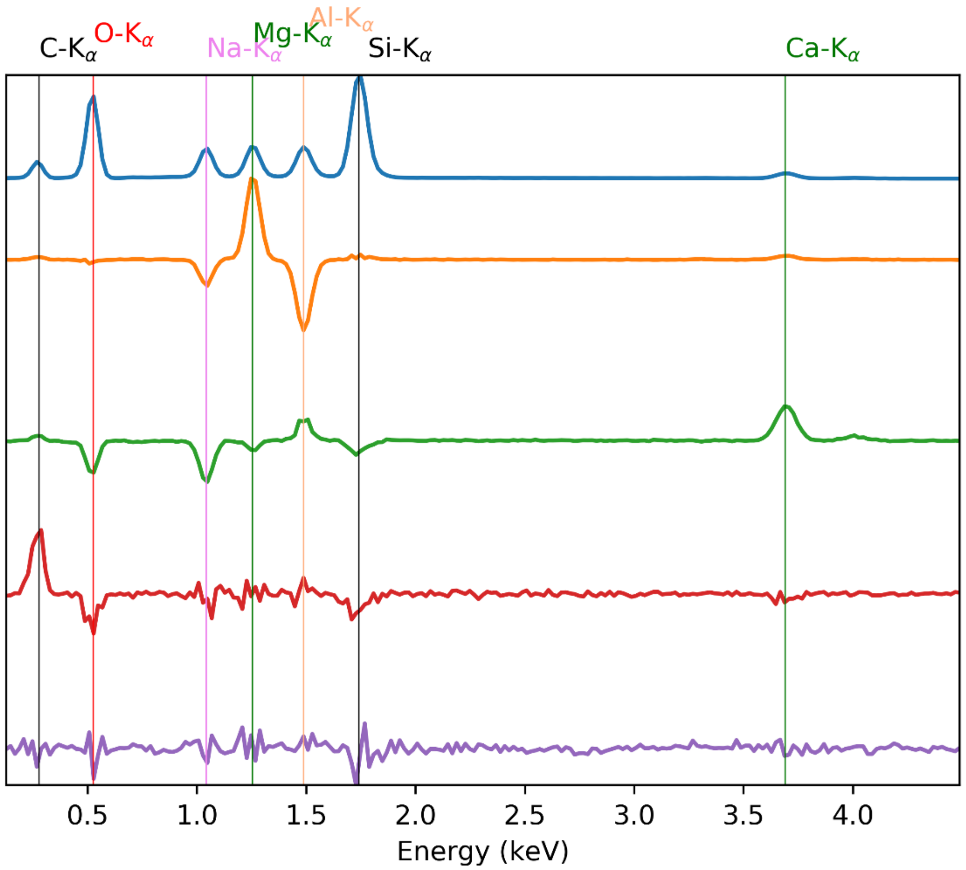
Offline "phase mapping"

- **Many algorithms exist to solve:** $\mathbf{D}_{(x,y),E} = \mathbf{W}_{(x,y)} \times \boldsymbol{S}_E$
- **Assumptions implicit in each affect their suitability for EDS**

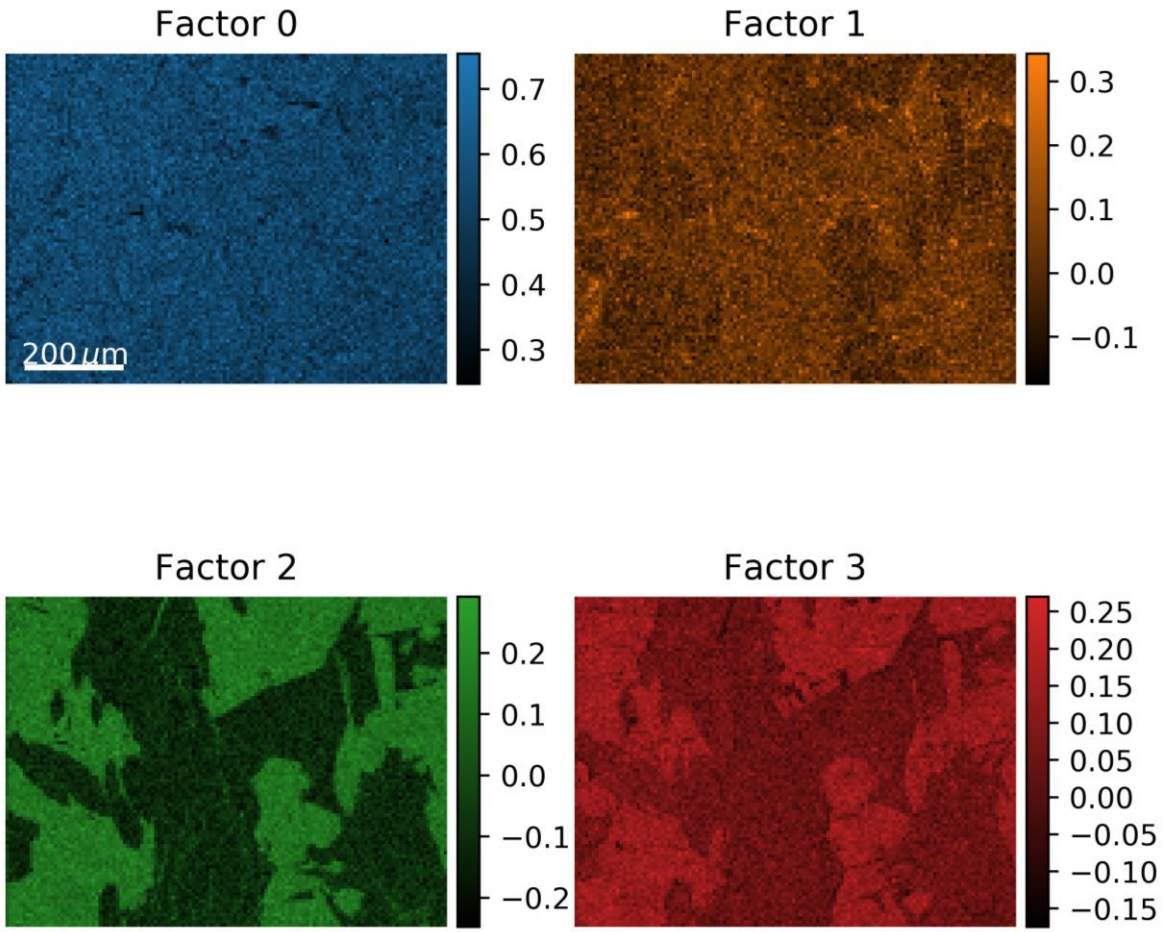
- **Other methods:**
  - Geometric methods – Vertex component analysis (VCA), Minimum volume simplex analysis (MVSA), Simplex identification via split augmented Lagrangian (SISAL), and others…
  - Monte Carlo methods – Bayesian linear unmixing (BLU)
  - Clustering methods – *k*-means, Gaussian mixture modeling (GMM)

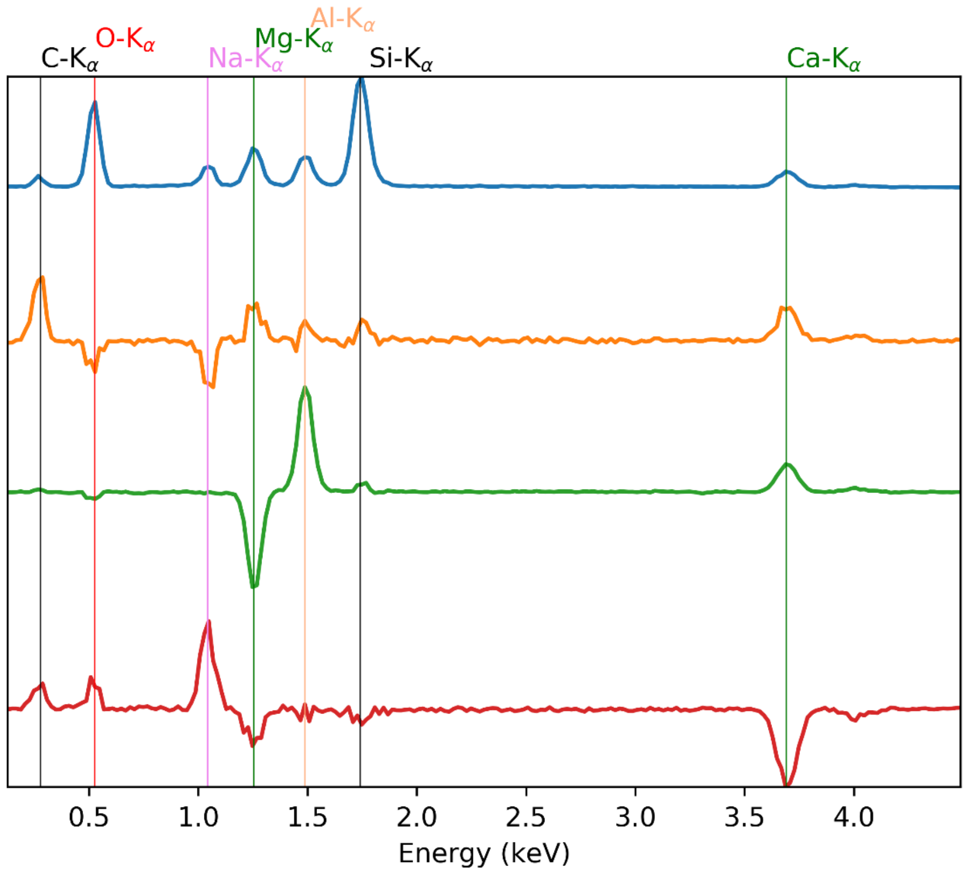
Principal component analysis (n = 5)





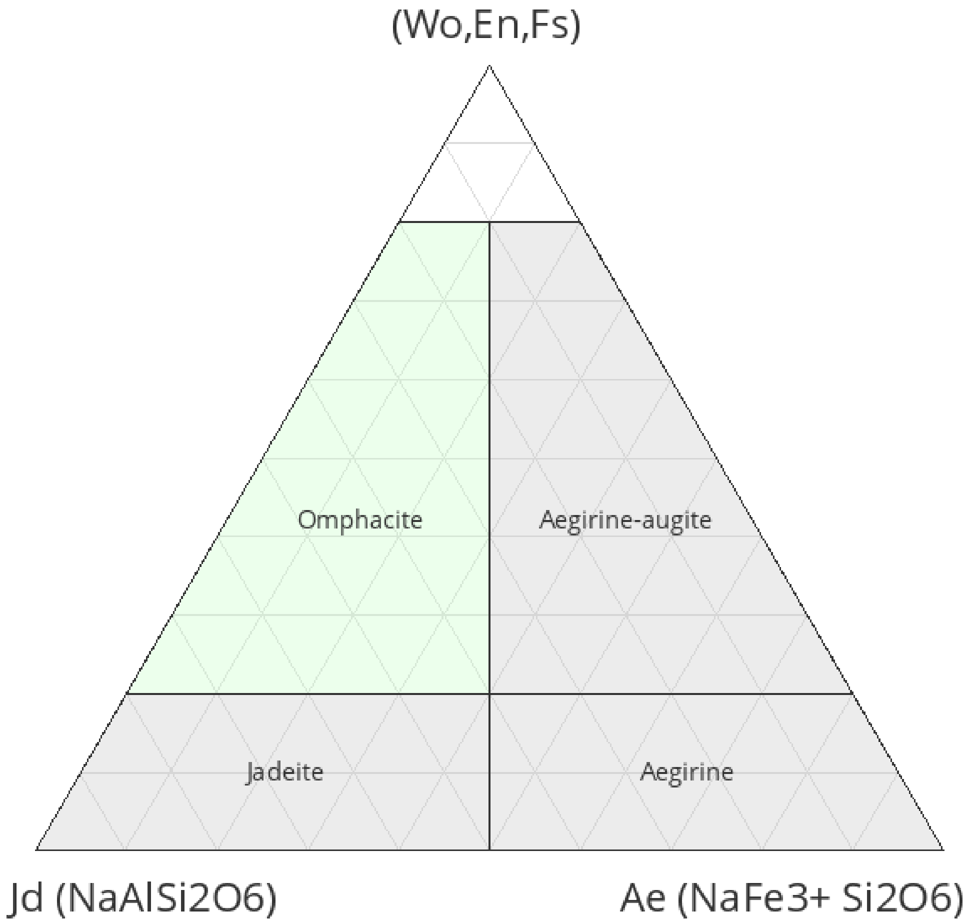
Adding independent component analysis (ICA)





Omphacite - (Ca,Na,Mg)(Mg,Fe^{2+}, Fe^{3+},Al)Si_2O_6

- **Omphacite is solid solution of:**
  - Jadeite – Na(Al,Fe^{3+})Si_2O_6
  - Augite – $(Ca_xMg_yFe_z)(Mg_{y1}Fe_{z1})Si_2O_6$
  - Aegirine – $NaFe^{3+}Si_2O_6$



https://www.mindat.org