Image sequence (stack) analysis with PCA_GUI in aXis2000

Adam Hitchcock (aph@mcmaster.ca)

1) Principles (slides courtesy Chris Jacobsen)
2) Example - core-shell microsphere (Stover) - 50:50 shell
    * forward fitting using SVD & OD1 reference spectra
    * PCA_GUI analysis
    * Comparison

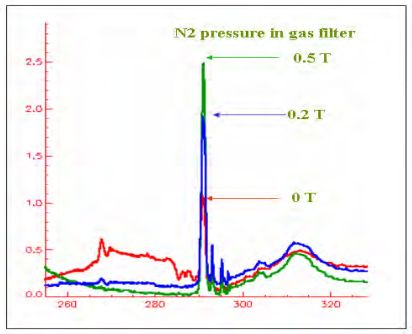
REFERENCES:

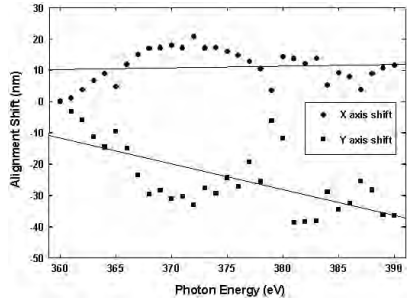
- C. Jacobsen, M. Feser, M. Lerotic, S. Vogt, J. Maser, and T. Schäfer, *Cluster analysis of soft x-ray spectromicroscopy data*, Journal de Physique IV **104**, 623 (2003)
- M. Lerotic, C. Jacobsen, T. Schäfer, and S. Vogt, *Cluster analysis of soft x-ray spectromicroscopy data*, Ultramicroscopy **100**, 35 (2004)
- M. Lerotic, C. Jacobsen, J. B. Gillow, A. J. Francis, S. Wirick, S. Vogt, and J. Maser, *Cluster analysis in soft X-ray spectromicroscopy: Finding the patterns in complex specimens*, J. El. Spec. Rel. Phen. 144, 1137 (2005)
- H. Ade & A.P. Hitchcock, , *NEXAFS microscopy and resonant scattering: Composition and orientation probed in real and reciprocal space*, Polymer 49, 643-675 (2008) - describes this example

Some challenges of quantitative mapping

1. Quality of data (spectra, linescan, **image sequence**)

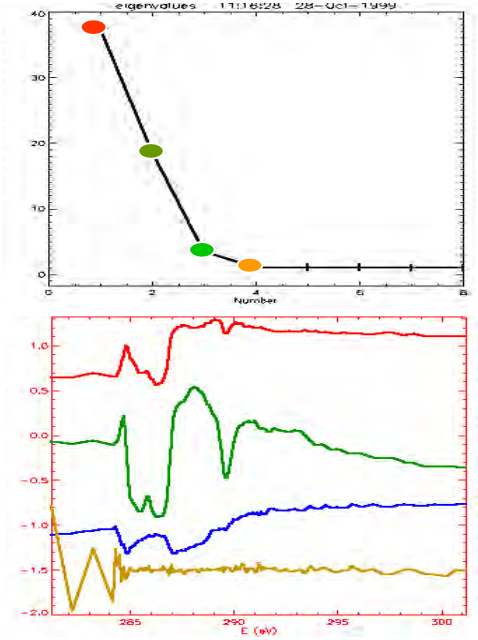
Issue	Recommendation
linearity of absorption	avoid saturation (OD < 2-3)
spectral distortion	reduce or eliminate second order
energy calibration	check regularly; calibrate
alignment	interferometry; careful alignment
radiation damage	use as small a dose as possible
linear E-scale	check with known spectra





2. Suitability & quality of reference spectra

Situation	Recommendation
chemistry well known	record spectra of same or similar pure material (eg monomer unit)
chemistry poorly known but spatially well isolated	internal models guided by external models
chemistry complex	Principle component analysis to place limits on number of components
chemistry unknown	MSA and cluster analysis Lerotic & Jacobsen, J. El. Spec. 2005, 144, 1137
chemistry unknown	internal models by trial and error



Slide courtesy of
Chris Jacobsen
SUNY Stony Brook

Rotating principal components to obtain real spectra using cluster analysis

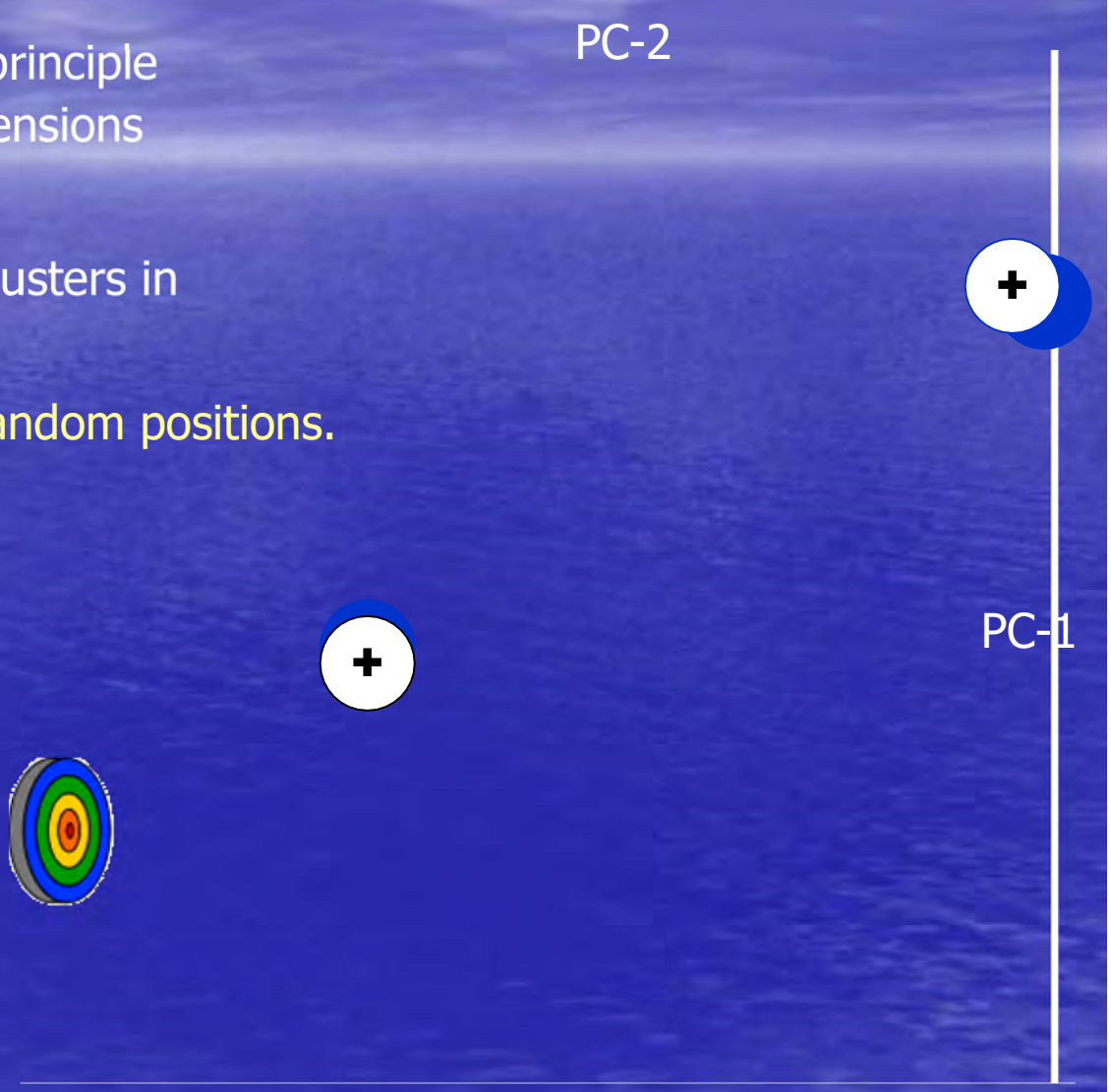
Problem: principal components are abstract. They are mixtures of the actual real spectra of the compounds present.

- PCA provides an orthogonalized representation of the data with less noise, fewer coordinates. Working with only significant components is an effective noise filter.
- We can find groupings of the data in the principal component coordinate system which relate to individual chemical components
- How ? => Cluster analysis or pattern matching

Slide courtesy of
Chris Jacobsen
SUNY Stony Brook

Cluster analysis: Euclidian distance learning algorithm

- Data are in multidimensional principle component space; only 2 dimensions shown here.
- Ideally data are arranged in clusters in this space!
- Put down cluster centers at random positions.



Slide courtesy of
Chris Jacobsen
SUNY Stony Brook

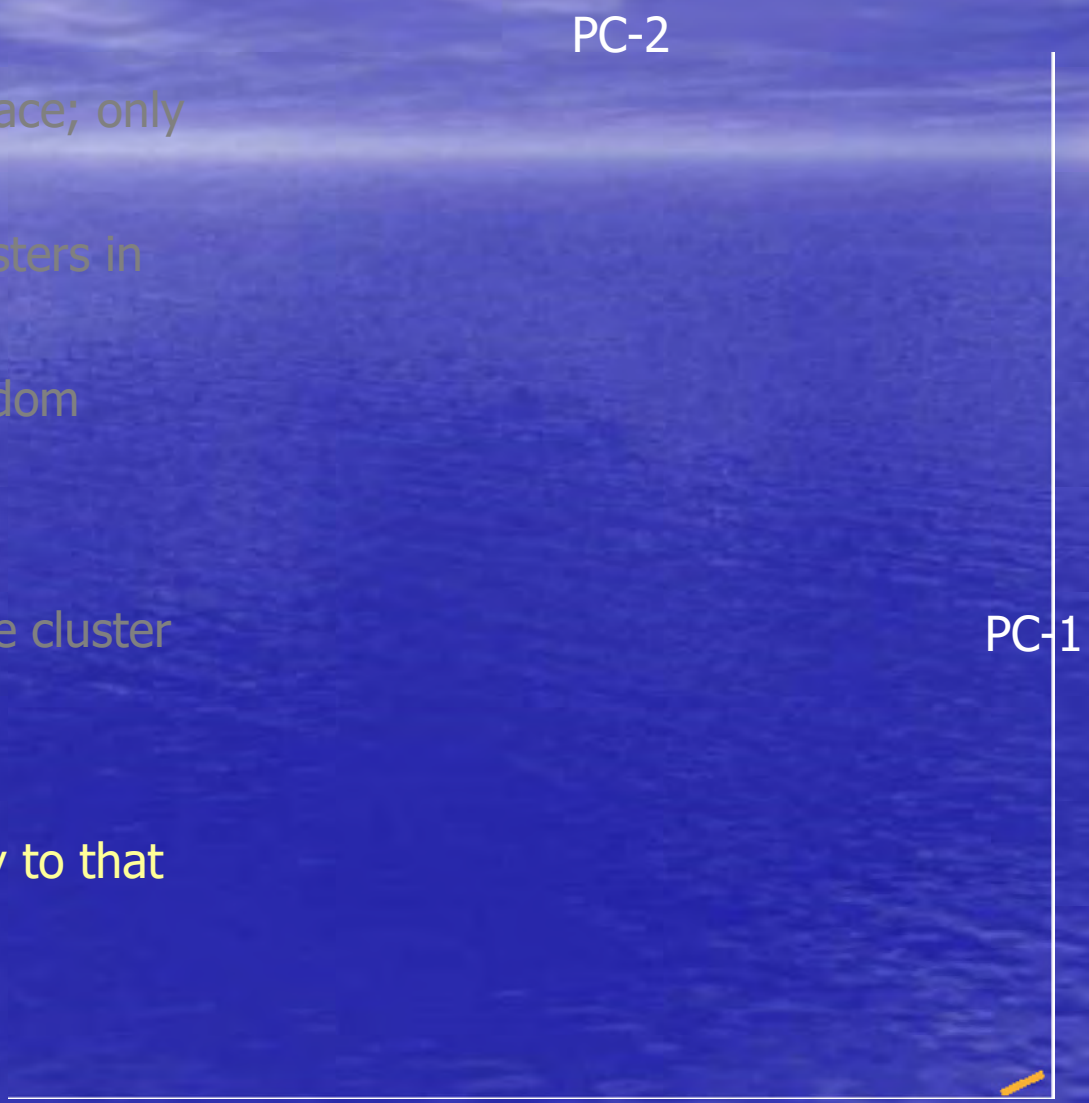
Cluster analysis: Euclidian distance learning algorithm

- Data are in multidimensional space; only 2 dimensions shown here.
- Ideally data are arranged in clusters in this space!
- Put down cluster centers at random positions.
- Iterate:
  - Calculate distances from one cluster center to all data points.
  - Pick shortest distance.

Slide courtesy of
Chris Jacobsen
SUNY Stony Brook

Cluster analysis: Euclidian distance learning algorithm

- Data are in multidimensional space; only 2 dimensions shown here.
- Ideally data are arranged in clusters in this space!
- Put down cluster centers at random positions.
- Iterate:
  - Calculate distances from one cluster center to all data points.
  - Pick shortest distance.
  - Move cluster center partway to that nearest point.



Slide courtesy of
Chris Jacobsen
SUNY Stony Brook

Cluster analysis: Euclidian distance learning algorithm

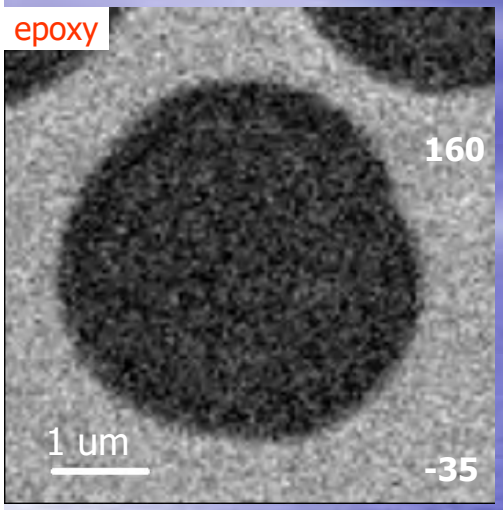
PC-2

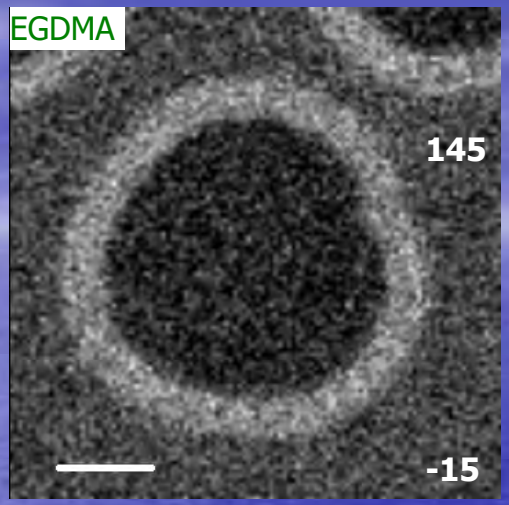
- Data are in multidimensional space; only 2 dimensions shown here.
- Ideally data are arranged in clusters in this space!
- Put down cluster centers at random positions.
- Iterate:
  - Calculate distances from one cluster center to all data points.
  - Pick shortest distance.
  - Move cluster center partway to that nearest point.

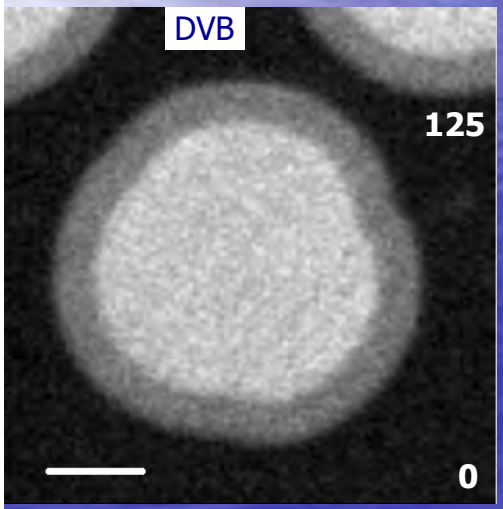
PC-1

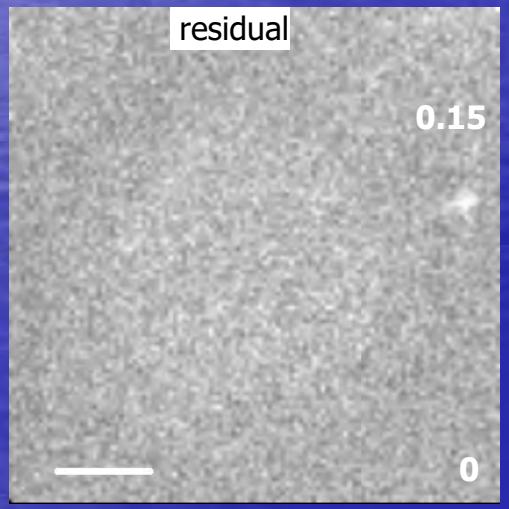
Example: core-shell microspheres

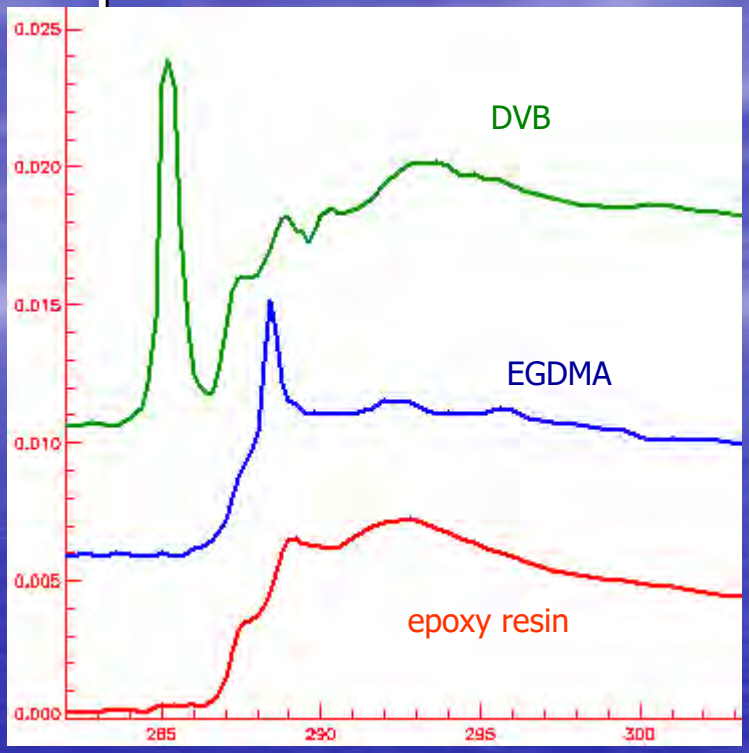
Datafile: cs50c-od.ncd/dat (stxm5322: 2003_09_28_042)

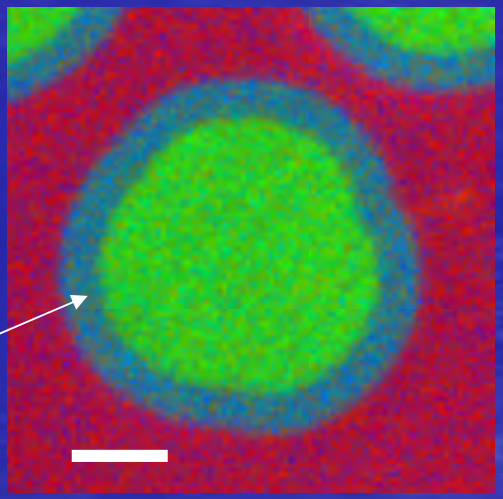










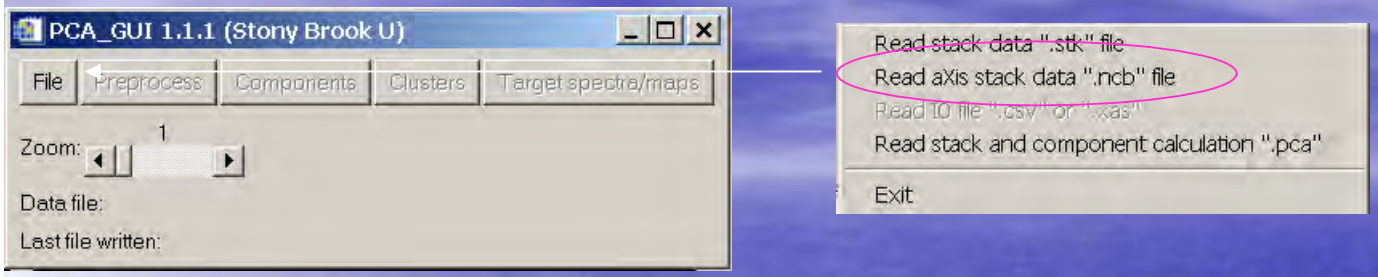


Turquoise color (B&G) indicates the shell is a MIX of DVB and EGDMA

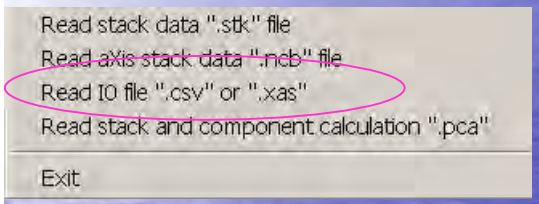
Getting the stack into PCA_GUI

0. LAUNCH stacks~statistical analysis~PCA_GUI (CJJ Dec 2005)

1. Read in the TRANSMITTANCE file - **cs50-c-1.ncb**/dat



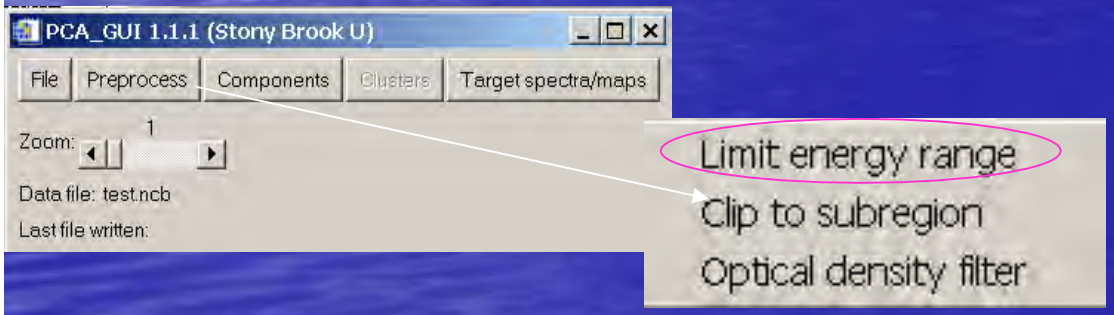
2. then, read in the Io, as the *.xas file:

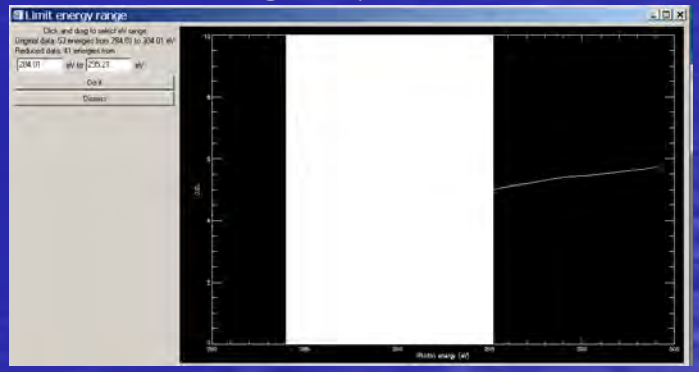


If you have only the stack in OD format, there is a commend on the stacK_process GUI (accessed by stacks~analyze~aXis_binary) which converts OD stacks to transmittance stakcs, with a whitel (I(E)=1.0 for all E) Io file

ONLY if you have both (I) and (Io) components successfully read in, do the next menu items become available

3. You can use **PreProcess** to reduce the ENERGY, SPATIAL, or INTENSITY range if you wish

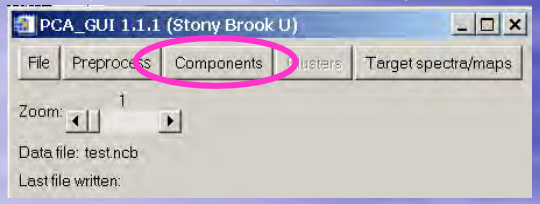


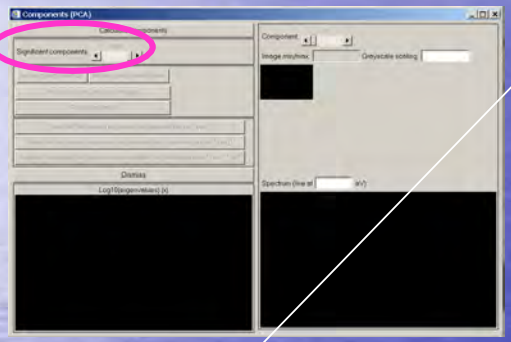


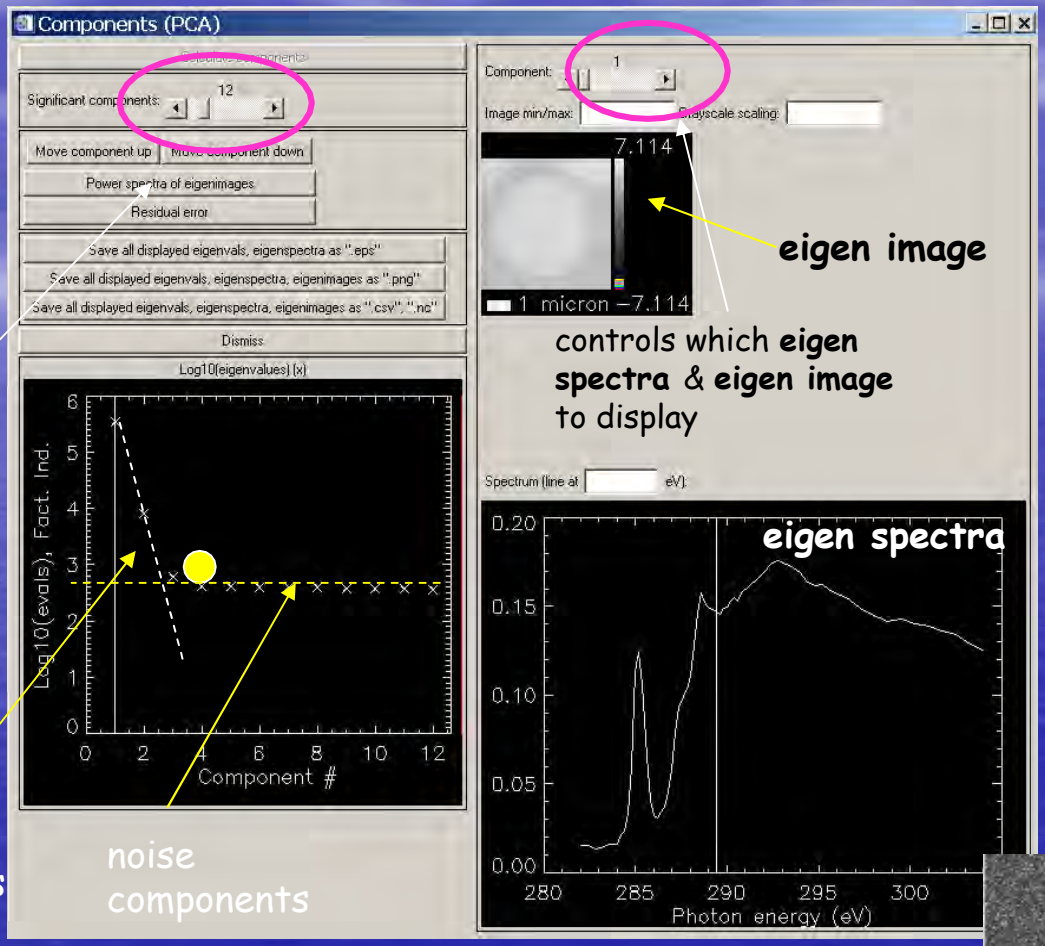
in this case I chose to keep the full data range of the recorded data It may be worthwhile to only include the energy range where the spectra of the chemical species differs the most, and to exclude any regions of noise, especially systematic noise

Using PCA_GUI – generating principle components

4. compute principle components







controls range of eigenvalues to display

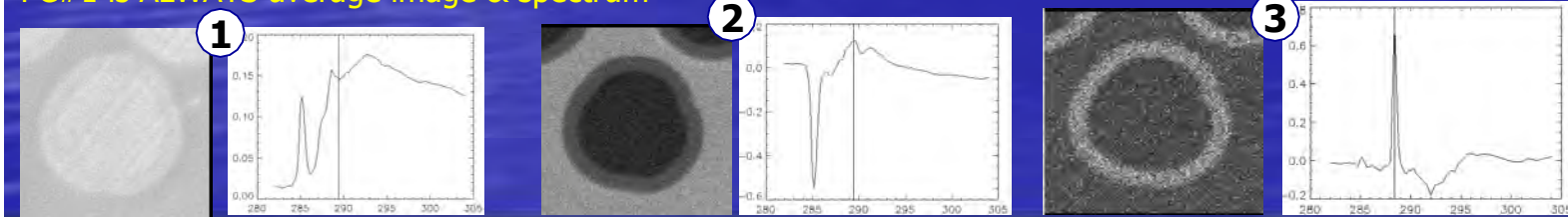
significant components

noise components

In this case components 1 -3 are definitely significant - but mix together all 3 components

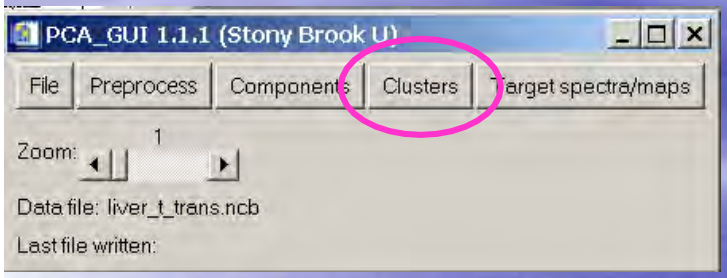
components 4-64 are noise; searching to component 15 just shows no pattern in the eigen-image, as for PC#4 If you regenerate the stack from PC#1, PC#2, PC3 a very clean dat a set can be produced - this is a type of noise filtering

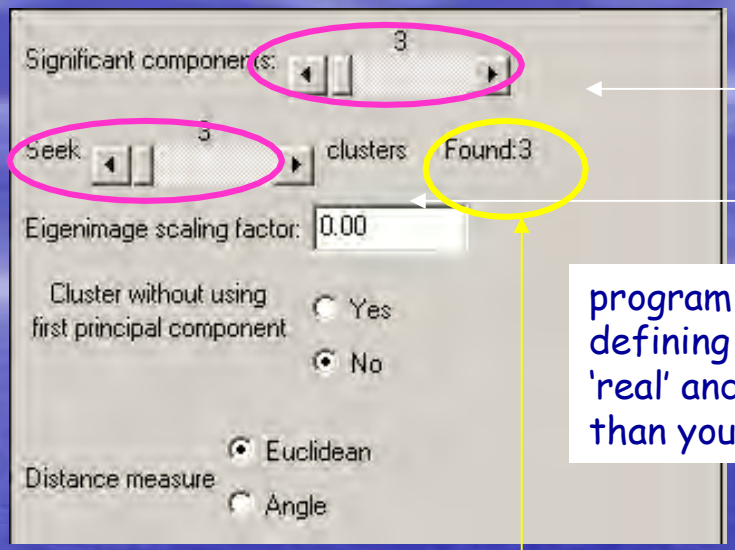
PC#1 is ALWAYS average image & spectrum



Using PCA_GUI cluster analysis to rotate PC

4. compute cluster components





I choose 3 as they are the only ones with meaningful signal

program has a way of defining which clusters are 'real' and often finds fewer than you ask for

this choice FINDS 3 components

This is where the 'art' comes in you have to select

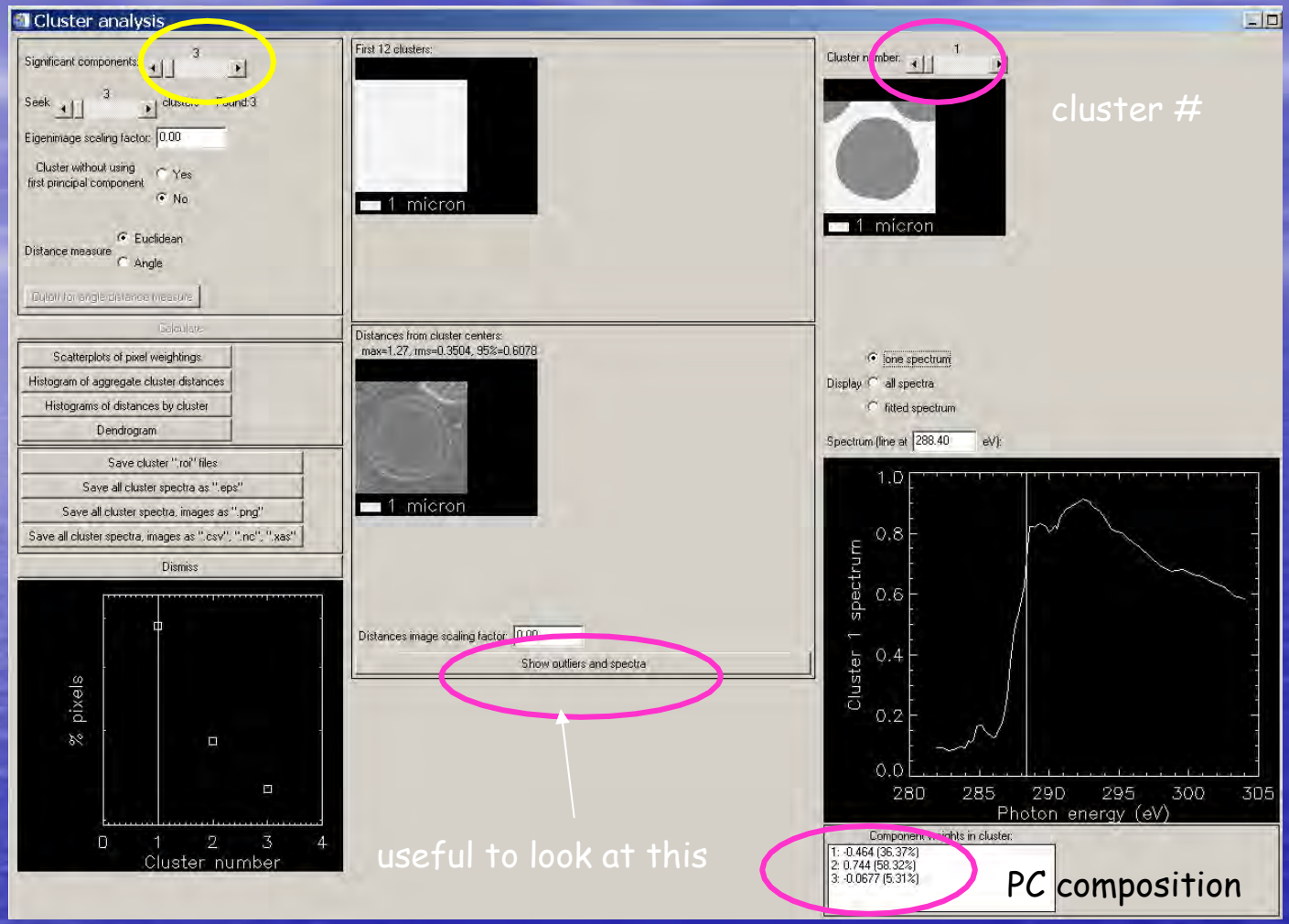
* # of components to include
* WHICH components to include
  (you can move components up or down in the PC# listing in the 'components' panel)
* how many clusters to seek (= # of chemical components)
* method (Euclidian (r) or Angle)
* parameters of each method

All these choices affect the final result to some extent, which requires either experience, or some playing around to be happy that the set of parameters you choose is 'appropriate' for the data set. This is one of the 'Achilles heels' of the unsupervised approach - it needs supervision !!

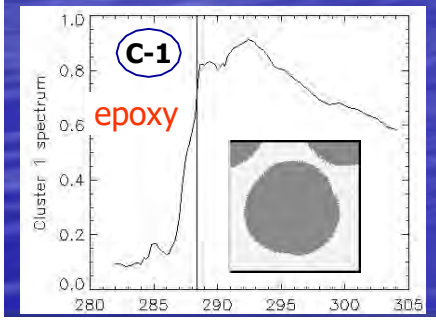
I tend to use Euclidean, but Chris and his student Lerotic, have argued persuasively that angle is better

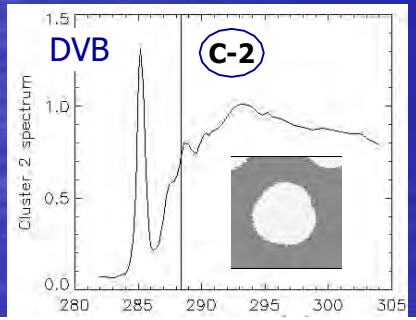
Using PCA_GUI cluster analysis to rotate PC

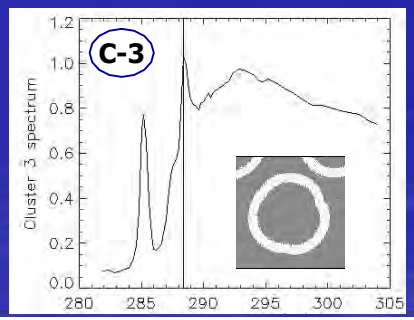
5. examine results



useful to look at this



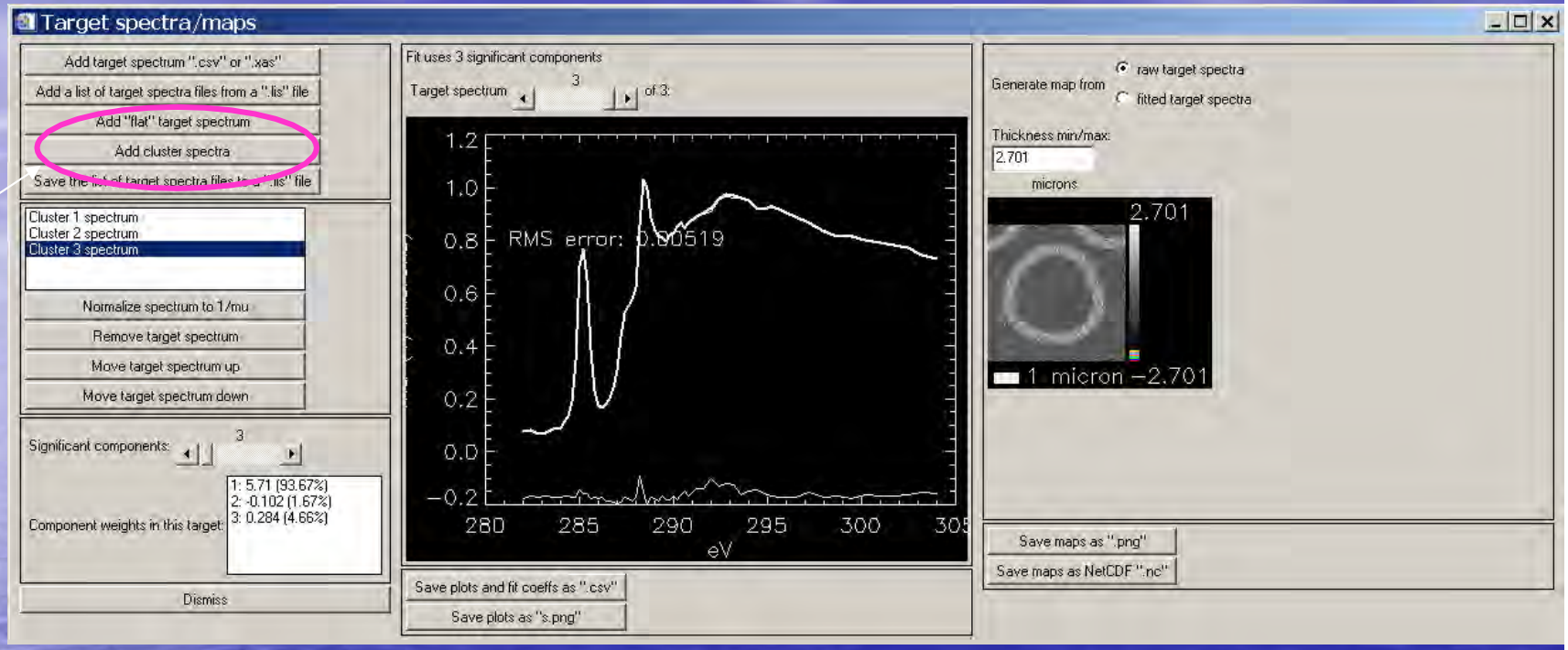




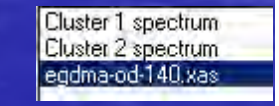
Cluster 3 is the SHELL - which is spatially a well-defined component, but we know from the synthesis and comparison to pure-materials reference spectra that it is a mix of DVB and EGDMA

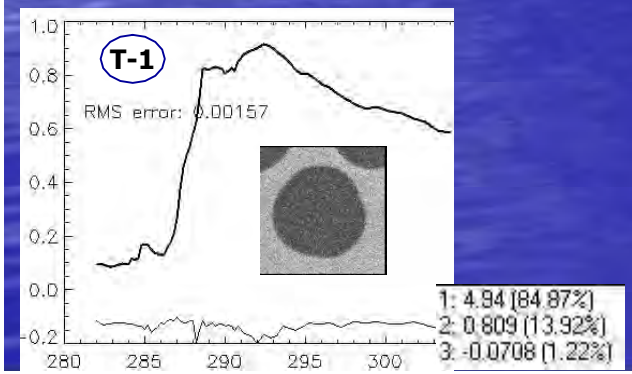
Refining the analysis: target analysis

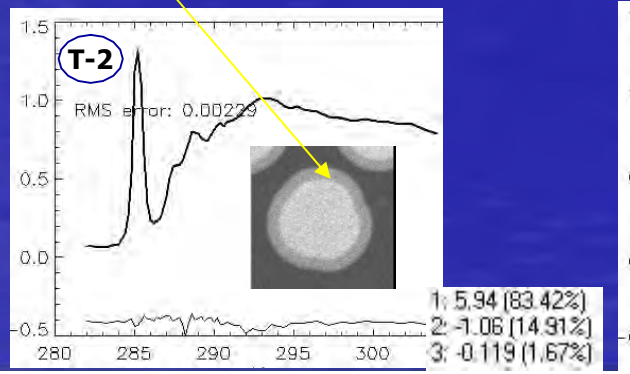
Auto adds the set of cluster spectra found from previous step

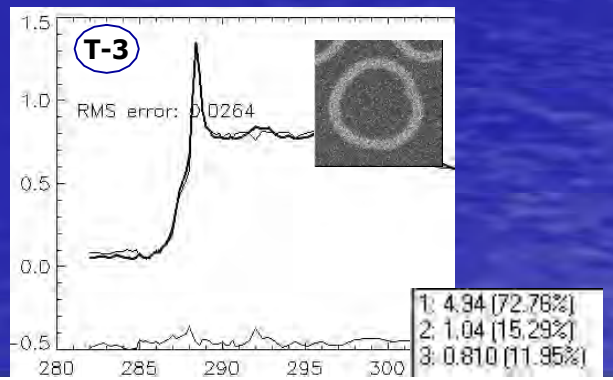


If you remove cluster#3 spectrum and add the spectrum of EGDMA (egdma-od-140.xas) then the analysis gets much better - it only finds EGDMA in the shell, and now finds **DVB** in the shell as well



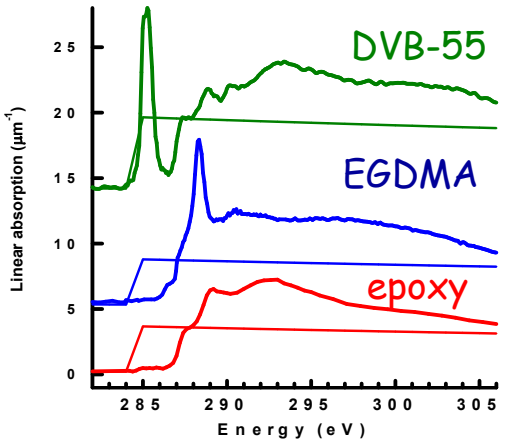


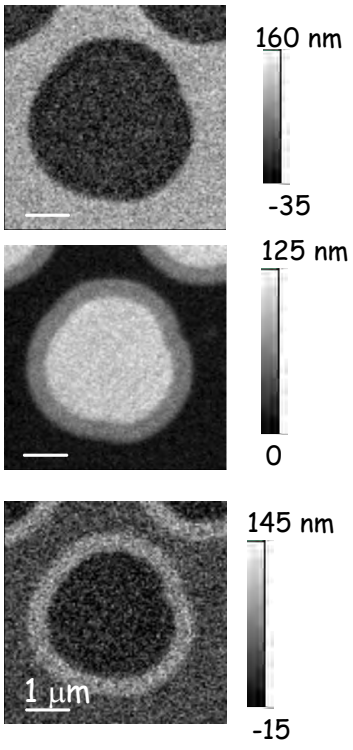


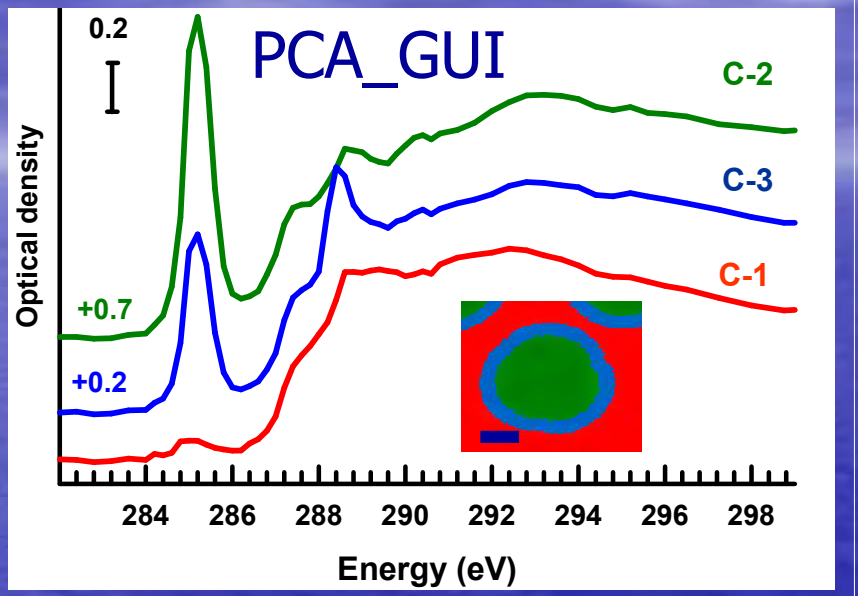


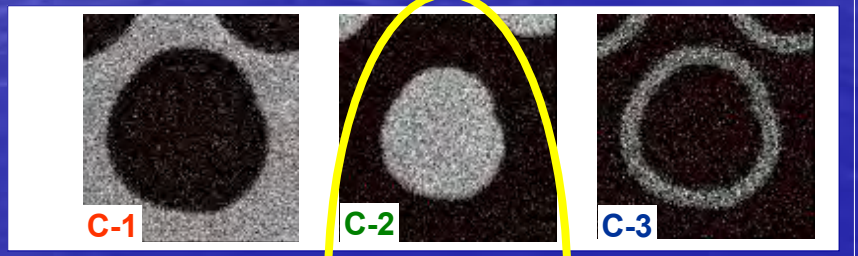
Comparison of SVD and PCA_GUI

SVD









introduce EGDMA as a target spectrum