

Q & A

Paolo Berzaghi

Q: is any reason to upgrade from WinISI II v. 1.50 to WinISI v. 1.60?

- ????????
- Don't have 1.6;
- Used once during WinISI training, didn't see differences in the calibration, Center, Select;
- Does Foss has something to add?

Tough questions

- What is Paolo's procedure for taking all of the data he receives from the NIRSC to make a calibration?
- what is the procedure for selecting calibration samples that are representative of the samples selected that

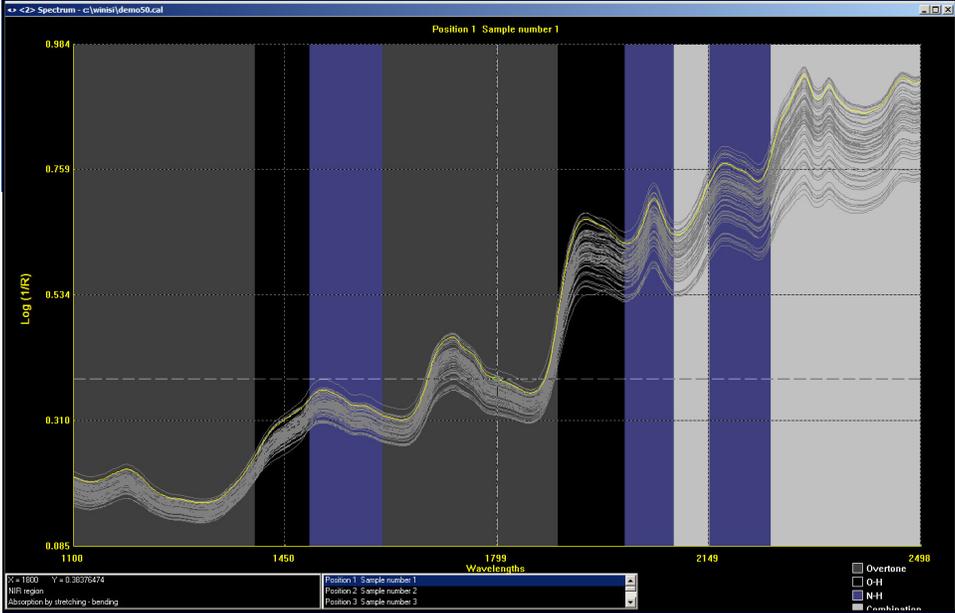
TOP SECRET!!!

Tough questions

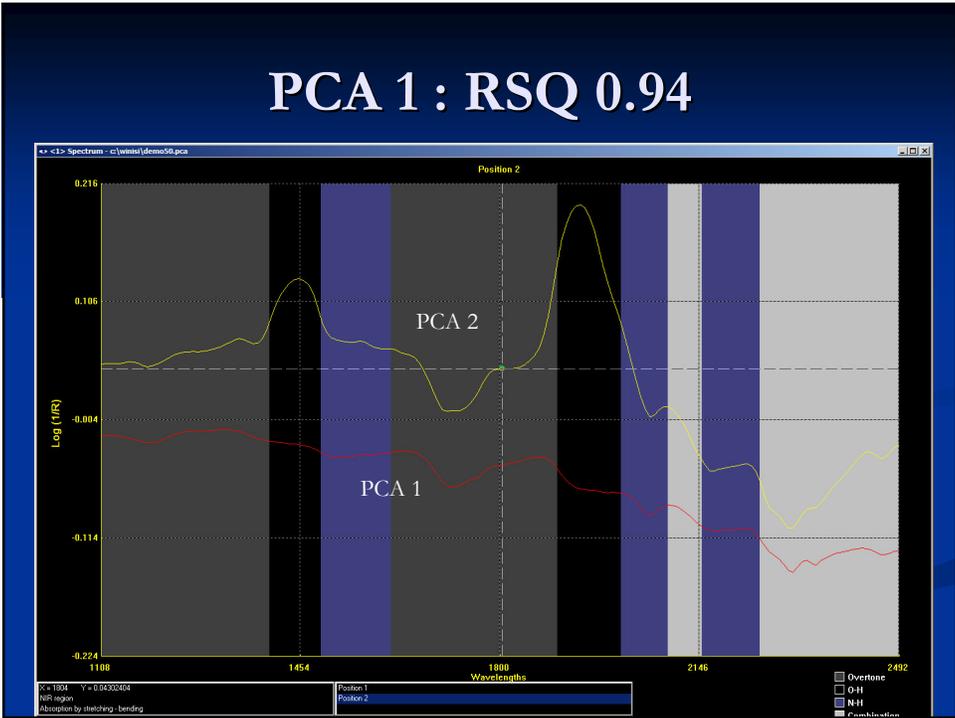
Q: What is Paolo's procedure for taking all of the data he receives from the NIRSC to select samples and make a calibration?

A: In the past it was PCA only. In the past years we had more specific projects (NDFD, starch) I've had to switch to PLS1 to be more effective. PCA vs. PLS1????

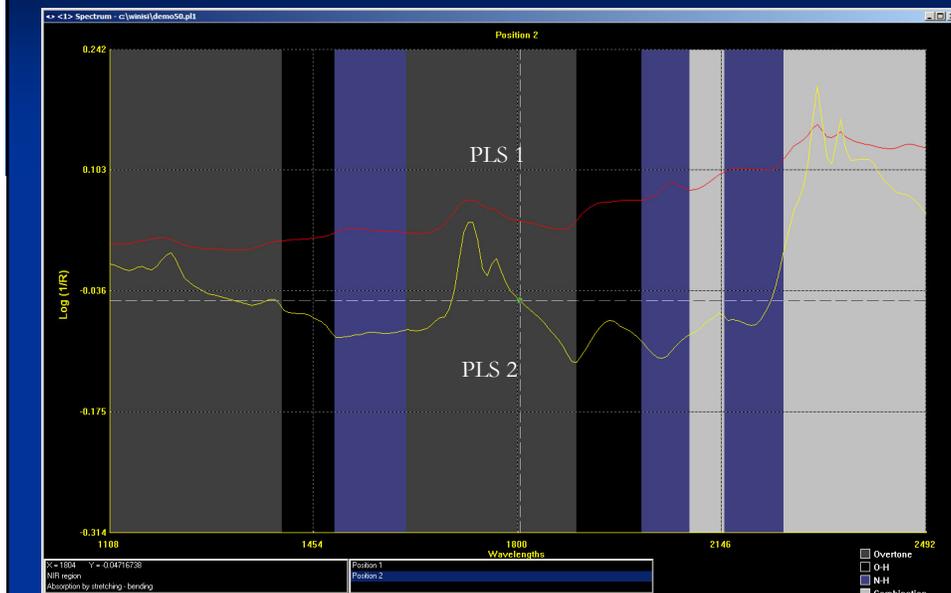
Demo.cal (5000)



PCA 1 : RSQ 0.94



PLS 1 : RSQ 0.64



PCA vs PLS

- PCA – it models the shape of spectra regardless of sample composition
- PLS1 – it models the shapes of spectra which are RELEVANT to sample composition;
- The PLS1 allow us to select those samples that are more important to develop calibration for a specific constituent

Steps to get a good calibration

- Get the right samples!!!!!!
- High quality spectra and high quality chemistry;
- It is cheaper to get good quality spectra than chemistry ---- Take duplicates and average them.
- Prepare a good validation file to fine tune the calibration.

Easy question?

Q: What is a validation set and how do you use a validation set?

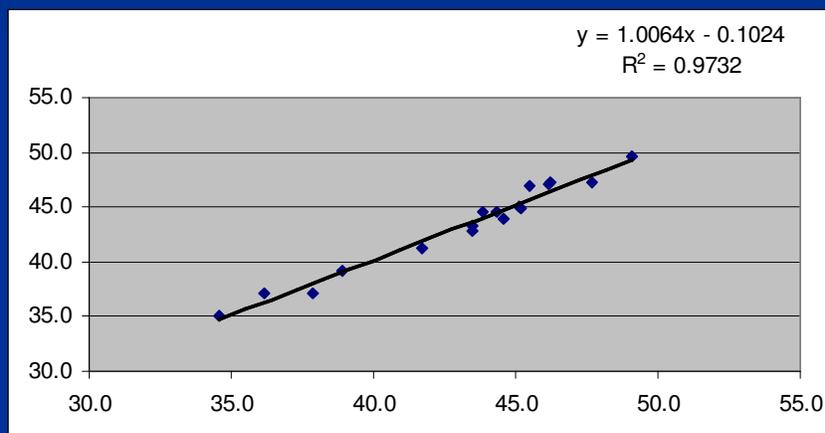
A: a validation set is like a navigation system, without it you may get lost, if you are lucky or you studied the maps (Shenk's notes) you will get where you want to be.

Easy question?

- A validation is a data set of spectra and chemistry not used in calibration. The values predicted based on the spectra are compared to its chemistry defining the accuracy of the calibration.
- Validation samples have to be representative of population variability
- A random validation set may “randomly” give you the wrong answer

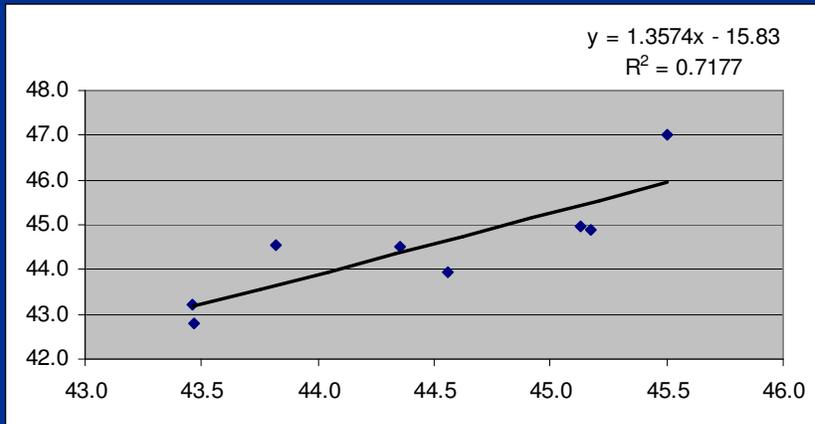
Validation set

Bias=-0.17; SEP=0.68



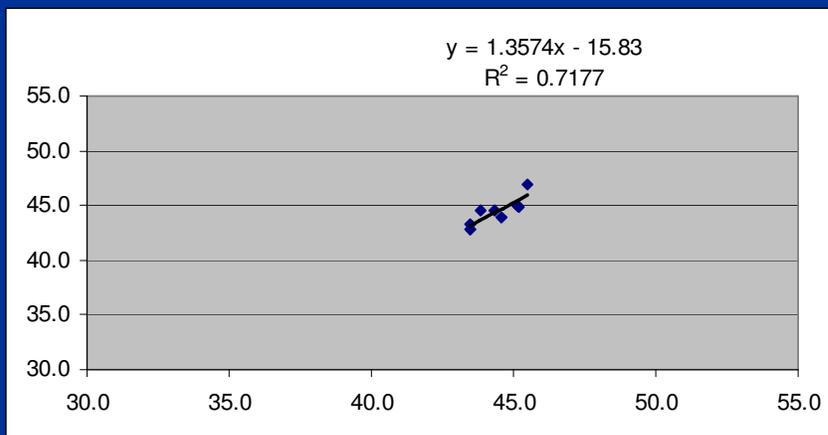
Validation set

Bias=-0.17; SEP=0.47



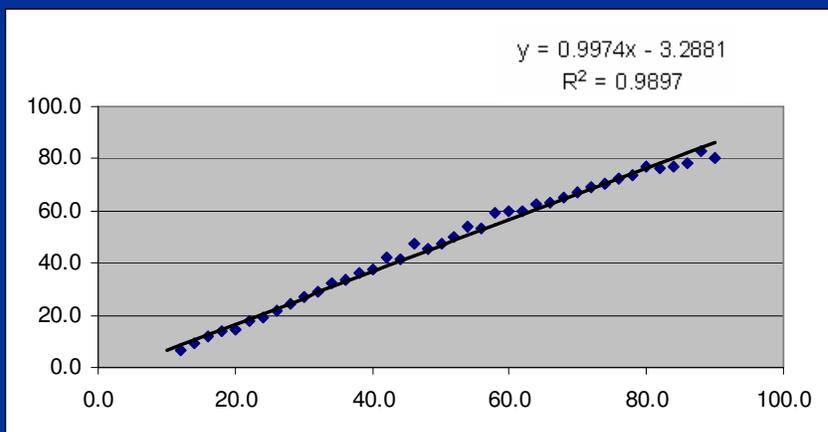
Validation set

Bias=-0.17; SEP=0.47



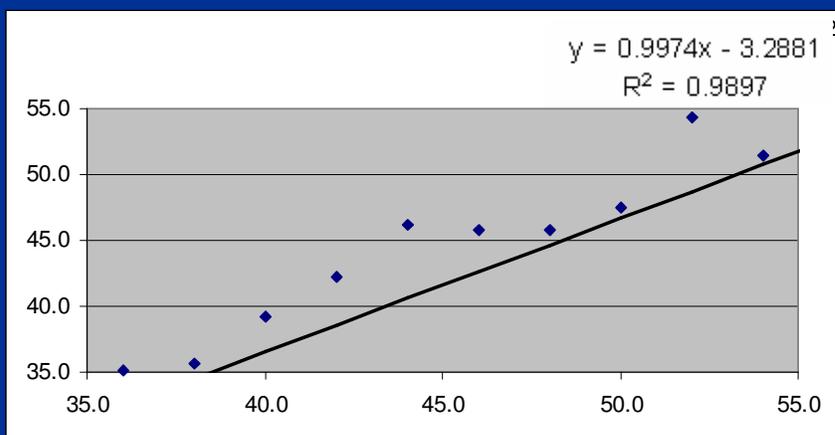
Validation set

Bias=-3.07; SEP=3.87



Validation set

Bias=-3.07; SEP=3.87



Monitor results



Predict Constituents Statistics

CP vs. CP

SEP:	0.882	Number of Samples:	50
Means:	21.611 21.526	Standard Deviations:	3.561 3.648
Bias:	0.086	Bias Limit:	0.580
SEP[C]:	0.887	SEP[C] Limit:	1.256
Slope:	0.947	RSQ:	0.941
Ave. Global H:	0.912	Ave. Neighbor. H:	0.327

Pos.	Sample #	LAB	ANL	Residual	Bias	Global H	Neigh. H
1	C73242006	21.430	20.236	1.194	1.108	1.586	0.392
2	C69244039	21.260	20.750	0.510	-0.424	1.238	0.475
3	N80140002	22.580	23.169	-0.589	-0.675	1.157	0.298
4	60341026	27.230	25.688	1.542	1.456	1.413	0.493
5	41445013	21.160	21.007	0.153	0.067	1.812	0.259
6	41143414	20.910	22.687	-1.777	-1.863	0.930	0.379
7	C71242011	19.350	19.994	-0.644	-0.730	0.945	0.000
8	R60343009	25.510	26.104	-0.594	-0.680	1.661	0.311
9	R50444026	24.170	23.934	0.236	0.151	1.587	0.390
10	A17143024	20.490	20.025	0.465	0.379	0.595	0.384
11	A17344030	24.500	23.457	1.043	0.957	1.092	0.559
12	C68641015	25.250	23.693	1.557	1.471	1.511	0.323
13	171	21.100	20.021	1.079	0.993	1.016	0.494
14	24132	15.940	16.087	-0.147	-0.233	0.592	0.246
15	4525	14.000	12.968	1.032	0.946	2.128	0.575
16	10713	22.700	22.992	-0.292	-0.378	1.421	0.328

Close Output Options XY Plot Residual Plot Correct Redo (0) Undo Undo All HELP

Validation file

Develop equations with the full spectrum

Input Files		Output Files	
Calibration file	07ahy50.cal	Equation file	07ah50.eqa
Repeatability file	std_ah.nir	Loadings file	.lod
Validation file	nftaah.cal	Prediction file	.anl

Choose a regression method

Modified PLS
 PLS
 PCR

Calculate Options... Preferences... Exit to the Project Manager Help

Validation file

Modified PLS Regression Statistics						
Input File	calah.cal		REP File	None		
Validation File	valah.cal		Equation File	None		
Math Treatment	1, 4, 4, 1		Number of variables	173		
Scatter Corrected	SNV and Detrend		Downweight outliers	No		
Constituent	PROTEIN		Number of samples	620		
Mean	20.756		Range	7.82 - 27.52		Standard deviation 2.990
	SEC	R ²	F	SECV	1-VR	
1	1.816	0.631	1060.20	1.832	0.628	
2	1.098	0.865	1073.90	1.139	0.856	
3	1.060	0.874	45.85	1.102	0.865	
4	0.951	0.899	150.80	1.007	0.888	
5	0.876	0.914	109.82	0.936	0.903	
6	0.842	0.921	51.85	0.906	0.909	
7	0.758	0.936	145.35	0.842	0.921	
8	0.700	0.945	105.99	0.795	0.930	
9	0.665	0.951	67.10	0.773	0.934	
10	0.639	0.954	51.79	0.732	0.941	
11	0.614	0.958	51.88	0.711	0.944	
12	0.593	0.961	44.54	0.697	0.946	
13	0.577	0.963	34.53	0.693	0.947	

Validation file

- Validation file is one of the most important file you must have;
- It MUST have the best quality data you can have because based on that data:
 - We define accuracy
 - We fine tune calibrations

Repeatability file

Q: Is there a repeatability file created using NIRSC instruments that is available to NIRSC labs for calibration purposes? If there is, who would I have to contact to get it?

A: Contact Paolo!!

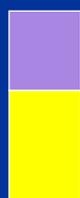
Repeatability file

Q: Is there a repeatability file created using NIRSC instruments that is available to NIRSC labs for calibration purposes? If there is, who would I have to contact to get it?

Repeatability file

■ Calibration data set

Rep file



Temperature variation (in rep. file)



Intrument variation (in rep.file)



NO rep

Analyze Constituents Statistics										
ADF vs. ADF										
SED:	1.180		Number of Samples:		20					
Means:	27.595	26.577	Standard Deviations:		1.596	1.830				
Bias:	1.018		Bias Limit:		0.000					
SED[C]:	0.612		SED[C] Limit:		0.000					
Slope:	0.825		RSQ:		0.893					
Ave. Global H:	1.457	1.272	Ave. Neighbor. H:		0.712	0.577				
Pos.	Sample No.	ANL	ANL	Residual	Bias	GH1	NH1	GH2	NH2	
1	1	27.30	25.12	2.18	1.16	1.41	0.65	1.17	0.53	
2	2	25.12	23.43	1.69	0.67	1.65	0.82	1.42	0.68	
3	3	25.34	23.91	1.44	0.42	1.09	0.60	0.98	0.47	
4	4	26.81	25.95	0.85	-0.16	1.08	0.62	0.92	0.48	
5	5	27.44	26.15	1.29	0.27	1.56	0.61	1.36	0.51	
6	6	26.11	25.24	0.87	-0.15	1.64	0.61	1.51	0.60	
7	7	28.45	27.34	1.11	0.09	1.17	0.68	0.83	0.32	
8	8	28.02	27.28	0.74	-0.28	1.73	0.67	1.64	0.52	
9	9	26.11	24.52	1.59	0.58	1.08	0.60	0.86	0.34	
10	10	26.06	24.98	1.09	0.07	1.16	0.51	0.97	0.42	
11	11	30.48	29.58	0.90	-0.11	1.82	0.94	1.69	0.77	
12	12	27.73	25.81	1.92	0.91	1.33	0.54	1.21	0.51	
13	13	29.22	29.29	-0.07	-1.08	1.72	0.85	1.52	0.74	

With REP targeting particle size

Analyze Constituents Statistics

ADF vs. ADF

SED: 0.588 Number of Samples: 20
 Means: 26.432 26.089 Standard Deviations: 1.419 1.586
 Bias: 0.342 Bias Limit: 0.000
 SED(C): 0.490 SED(C) Limit: 0.000
 Slope: 0.852 RSQ: 0.908
 Ave. Global H: 1.457 1.272 Ave. Neighbor. H: 0.712 0.577

Pos.	Sample No.	ANL	ANL	Residual	Bias	GH1	NH1	GH2	NH2
1	1	26.33	24.99	1.33	0.99	1.41	0.65	1.17	0.53
2	2	24.06	23.21	0.85	0.51	1.65	0.82	1.42	0.68
3	3	24.39	23.87	0.51	0.17	1.09	0.60	0.98	0.47
4	4	25.65	25.47	0.18	-0.16	1.08	0.62	0.92	0.40
5	5	26.33	25.91	0.42	0.08	1.56	0.61	1.36	0.51
6	6	25.15	25.10	0.05	-0.29	1.64	0.61	1.51	0.60
7	7	27.28	27.05	0.23	-0.11	1.17	0.68	0.83	0.32
8	8	26.89	26.64	0.25	-0.09	1.73	0.67	1.64	0.52
9	9	24.96	24.35	0.61	0.27	1.08	0.60	0.86	0.34
10	10	25.38	24.76	0.62	0.28	1.16	0.51	0.97	0.42
11	11	28.86	28.69	0.17	-0.17	1.82	0.94	1.69	0.77
12	12	26.70	25.43	1.27	0.93	1.33	0.54	1.21	0.51
13	13	27.91	28.41	-0.50	-0.84	1.72	0.85	1.52	0.74

Close Output Options XY Plot Residual Plot Redo (0) Undo Undo All Help

Repeatability file

- Rep file have to target specific factors for a lab or a network like the NIRC;
- For NIRC we need REP file to increase robustness for particle size, temperature and instrument alignment;
- For your own eqa in your own lab you probably need a rep file targeting different factors

Repeatability file

- Rep file target specific factors for a lab or a network like the NIRC;
- For NIRC we need REP file to increase robustness for particle size, temperature and instrument alignment;
- For your own eqa in your own lab you probably need a rep file targeting different factors;
- Use sample selection and rep file to increase robustness of eqa

More questions?