Some considerations about near Infrared applications.

Pierre Dardenne
dardennepaj@gmail.com

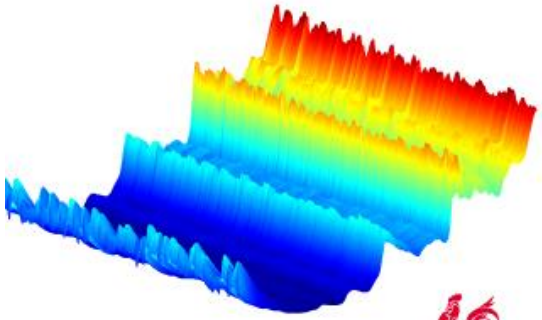
Vibrational Spectroscopy and Chemometrics

Training Session

13 March–17 March 2017

cra-w

Anniversary 10th Edition



Wallonie

Chemometrics applied to vibrational data

Exploratory analysis

Data visualisation

Principal component analysis

Outlier detection

Uncertainty estimation

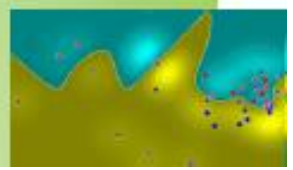
Quantification and classification

Multivariate calibration

Partial Least squares PLS

Multiple linear regression MLR

Support vector machines SVM



Tom Fearn
Paolo Berzaghi
Mark Westerhaus

Abbas, Ouissam (MIR spectroscopy)

Baeten, Vincent (Raman / Sampling)

Dardenne, Pierre (NIR considerations)

Fernández Pierna, Juan Antonio (Chemometrics)

Vermeulen, Philippe (Hyperspectral Imaging)

Vincke, Damien (Hyperspectral Imaging)

Lecler, Bernard (Transfer/Standardization)

Minet, Olivier (NIR networks)

Sinnaeve, Georges (NIR online)

Vibrational Spectroscopy and Chemometrics

Training Session

13 March–17 March 2017

Chemometrics applied to vibrational data

Exploratory analysis

Data visualisation

Principal component analysis

Outlier detection

Uncertainty estimation

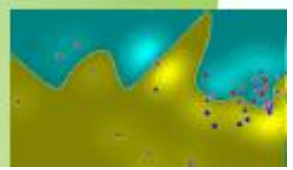
Quantification and classification

Multivariate calibration

Partial Least squares PLS

Multiple linear regression MLR

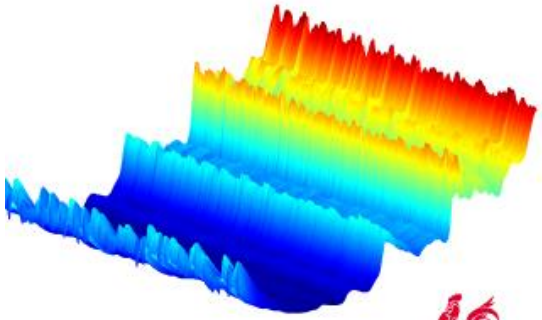
…pport vector machines SVM



TIPS & TRICKS

cra-w

10th Edition



Wallonie

Tom Fearn
Paolo Berzaghi
Mark Westerhaus

Abbas, Ouissam (MIR spectroscopy)

Baeten, Vincent (Raman / Sampling)

Dardenne, Pierre (NIR considerations)

Fernández Pierna, Juan Antonio (Chemometrics)

Vermeulen, Philippe (Hyperspectral Imaging)

Vincke, Damien (Hyperspectral Imaging)

Lecler, Bernard (Transfer/Standardization)

Minet, Olivier (NIR networks)

Sinnaeve, Georges (NIR online)

1

The concept of the mixture model

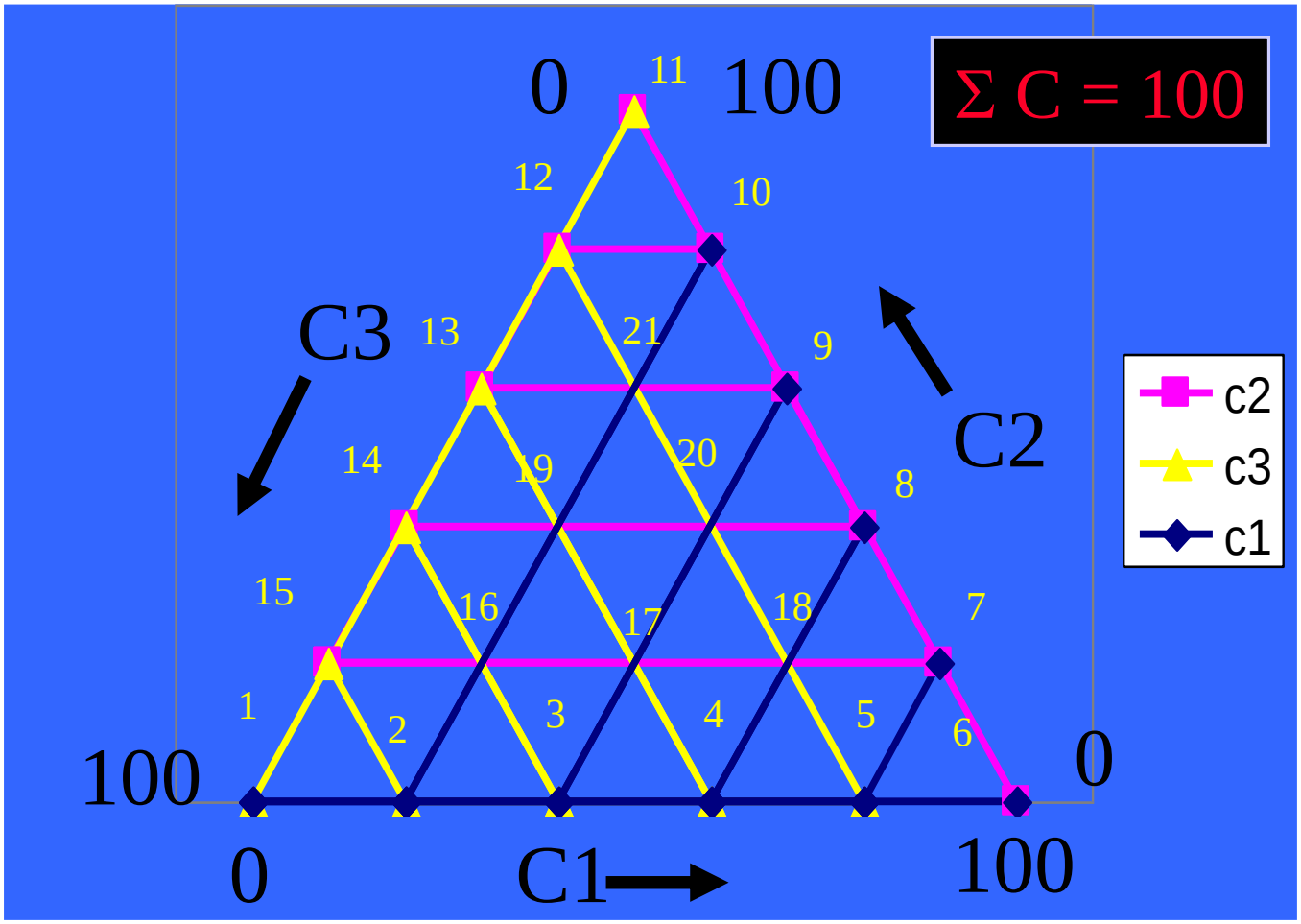
Multivariate calibration and chemometrics for near infrared spectroscopy: which method?

Pierre Dardenne, George Sinnaeve and Vincent Baeten
Département Qualité des Productions Agricole, Centre de Recherches Agronomiques de Gembloux–CRAGx, 24 Chaussée de Namur, B-5030 Gembloux, Belgium

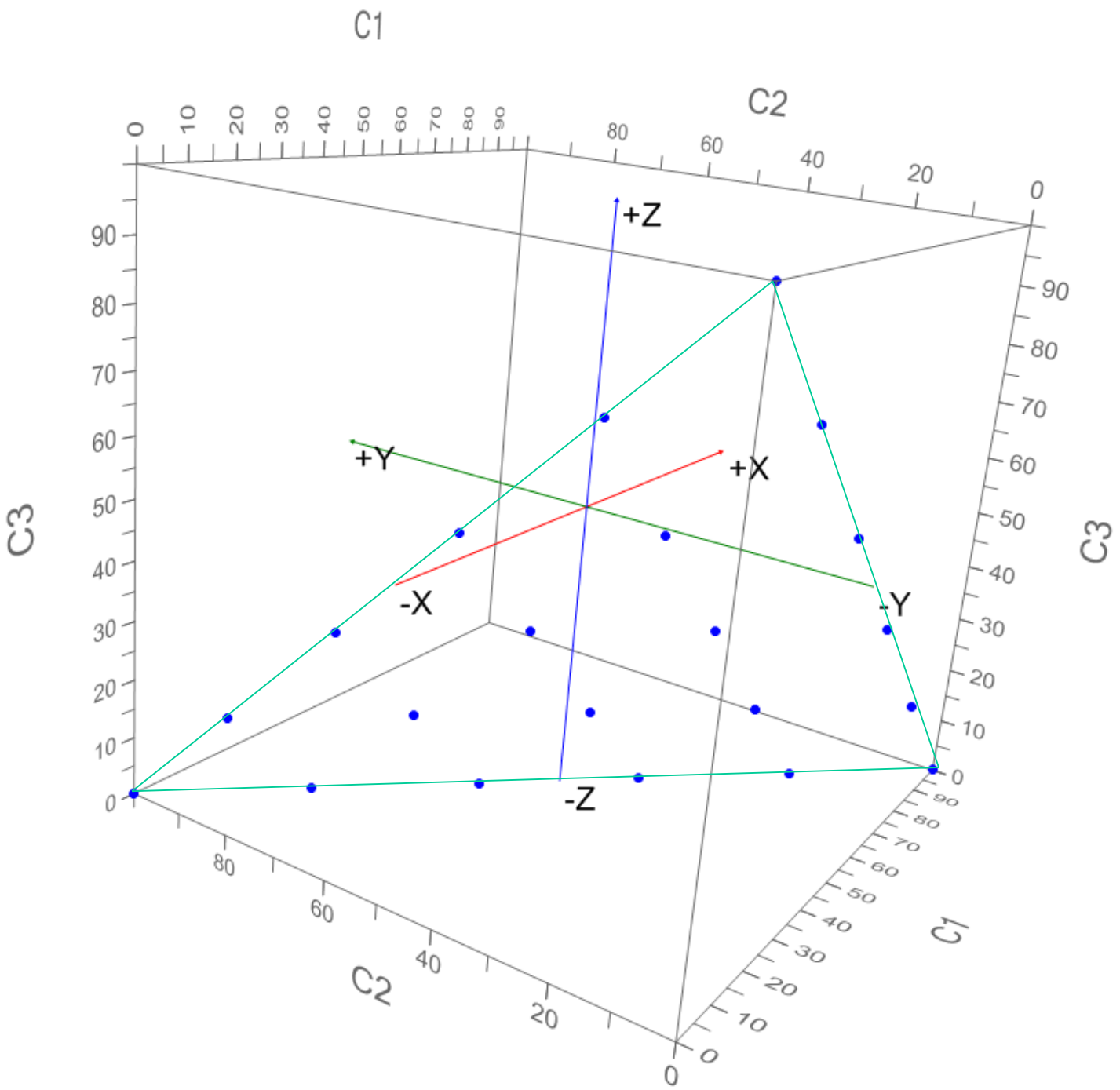
J. Near Infrared Spectrosc. **8,** 229–237 (2000)

3 Constituents : 2 dimensions

N=21

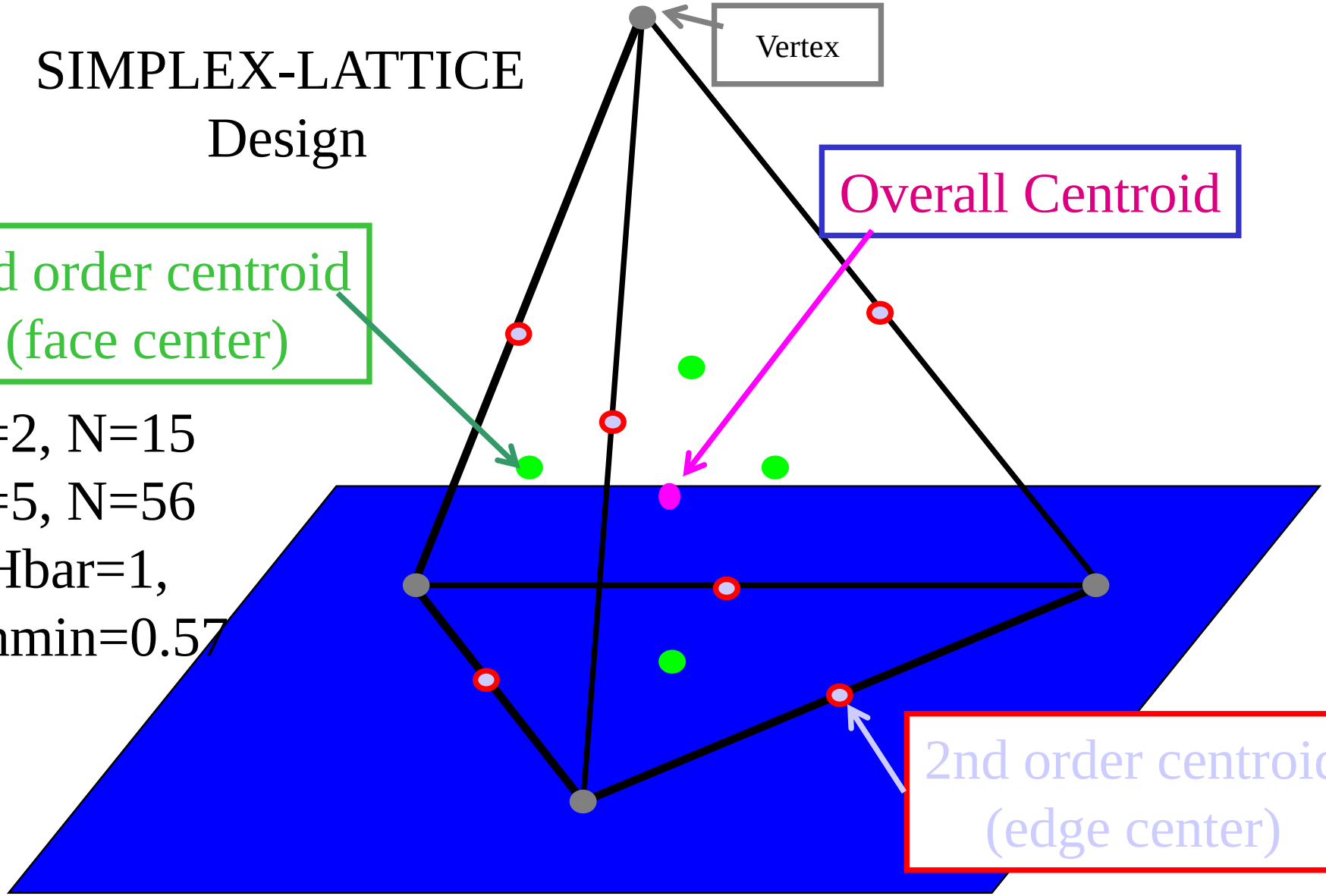


GHbar=1, NHmin=0.60


C1
C2
C3
+Z
+Y
+X
-X
-Y
-Z

4 Constituents : 3 dimensions

SIMPLEX-LATTICE
Design



R=2, N=15
R=5, N=56
GHbar=1,
Nhmin=0.57

p Constituents :

Excel: =COMBIN( )

p-1 dimensions

Ranges /5 → r=5

$$Nspl = C_{r+p-1}^{p-1} = \frac{(r+p-1)!}{r!(p-1)!}$$

$$C_{5+3-1}^{3-1} = \frac{(5+3-1)!}{5!(3-1)!} = \frac{7!}{5!*2!} = \frac{5040}{120*2} = 21$$

p Constituents :

p-1 dimensions

Ranges /5 → r=5

$$Nspl = C_{r+p-1}^{p-1}$$

$$C_{r+p-1}^{p-1} = \frac{(r+p-1)!}{r!(p-1)!}$$

GHbar=1, NHmin=0.57

Const.	Dim.	Nspl
3	2	21
4	3	56
5	4	126
6	5	252
7	6	462
8	7	792
9	8	1287
10	9	2002
11	10	3003
13	12	6108
15	14	11628
:	:	:
25	24	118755

An example: calibrations for Milk Powder (Whole and Skimmed)

SMP – WMP

Protein :	**20 – 40 %**
Fat :	**0 – 30 %**
Carbohydrate :	**35 – 55 %**
Minerals :	**5 - 10 %**
Moisture :	**0 – 5 %**
Additives : ?	**<1%**

+ color
+ particle size
+ temperature
+ instrument
+……

Out of the constraint Σ=100

→9 dim. → 2000 samples required to fill the space with 6 points evenly spread for each dimension

The questions are:

How long does it last to complete a calibration?

How many samples are needed to obtain a robust model?

NIR EQUATION STABILITY OVER TIME

Product: WHOLE PLANT MAIZE SILAGE

Constituents:

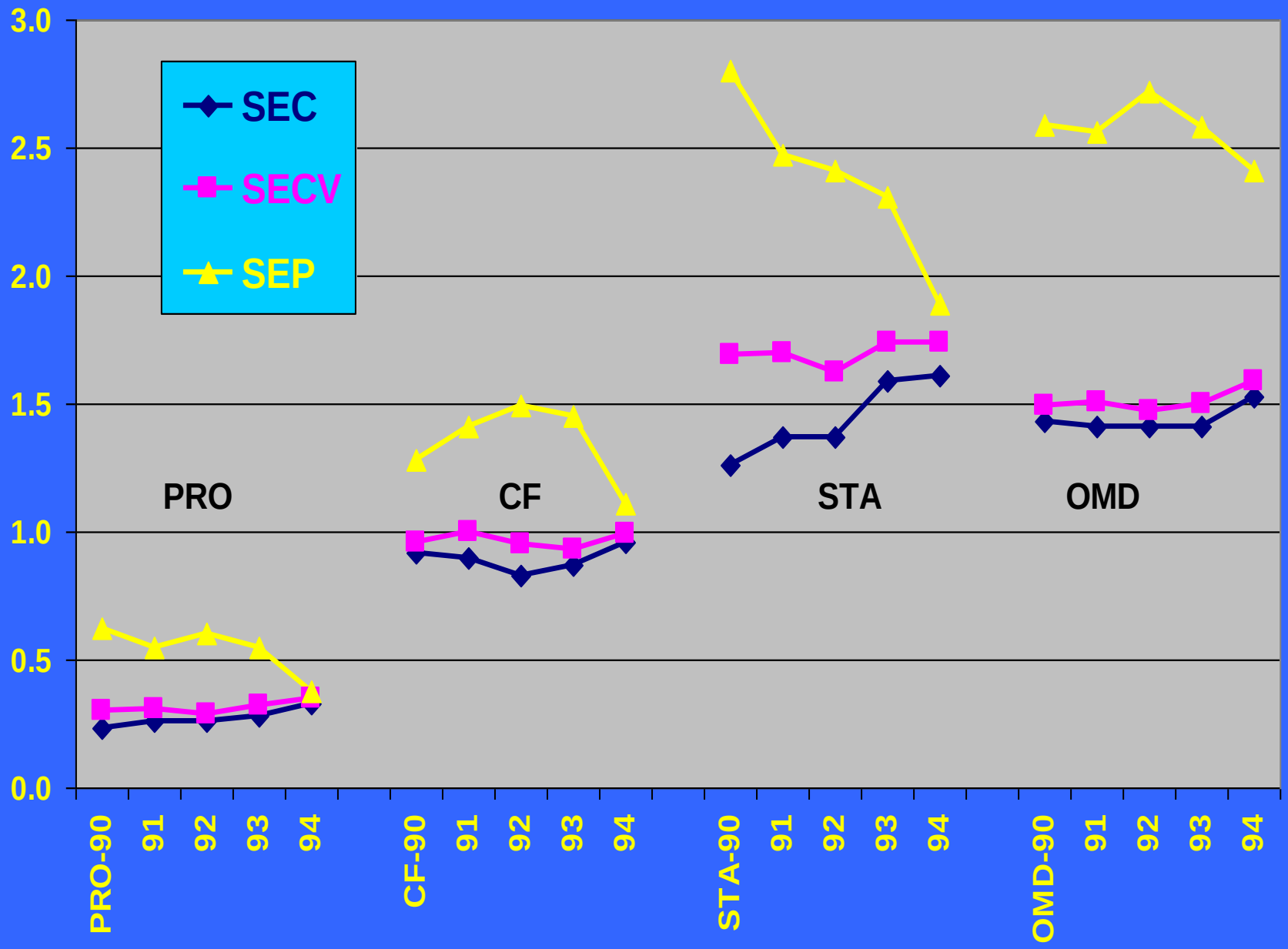
PROTEIN (1600) **CRUDE FIBER (1700)** **STARCH (1950)** **OMD (2400)**

	90	91	92	93	94	95
Model 1	Calib	Validation				
Model 2	Calibration		Validation			
Model 3	Calibration			Validation		
Model 4	Calibration				Validation	
Model 5	Calibration					Valid

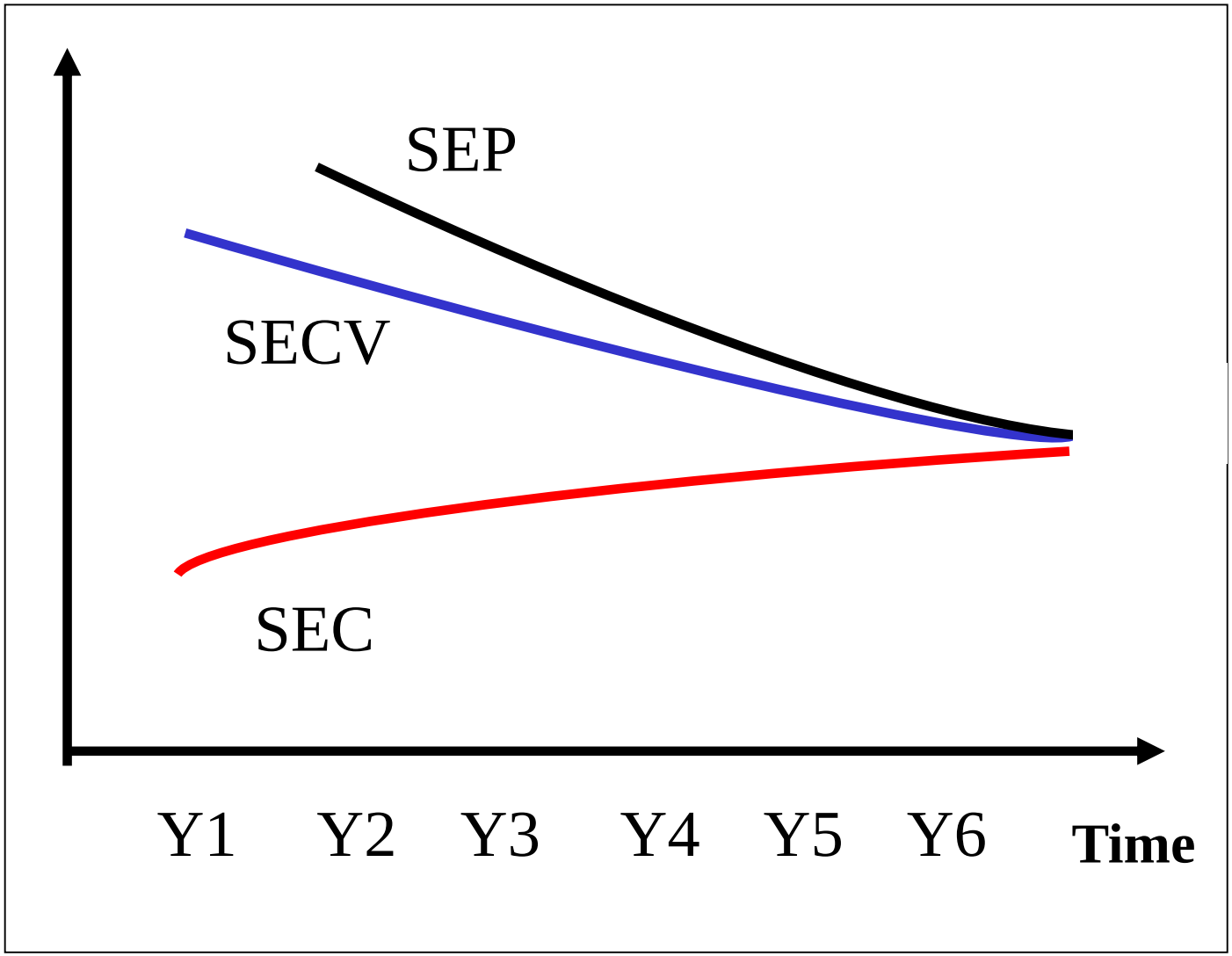
Validation is INDEPENDENT

Dardenne *et al.*,NIR News vol.7 N°5 (1996)

NIR EQUATION STABILITY OVER TIME



GENERAL TRENDS OF MODEL STATS



In NIR models

ADD NEW INFORMATION (new samples) & RECALIBRATE:

1) While SECV # SEC

Rule of thumb: (SECV <**1.1***SEC)

2) While SEP # SECV

Rule of thumb: (SEP < **1.3** * SECV)

Agriculture Handbook, N°643 (1989)

Marten, G. C., J. S. Shenk, and F. E. Barton II,

CONCLUSION Multivariate models

- 1 multivariate space defines a lot of space

- NIR needs numerous DIFFERENT samples to cover the actual variability (the rule of 10 #/term !!)

- It takes time (years) to fill evenly the whole space to make the models stable and robust

- Extrapolation is dangerous (empirical models)

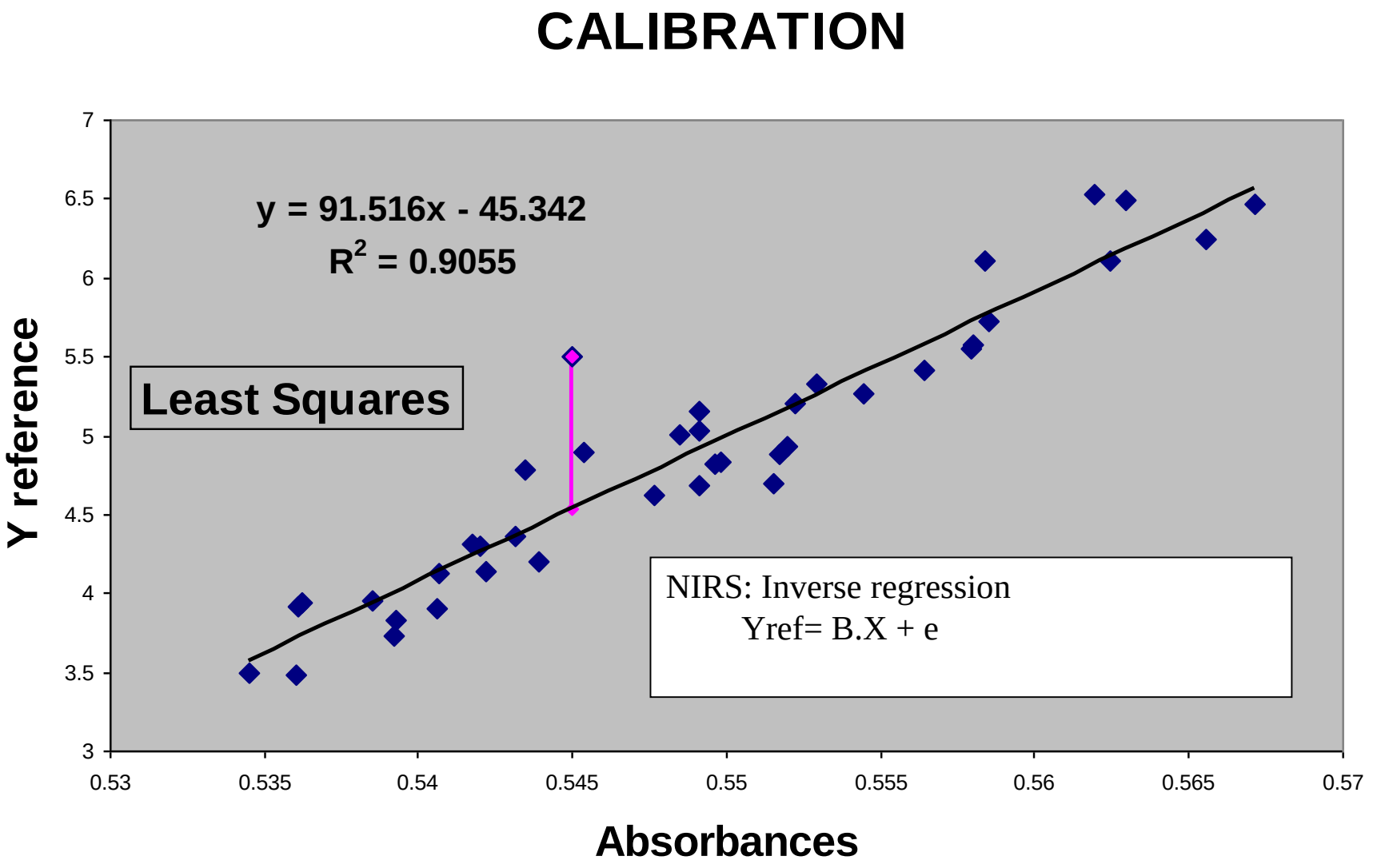
- Always good performance if the unknown samples are already represented in the calibration set (neighbors) : → Local approaches

2

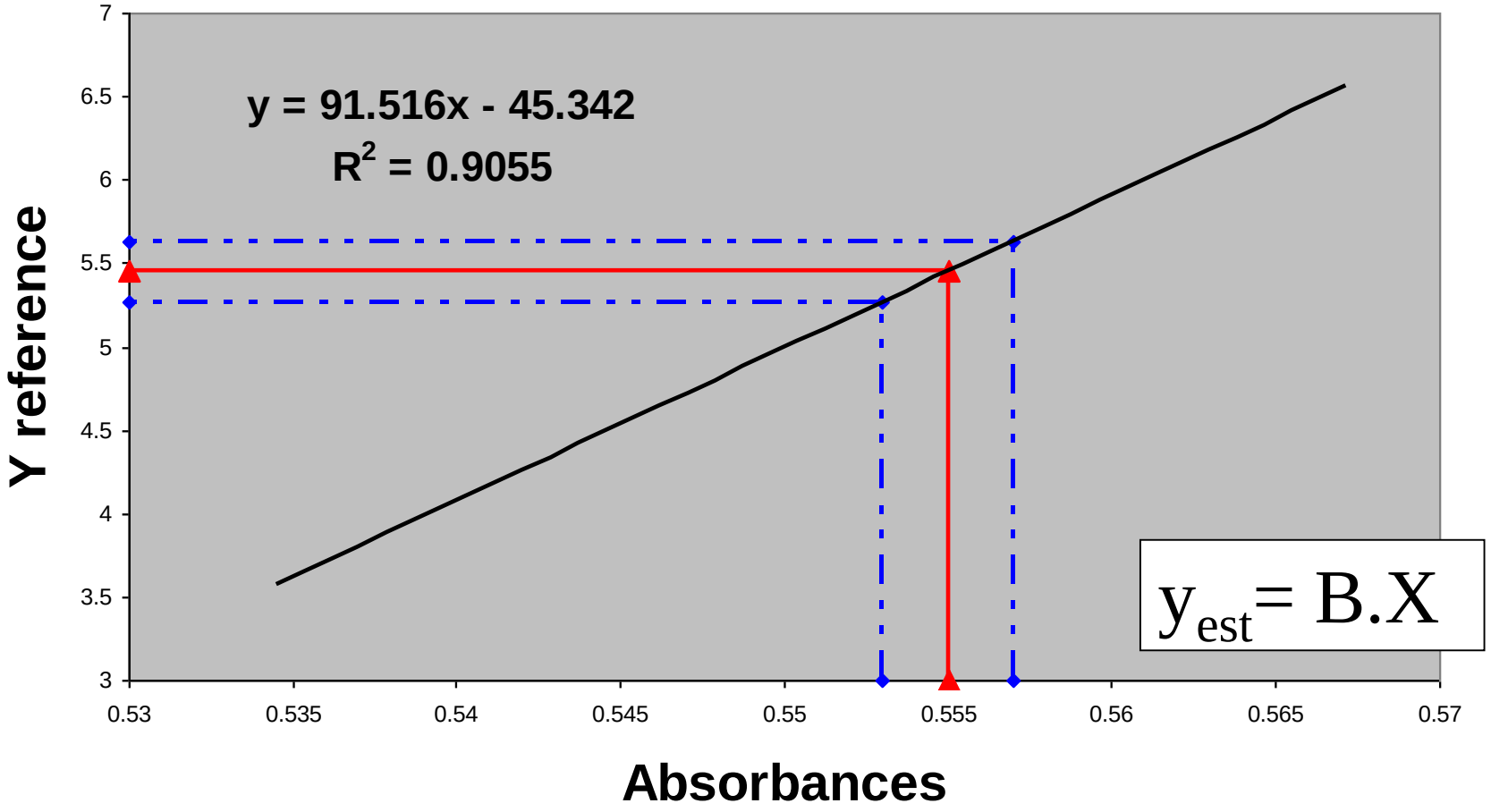
UNCERTAINTY

$$s\left(\hat{y}_i - y_i\right)?$$

uncertainty

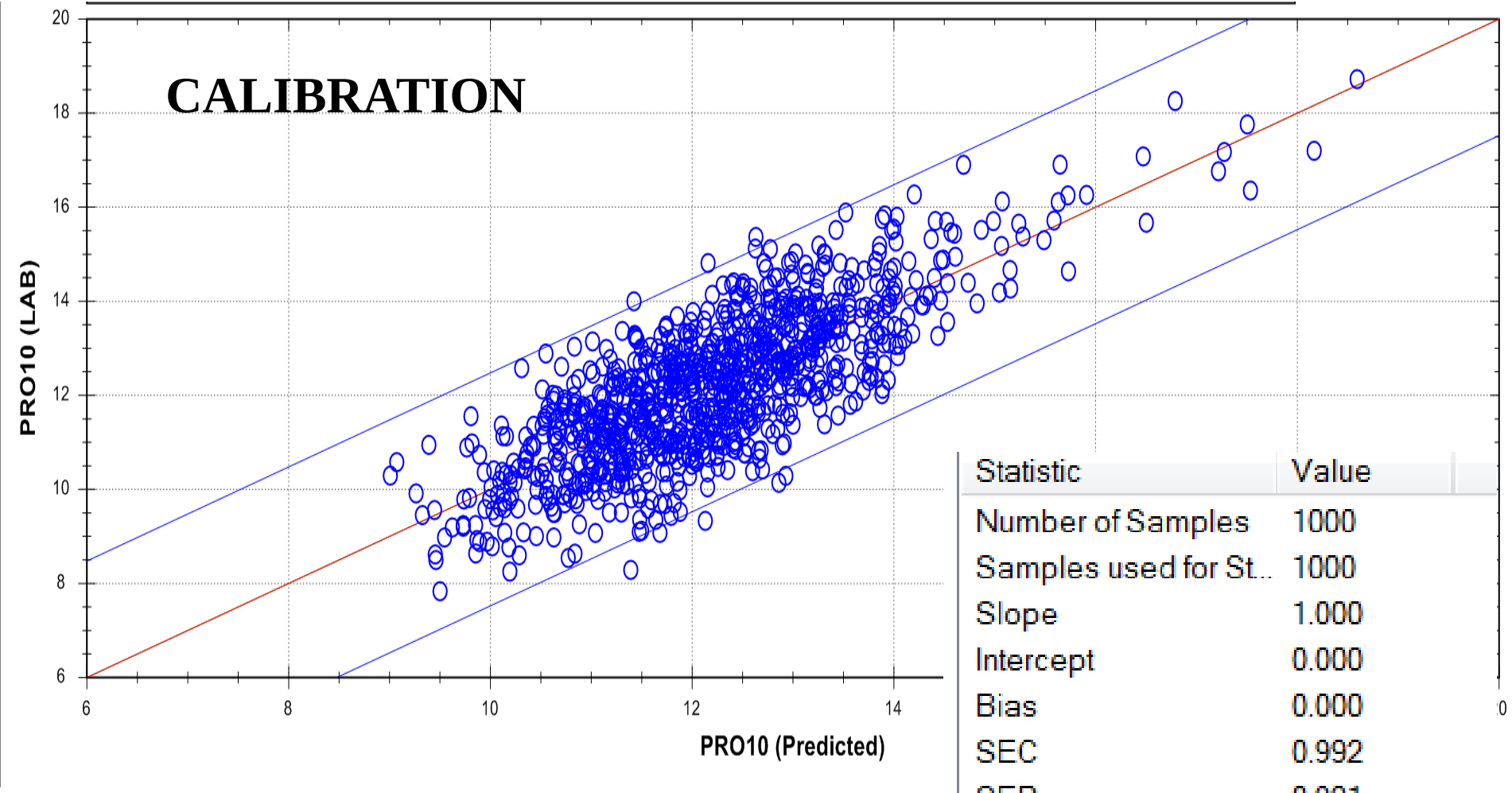


PREDICTION


y = 91.516x - 45.342
R2 = 0.9055
Y reference
Absorbances
7
6.5
6
5.5
5
4.5
4
3.5
3
0.53
0.535
0.54
0.545
0.55
0.555
0.56
0.565
0.57
yest= B.X

2000 spectra of ground wheat
Spilt ½ -> CAL and VAL

CAL y values are modified by a vector of Gaussian noise of mean =0 and sd of 0.1, 0.2, …1.0, ..1.5

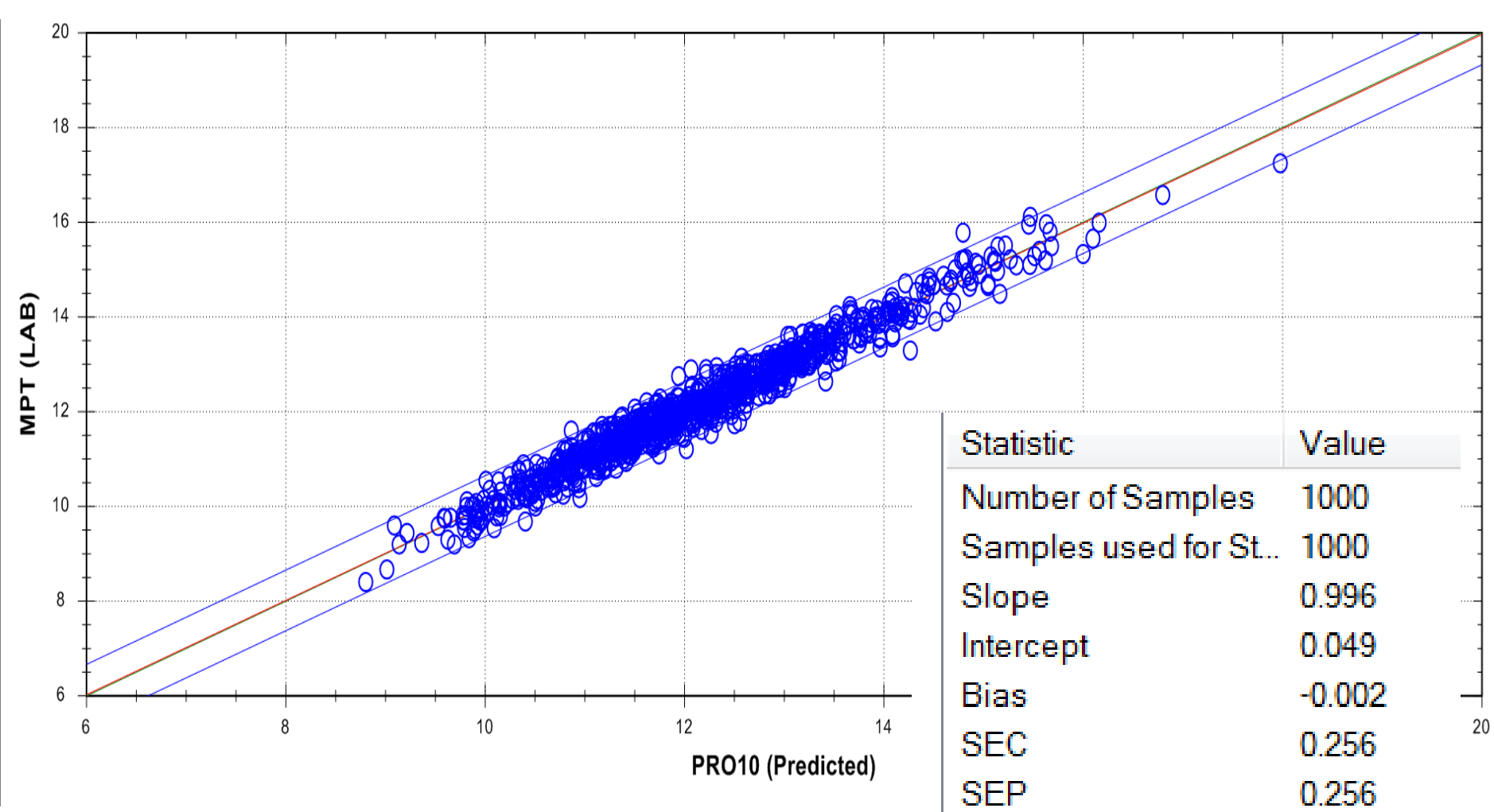


Statistic	Value
Number of Samples	1000
Samples used for St..	1000
Slope	1.000
Intercept	0.000
Bias	0.000
SEC	0.992
SEP	0.991
SEP(C)	0.991
RSQ	0.618
Predicted Average	12.179
Actual Average	12.179
Predicted SD	1.260
Actual SD	1.604

Y modified with noise N(0,1)

mean = 0

sd = 1

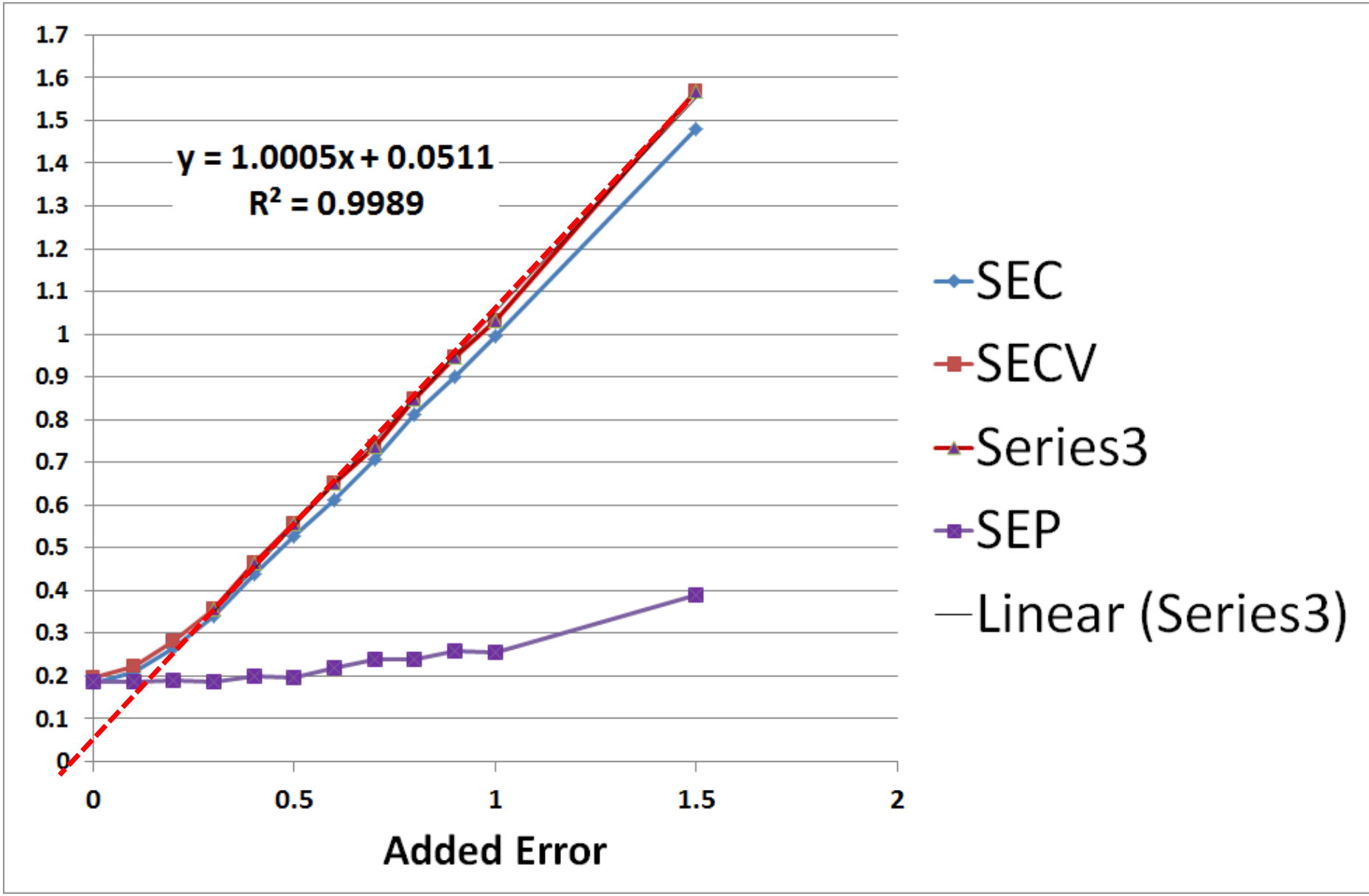


Statistic	Value
Number of Samples	1000
Samples used for St..	1000
Slope	0.996
Intercept	0.049
Bias	-0.002
SEC	0.256
SEP	0.256
SEP(C)	0.256
RSQ	0.960
Predicted Average	12.150
Actual Average	12.148
Predicted SD	1.268
Actual SD	1.288

TEST SET N= 1000

Added

Error	SECV	SEP
0.0	0.20	0.19
0.1	0.22	0.19
0.2	0.28	0.19
0.3	0.36	0.19
0.4	0.47	0.20
0.5	0.56	0.20
0.6	0.65	0.22
0.7	0.74	0.24
0.8	0.85	0.24
0.9	0.95	0.26
1.0	1.03	0.26
1.5	1.57	0.39


y = 1.0005x + 0.0511
R² = 0.9989
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.5
1
1.5
2
Added Error
SEC
SECV
Series3
SEP
Linear (Series3)

Uncertainty of the NIR analyses

$$s(\hat{y}_i - y_i) = \left[(1 + h_i).SEC^2 - S^2_{ref}\right]^{1/2}$$

Faber and Bro, *Chemom., Intell. Lab. Syst.* 61, 133 (2002)
Fernandez Pierna & *al.*, *Chemom., Intell. Lab. Syst.* 65,281 (2003)

$$SEP^2_{actual} = SEP^2_{observed} - SEL^2_{ref}$$

$$SEP_{actual}^2 = SEP_{observed}^2 - SEL_{ref}^2$$

An example:

Protein in wheat: SEL = 0,15

SEPobserved = 0,21

SEPactual = SQRT(0,21^2 – 0,15^2) = 0,15

2 FIGURES OF MERIT:

RMSEP

SEL

JNIRS

Journal of Near Infrared Spectroscopy
25(2) 85–90

DOI: 10.1177/0967033517702395
journals.sagepub.com/home/jns
SAGE

Original Article

Tutorial: Items to be included in a report on a near infrared spectroscopy project

Phil Williams[1], Pierre Dardenne[2] and Peter Flinn[3]

What is the MOST important in a model ?

Not the stats : sec, sep, R2,… = gauges

B coefficients = engine

$$\hat{Y} = X.b$$

$$n.1 = (n.m) * (m.1)$$

CONCLUSION UNCERTAINTY

In many NIR applications, NIR values are more accurate than the reference method ones.

But it is not easy to prove it!

Do not try to obtain the smallest SEC (outliers)

Just try to compute stable Bcoefficients from spectra containing the expected future variation.

3

LOD

!., J. Near Infrared Spectrosc. 17, 59–67 (2009)

Received: 6 January 2009 ■ Revised: 7 March 2009 ■ Accepted: 10 March 2009 ■ Publication: 23 March 2009

JNIRS
JOURNAL
OF
NEAR
INFRARED
SPECTROSCOPY

Rapid detection of melamine in milk powder by near infrared spectroscopy

than 1 ppm.[21] In order to overcome the limitation of the methods which had already been evaluated, this study tried to establish a fast detection method for pure melamine in milk powder by near infrared (NIR) spectrometry together with LS-SVM and has shown some promise. The detection limit of this method was lower than 1 ppm. It conforms to the national standard for melamine in infant milk products. The

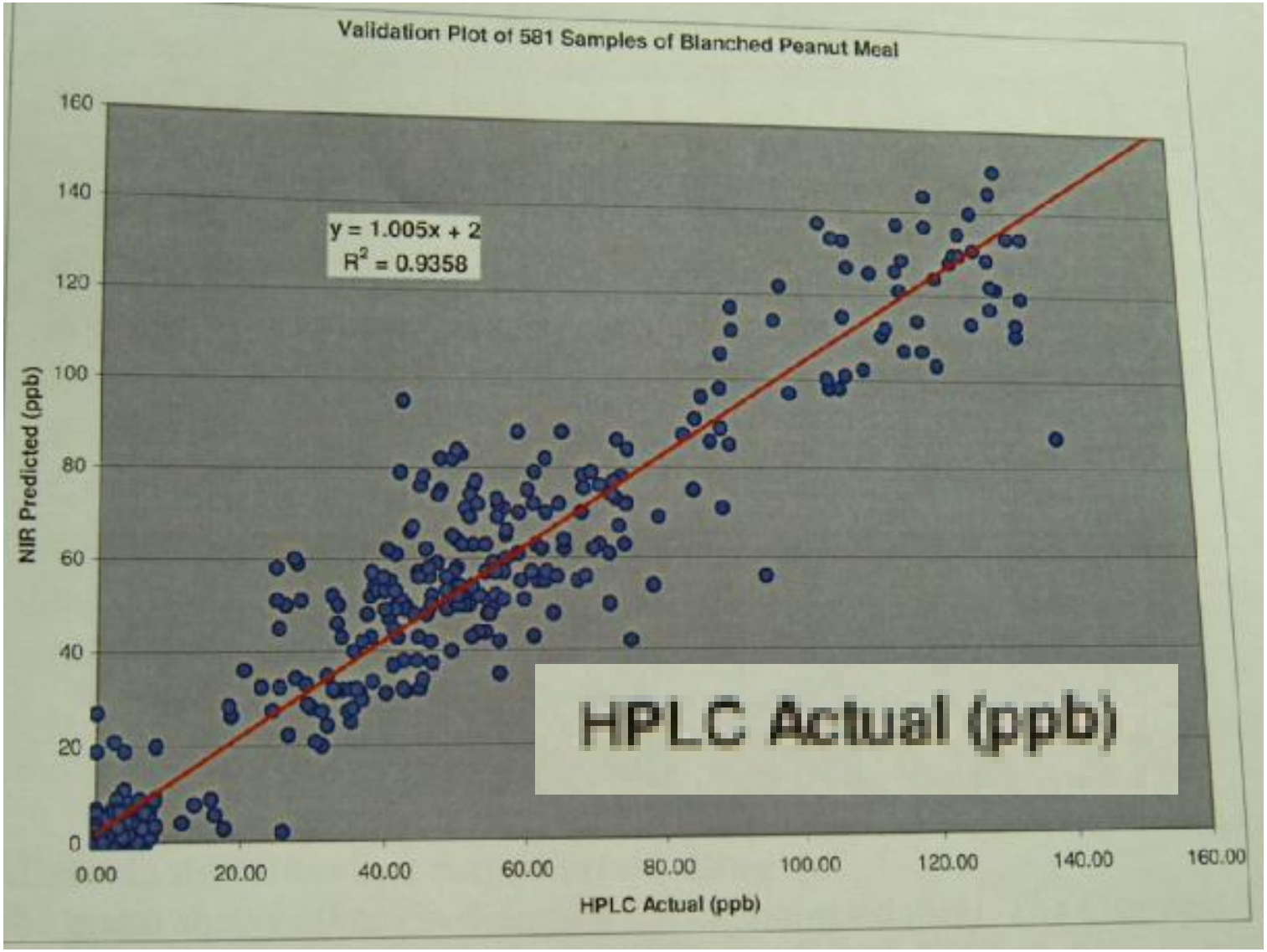
 J. Agric. Food Chem. **2009,** *57,* 3974–3980
DOI:10.1021/jf900587m

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Melamine Detection in Infant Formula Powder Using Near- and Mid-Infrared Spectroscopy

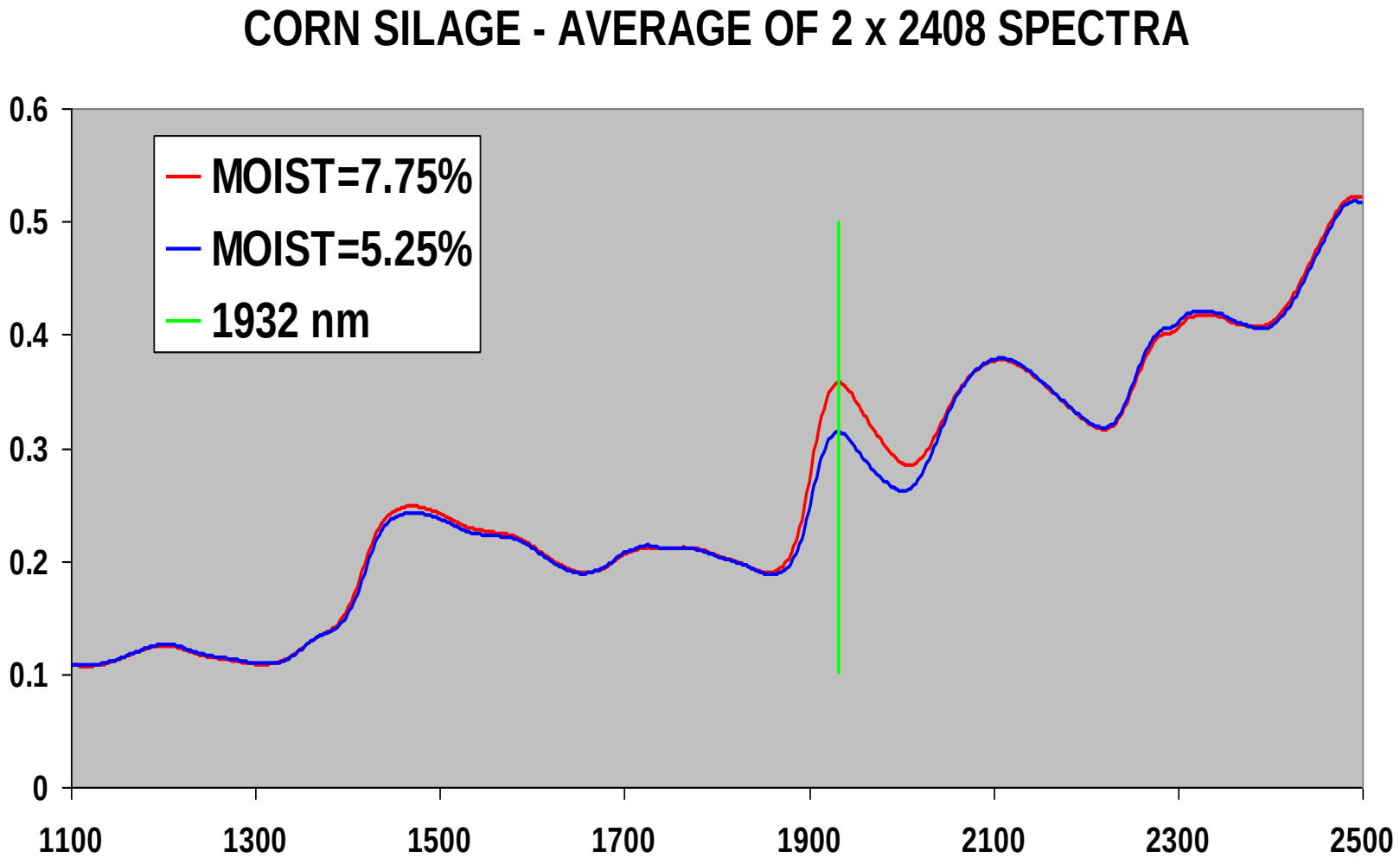
models are not matrix independent. New calibration models are required for application to different brands or formulations of infant formula powders or other food products. It is expected that the analytical approach for quantifying or detecting melamine using NIR or FTIR spectroscopy would be applicable to other products. The NIR and FTIR methods meet the need for a rapid, simple, and available technique for detecting 1 ppm melamine in infant formula powder.

Rapid and Non-destructive Mycotoxin Assay with NIRS, *Russel Wilkie* (ICNIR2005, Aukland)



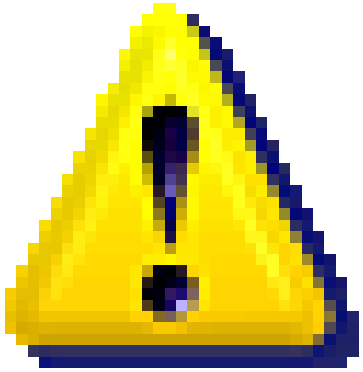
LOD of NIRS

CORN SILAGE DATA SET 4816 spectra sorted/H2O %



LOD of NIRS

	1932
MOIST=7.75%	0.357243
MOIST=5.25%	0.313645
2.5	0.043598
1	0.017439



If <0.1%

A units	Water %	ppm	ppb
0.020 000	1	10 000	10 000 000
0.002 000	0.1	1 000	1 000 000
0.000 200	0.01	100	100 000
0.000 020	0.001	10	10 000

Villeneuve d'ascq, 5-6 Décembre 2012

Polytech'Lille, Université Lille1, FRANCE

Connexion

NAVIGATION

- Accueil
- Lieu de la conférence
- Conférenciers invités
- Liste de communications acceptées
- Programme CHIMIOMETRIE 2012
- Comités
- Résumé et inscription
- Publication dans CHEMOLAB
- Challenge 2012
- Prix et récompenses
- Hébergement
- Exposants / Partenaires / Sponsors
- Devenir partenaire / Sponsoriser cette conférence

CHALLENGE 2012

Le Challenge est ouvert !

Le challenge organisé par le Groupe Français de Chimiométrie se présente sous la forme d'une énigme. Il s'agit de trouver les réponses d'un fichier "test" à partir d'un fichier dit d'étalonnage. Les données de ce dernier fichier peuvent être polluées (données manquantes, données aberrantes, etc.).

Le fichier d'étalonnage et le fichier "test" sont mis à votre disposition ci-dessous. L'analyse des solutions se fait au cours du congrès. Les congressistes ayant proposé les meilleurs solutions seront invités à présenter brièvement leur méthodologie lors d'une session spéciale CHALLENGE. Les meilleures solutions seront primées.

NOUVEAUTÉ : Afin de favoriser la participation des jeunes chimiométriciens (moins de 30 ans), le comité scientifique a décidé cette année de créer une catégorie spécifique dite "Junior". Ces contributions au Challenge seront ainsi étudiées séparément de celles des séniors.

Téléchargez le fichier de données : ICI

Description du CHALLENGE CHIMIOMETRIE 2012 :

270 spectra in a CAL set with known concentration of a contaminant.

276 spectra in a TEST set to be predicted.

CAL SET (only soya meal (SM))

240 with 4 concentrations (0.10, 0.25, 0.50, 1.00%)

+30 randomly selected from the clean set

270 spectra

TEST SET

T1 = 40 SM randomly selected with high contamination (2, 3, 4, 5%)

T2 = 50 randomly selected from the clean set

T3 = 43 soyameal spl from another origin + another instrument

T4 = 78 wheat gluten (39 in duplicates)

T5 = 65 maize gluten .

Total : 276

Foss XDS

Foss 6500
Clean & Conta.
Melamine
Cyanuric Acid

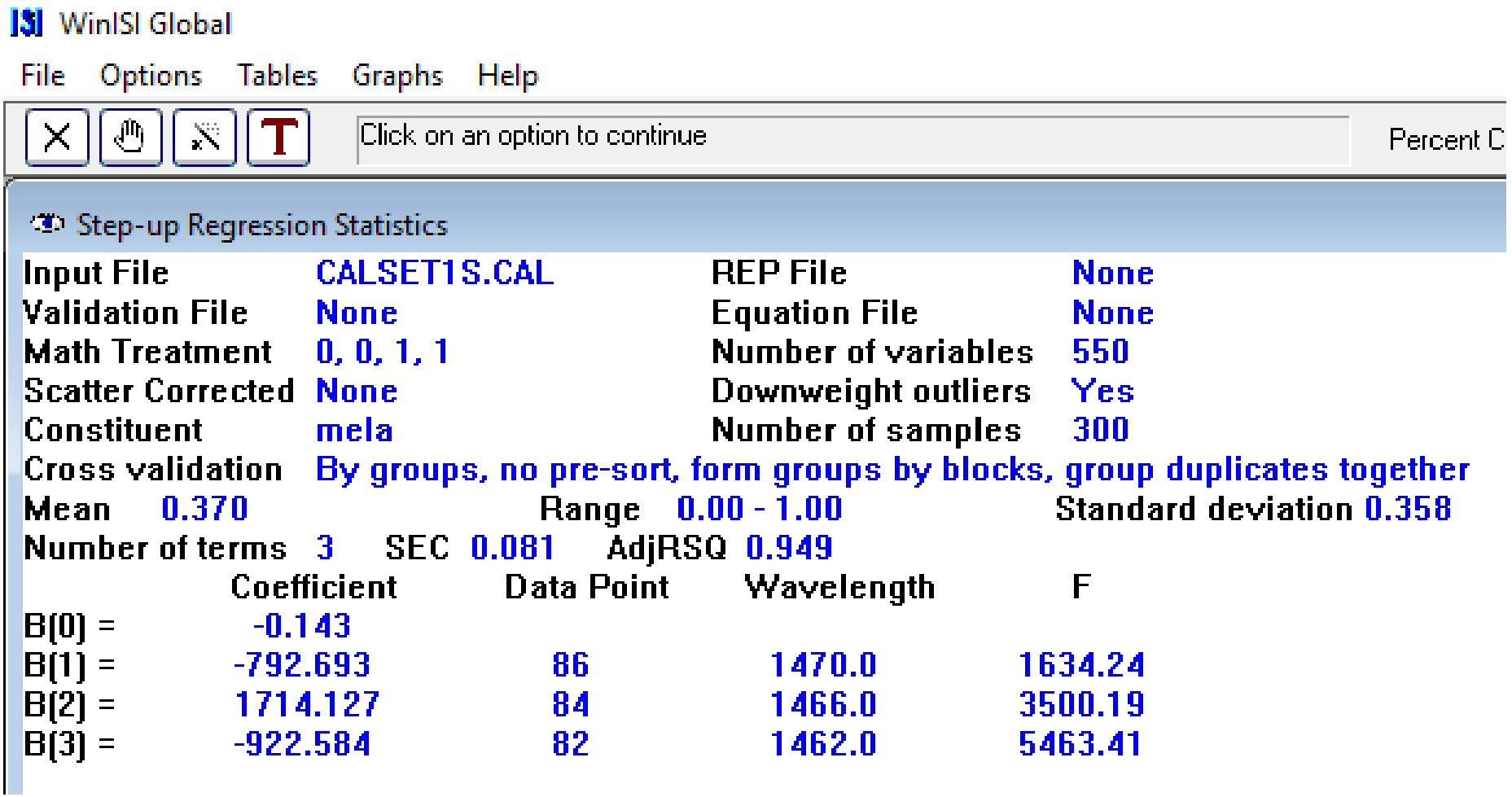
Step-up Regression Statistics

Input File	calset1s.cal	REP File	None
Validation File	None	Equation File	mlr2.eqa
Math Treatment	0, 0, 1, 1	Number of variables	550
Scatter Corrected	None	Downweight outliers	Yes
Constituent	mela	Number of samples	300
Cross validation	By groups, no pre-sort, form groups by blocks, group duplicates together		
Mean 0.370		Range 0.00 - 1.00	Standard deviation 0.358

Number of terms 3 SEC 0.082 AdjRSQ 0.948

	Coefficient	Data Point	Wavelength	F
B(0) =	-0.131			
B(1) =	-1768.915	83	1464.0	5102.85
B(2) =	2269.486	84	1466.0	5246.59
B(3) =	-501.577	87	1472.0	1979.13

MLR : Multi Linear Regression – manual step up

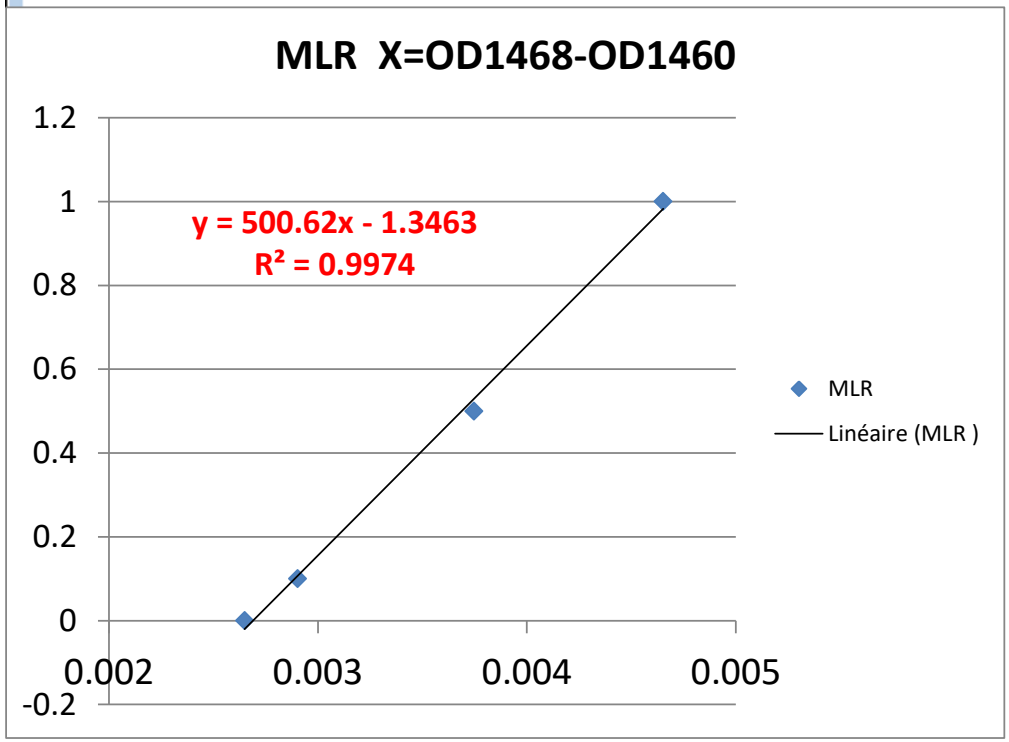


Average by group of 60 samples : 0.0, 0.10, 0.25, 0.50, 1%

Step-up Regression Statistics

Input File	totmeans6.cal	REP File	None
Validation File	None	Equation File	None
Math Treatment	0, 0, 1, 1	Number of variables	550
Scatter Corrected	None	Downweight outliers	Yes
Constituent	mela	Number of samples	5
Cross validation	By groups, no pre-sort, form groups by blocks, group duplicates together		
Mean 0.370	Range 0.00 - 1.00	Standard deviation 0.399	
Number of terms 2	SEC 0.032	AdjRSQ 0.994	

	Coefficient	Data Point	Wavelength	F
B(0) =	-0.209			
B(1) =	518.336	85	1468.0	613.22
B(2) =	-521.870	81	1460.0	614.07



1/500 = 0.002

A OD difference of
0.002 = 1% melamine

A Units	µLog	Melamine %	ppm	ppb
0.002	2000	1	10 000	10 000 000
0.000 2	200	0.1	1 000	1 000 000
0.000 02	20	0.01	100	100 000
0.000 002	2	0.001	10	10 000

MELAMINE ABSORBS 10 x LESS THEN WATER !!

O Abbas *et al.*, *J. Near Infrared Spectrosc.* **21,** 183–194 (2013)
Received: 16 October 2012 ■ *Revised: 26 February 2012* ■ *Accepted: 14 March 2013* ■ *Publication: 26 March 2013*

Detection of melamine and cyanuric acid in feed ingredients by near infrared spectroscopy and chemometrics

O. Abbas,* B. Lecler, P. Dardenne and V. Baeten

Food and Feed Quality Unit, Valorisation of Agricultural Products Department, Walloon Agricultural Research Centre (CRA-W), Chaussée de Namur 24, B-5030 Gembloux, Belgium. E-mail: o.abbas@cra.wallonie.be

Chemometrics and Intelligent Laboratory Systems 152 (2016) 157–162

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy☆

J.A. Fernández Pierna, D. Vincke, V. Baeten, C. Grelet, F. Dehareng, P. Dardenne *

Walloon Agricultural Research Centre (CRA-W), Valorisation of Agricultural Products Department, Chaussée de Namur n°24, 5030 Gembloux, Belgium

MID-IR MILK & Melamine → 250 ppm

NIR SOYAMEAL & Melamine → 1000 ppm
(60 samples at 0.1 = 4 misses)

Received: 18 January 2015 ■ *Revised: 2 May 2016* ■ *Accepted: 3 May 2016* ■ *Publication: 9 May 2016*

JNIRS JOURNAL OF NEAR INFRARED SPECTROSCOPY

Virtual Issue: Papers Presented at NIR-2015, October 2015, Foz do Iguassu, Brazil

LOCAL regression algorithm improves near infrared spectroscopy predictions when the target constituent evolves in breeding populations

F. Davrieux,[a,*] D. Dufour,[b,e] P. Dardenne,[c] J. Belalcazar,[d] M. Pizarro,[d] J. Luna,[d] L. Londoño,[e] A. Jaramillo,[e] T. Sanchez,[d] N. Morante,[d] F. Calle,[d] L.A. Becerra Lopez-Lavalle [d] and H. Ceballos[d]

carotenoids content of cassava roots

TCC values ranged from 0.11 µg g^{-1} to 29.0 µg g^{-1} (ppm)

>6000 samples and cross year validation

CONCLUSION LOD

When Y < 0.1 %, be sure what you analyze is (significantly) present in the spectra;
Look at the ratio absorptivity/noise

Analytes <0.1 % can be well predicted and provide useful results
But these analytes are likely correlated with something else into the matrix.

Hyperspectral Imaging can improve the LOD significantly
→ ppm ?

Conclusions

1. **Mixture model and multivariate space – Number of samples>>**
2. **Uncertainty**
3. **Limit of detection**

As Karl Norris and John Shenk recommendations

- Look at the spectra

- Make the models simple

- Challenge (validate) the models with new sets