

## Using NIRS to determine sample differences: range, homogeneity,...

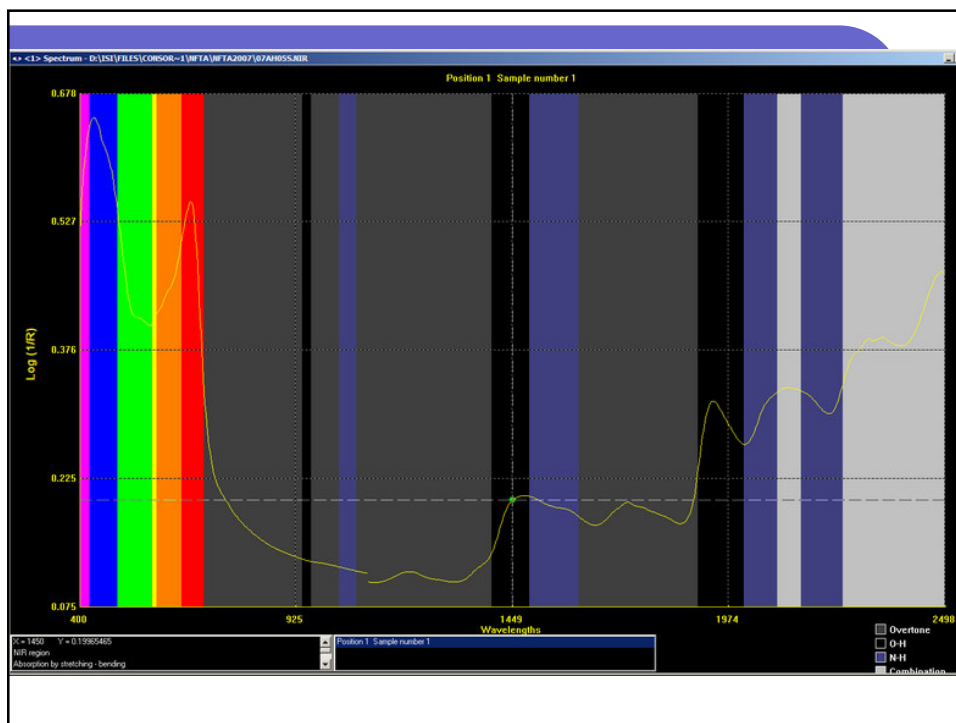
Paolo Berzaghi  
NIRS Consortium/Univ. of Padua, Italy

## Factors that affect spectra

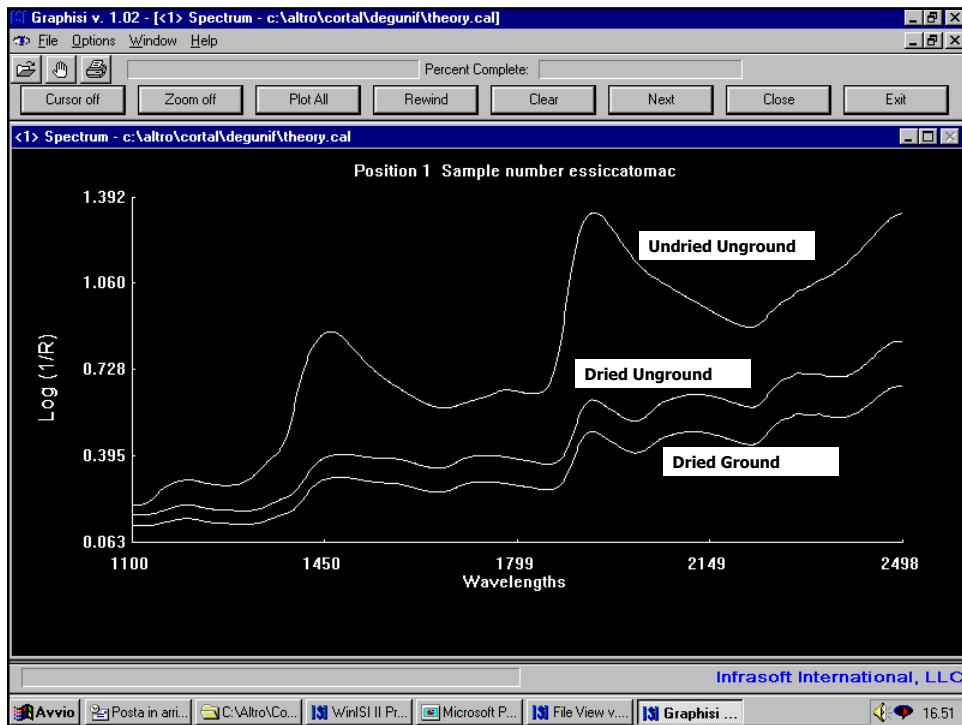
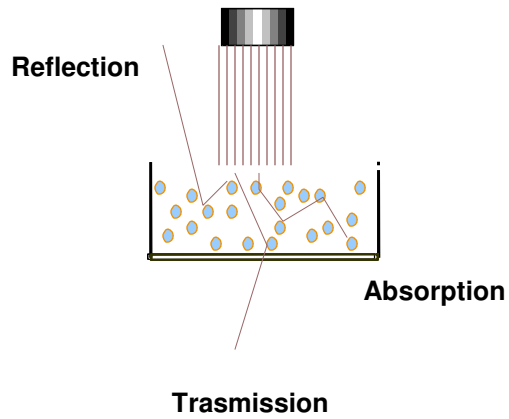
- *Sample prep: drying, grinding;*
- *Chemical composition;*
- Lab environment (Temp. and humidity);
- *Cups/cuvette sampling and packing;*
- Instruments within manufactures;
- Instruments between manufactures;
- .....

## Factors that affect spectra

- *Sample prep: drying, grinding;*
  - Drying will affect chemical composition like residual moisture and fiber content;
  - Particle size will affect the amount of light absorbed;

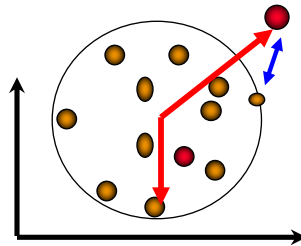


## Light - Matter interaction



## H values

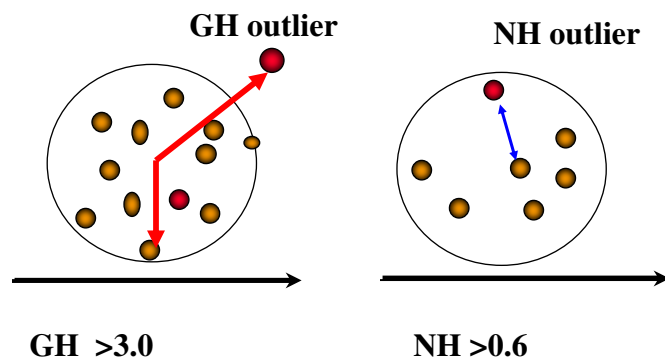
- Scores are used to calculate H values.



Global H number = distance from the population average [CENTER] (maximum value=3.0)

Neighborhood H number = distance from the closest sample [select] (value=0.6)

## PCA & LIB = their use



## Effect of sample grinding

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

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

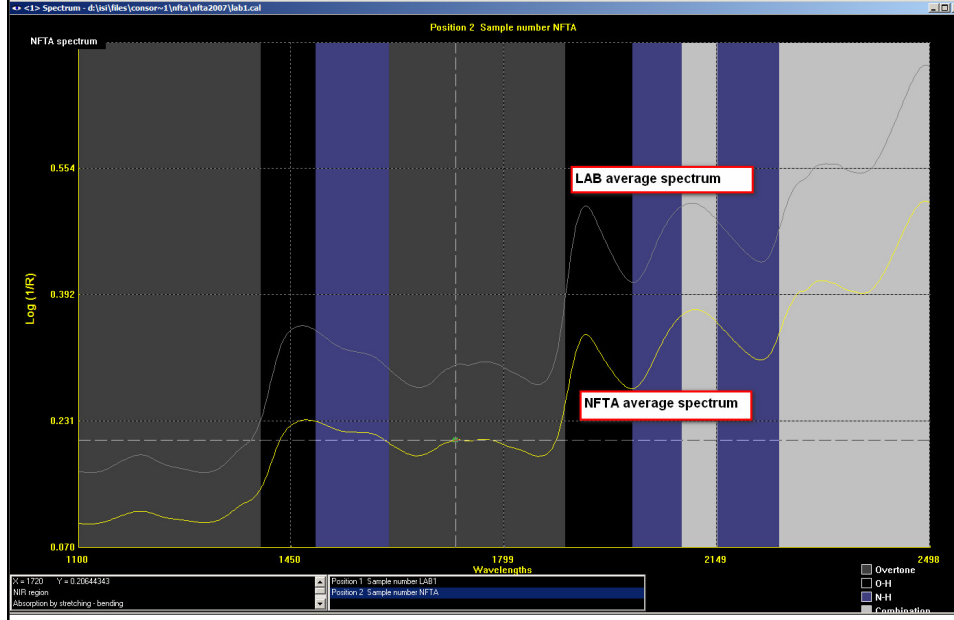
## Factors that affect spectra

### Real Problem:

*“Using NIRSC equation I get good results for NFTA, but predictions don't match composition for my samples”*

1. Your chemistry is not good!
  - Nope, we get straight A's with NFTA samples
2. Is something wrong with your sample prep?
  - Nope, we dry at 60°C and grind at 1mm.

## Factors that affect spectra



## Factors that affect spectra

1. Is something wrong with your sample grinding?
  - "Yea, I checked the grinder that is used for the NIR and the impact ring is worn almost completely smooth."

## Factors that affect spectra

- *Chemical composition;*
  - Calibration are regression that needs to be built with a range of chemical composition.
  - Odd chemical composition/sample won't contribute to improve calibration;

Look for high NH and GH values.


That doesn't mean high NH and GH have odd composition

## Effect of size of calibration on H values

The image shows two screenshots of a software window titled 'Predict Constituents Statistics'. The top window is labeled 'Small calibration' and the bottom window is labeled 'Calibration after updates'. Both windows show a dropdown menu set to 'CP vs. CP' and a table of statistical data. In both windows, the 'Ave. Global H' and 'Ave. Neighbor. H' values are highlighted with red boxes.

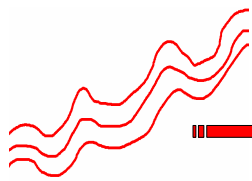
Statistic	Small calibration	Calibration after updates
SEP:	1.108	0.844
Means:	23.842 23.736	23.842 23.742
Bias:	0.106	0.100
SEP[C]:	1.110	0.844
Slope:	1.017	0.964
Ave. Global H:	1.347	1.201
Number of Samples:	77	77
Standard Deviations:	2.922 2.659	2.922 2.905
Bias Limit:	0.659	0.600
SEP[C] Limit:	1.429	1.299
RSQ:	0.856	0.918
Ave. Neighbor. H:	0.685	0.607

## Factors that affect spectra

- *Cups/cuvette sampling and packing;*
  - *It affects repeatability of spectra*  *results*
  - *How to check?*
    - *Look at predictions;*
    - *Look at spectra;*

## Cups/cuvette sampling and packing

*Look at predictions;*



$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

	#1	#2	#3
DM	93.9	94.2	94.2
CP	15.0	14.8	14.8
ADF	34.2	33.9	34.0
NDF	40.1	39.8	40.0

*They depends on*

- *the calibration equation;*
- *spectra;*



## Two prep procedures – same eqa(1)

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

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

## Two prep procedures – same eqa(2)

Analyze Constituents Statistics

ADF vs. ADF

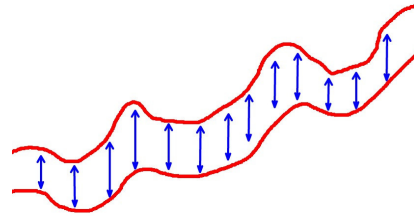
SED: 0.588      Number of Samples: 20  
 Means: 26.432 26.089      Standard Deviations: 1.419 1.586  
 Bias: 0.342      Bias Limit: 0.000  
 SED(C): 0.490      SED(C) Limit: 0.000  
 Slope: 0.852      RSQ: 0.908  
 Ave. Global H: 1.457 1.272      Ave. Neighbor. H: 0.712 0.577

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	0.48
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

## Root means square (RMS)

- Calculated like SEP (accuracy) or SED repeatability



$$RMS = \sqrt{\frac{\sum_1^n (y_{n1} - y_{n2})^2}{n}}$$

## Cups/cuvette sampling and packing

- To increase accuracy run duplicates;
- Check dups with RMS;
- Rescan dups with high RMS;
- Average either spectra or prediction;

SAMPLE	MEAN
1	Only one sample in group, no RMS possible
2	3100
3	2591
4	1972
5	2028
6	5601
7	3997
8	2181
9	4217
10	1276
17	4392
..	
Overall mean:	2252

## Summary

- Spectra contains all the information
  - First look at spectra, apply PCA and look at H's values or at RMS [difficult].
- Use the spectra for prediction but remember that eqa can be built for robustness and predictions may not reveal all of the problem.
- Looking at spectra and prediction give us a complete picture of what's going on.