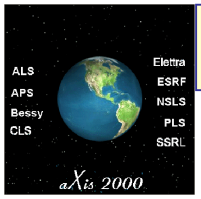
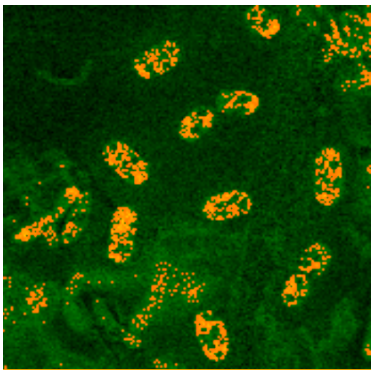
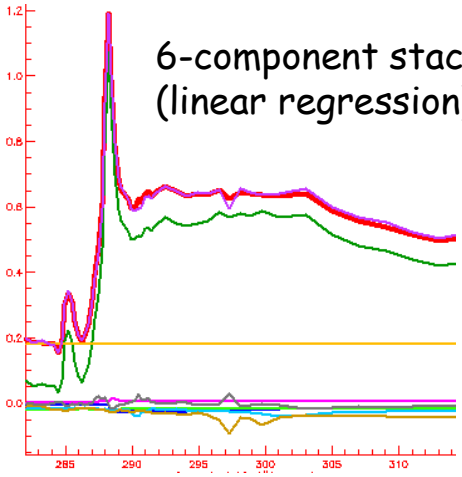


Spectral curve fitting

Users are strongly advised to CHECK stack mapping results by examining the quality of the fit in spectral space.



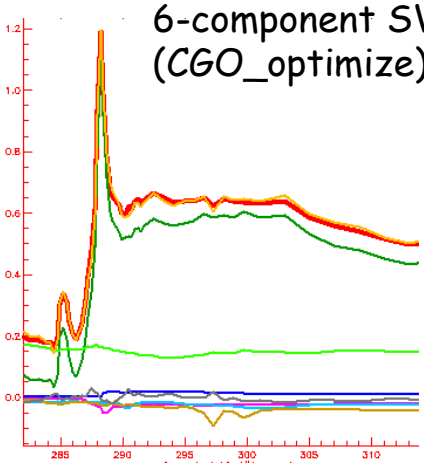
1. use images~generate mask~histogram to generate an ROI file, which are the pixels where the analysis suggests that component is present in a large amount.
2. launch stacks~analyze~axis binary; select the ROI and save the spectrum
4. read in the extracted spectrum
5. Use on of the 2 types of spectra~curve fit (SVD, or stack fit) to fit the extracted spectrum
6. If you are happy with the visual fit, you may conclude the model (the set of reference spectra and the choice of either SVD or stack fit) is valid. Generally it is a good ideal to compare 2 or more different choices to get a fell as to how dependent the maps (shapes and amounts) are on the model chosen
7. If the fit is bad, adjust the model. The deviations may suggest missing components, spectra which have flaws (mis-calibraiton), or cases where the fit is over-determined - too many similar components in the model



6-component stack fit (linear regression)

bio#4c-k-od1.txt
F-test = 3720.00
chi square = 0.00539
Fit components

prot	165.000
sacc	-5.26300
lipid	1.86600
silica	-3.65800
CaCO3	-8.41800
K	-6.23700
const	0.180000

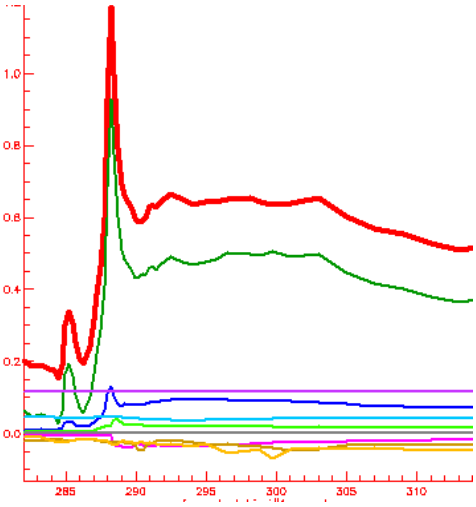


6-component SVD (CGO_optimize)

CG_Optimize: 31 loops
Std. dev. = 0.0117

prot	169.000
sacc	3.51900
lipid	-7.96600
silica	30.8000
CaCO3	-8.07000
K	-6.25700

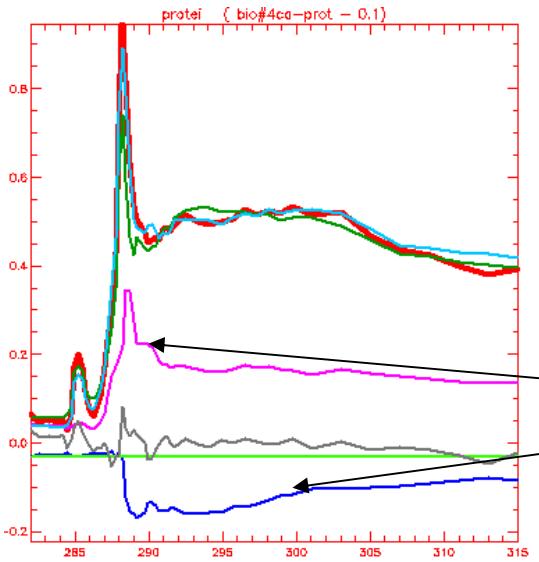
Add one more protein



F-test = 1.00000e+006
chi square = 9.89000e-010
Fit components

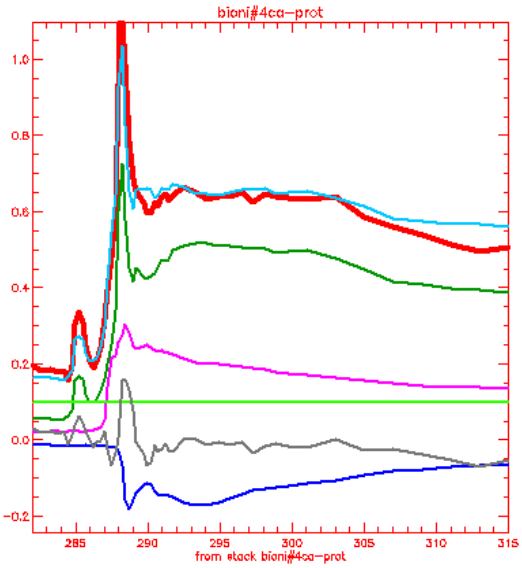
prot	142.000
prot2	25.9000
sacc	-7.25600
lipid	5.90800
silica	8.46300
CaCO3	-10.4430
K	-6.65200
const	0.116000

Spectral curve fitting (2)



F-test = 1340.00
chi square = 0.0289
Fit components
albumin 146.000
sacc -32.0490
lipid 53.4000
const -0.0310

Here one clearly sees the 'fighting' between the internal-saccharide, internal-lipid and the external protein components which leads to an un-acceptable large -ve coefficient for the saccahride

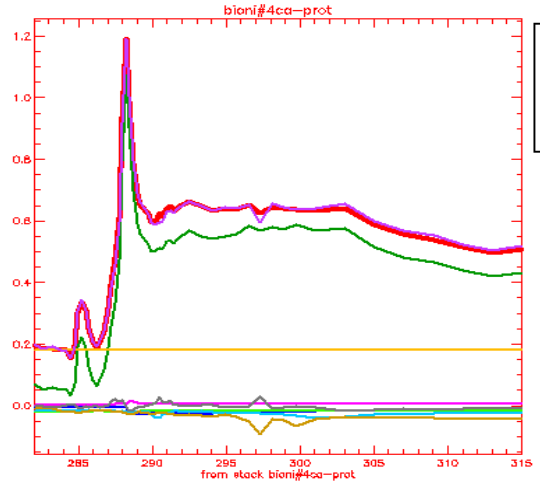


F-test = 341.000
M_correlation = 0.977
chi square = 0.119000
Fit components
alb 143.000
alginate -25.8650
lipid_1 51.6000
const 0.0999000

All external models

Clearly a poorer fit

for **F-test** - larger numbers are better
for **Mcorrelation** - numbers closer to +1.0 are better
- values close ot 0 are bad
for **chi-square** - smaller numbers are better, but be careful - larger numbers of reference spectra will always lead to smaller chi-square so SF always looks better than SVD, even when it isn't

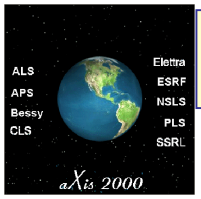


F-test = 3720.00
M_correlation = 0.998
chi square = 0.00539
Fit components
prot 165.000
sacc -5.26300
lipid 1.86600
silica -3.65800
CaCO3 -8.41800
K -6.23700
const 0.180000

All internal models

Clearly a better fit

Dropping out K and CaCO3 will further improve it



What do F-test parameters mean ?

http://geography.uoregon.edu/geogr/topics/interpstats.htm

Interpreting test statistics, p-values, and significance

Analysis	Test statistic	Null hypothesis	Alternative hypothesis	Results	p-value	significance	decision
Regression analysis	F (see note 5)	no relationship between response and predictor vars.	relationship between response and predictor vars.	big F	small p (<0.05)	yes (there is a relationship)	reject H_o, accept H_a
				small F	big p (>0.05)	no (there is not a relationship)	don't reject H_o

5) The null hypothesis here is that there is not a general relationship between the response (dependent) variable and one or more of the predictor (independent) variables, and the alternative hypothesis is that there is one. ***A big F, with a small p-value*, means that the null hypothesis is discredited**, and we would assert that there is a *general relationship between the response and predictors* (while a small F, with a big p-value indicates that there is no relationship).