- Poisson, C.; Jourdain, M. and Verrier, J.L.; Application of Near Infrared Spectroscopy to tobacco plant breeding : Accurate predictions of alkaloids, nitrogen, reducing sugars and ashes. Paper presented at the 41st Tobacco Workers' Conference, 2004. Seita, Altadis Group, Institut du Tabac, Domaine de la Tour, 24100 Bergerac, FRANCE. Telephone: 33(5).53.63.66.21. Fax: 33(5).53.63.66.08. E-mail: catherine.poisson@altadis.com
- The Bergerac Tobacco Institute uses Near Infrared Reflectance (NIR) Spectroscopy since 1999 for the screening of lines from tobacco plant breeding research programs.
- MPLS (Modified Partial Least Squares) predictive models, connecting NIR spectral data to chemical analysis results, were developed from tobacco powders. They allow the prediction of alkaloids, nitrogen, reducing sugars, ashes and powder dry matter from different tobacco types cultivated in France : Dark air-cured, Burley and Flue-cured. Equations were developed independently on each tobacco type resulting in 18 predictive models.
- The present calibration models were obtained using 2420 air-cured and flue-cured tobacco samples (crop years 1995 to 2000). 1756 spectra were included in calibration files whereas 664 spectra were used as model validation samples. The data thus include a great variability as far as spectral data and levels of chemical components are concerned : for example different varieties and stalk positions, year effect, including soil and weather effects.
- The models show a high degree of performance in comparison with classical chemical analyses.
- A PLS discriminant equation was also set up. It allows the classification of each sample in a group according to its NIR spectrum and the predictions with the appropriate models.
- This technology is fast, safe, simple to perform, non-destructive and inexpensive. It is perfectly well-adapted to tobacco breeding in order to improve characteristics implied in tobacco leaf chemical quality.
- The possibility of predicting, through this method, several other criteria is in prospect. Among those are other chemical components (i.e. related with aromatic properties), physical criteria (burning capacity and stem percentage) and smoke characteristics (tars, nicotine, biological activity…). (Reprinted with permission)
- Key terms: Near Infrared Reflectance Spectroscopy, tobacco, alkaloids, ashes, nitrogen, reducing sugars.

41st Tobacco Workers ' Conference, January 2004

Application of Near Infrared Spectroscopy to tobacco plant breeding : Accurate predictions of alkaloids, nitrogen, reducing sugars and ashes.



C. Poisson, M. Jourdain and J.-L. Verrier

Altadis
Institut du Tabac

Introduction

✓ **Use of Near Infrared Reflectance Spectrometry (NIRS) for the characterization of tobaccos issued from the Bergerac Tobacco Institute (ITB) breeding programs**

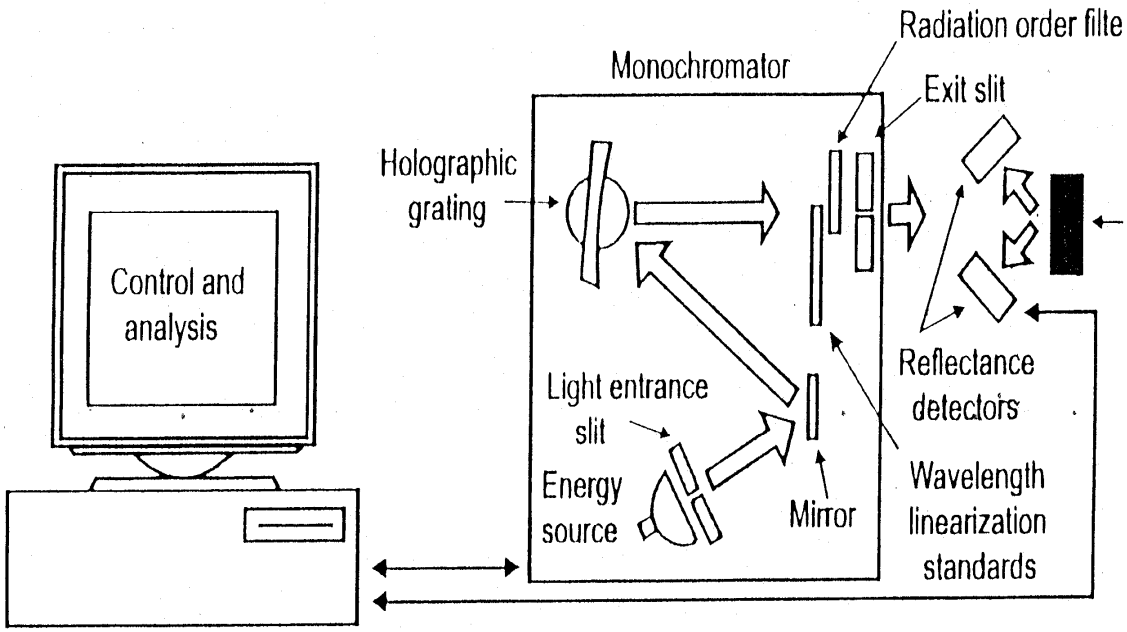
✓ **Chemical components routine-predicted since 1999** :

- **Total alkaloids (A),**
- **Ashes (Ash),**
- **Total Nitrogen (N),**
- **Total Reducing Sugars (TRS),**
- **Tobacco powder Dry Matter (DM).**

➢ **Model development and accuracy**

➢ **NIR spectrometry advantages**

The spectrophotometer



✓ **FOSS NIRSystems 6500 : from 400 to 2500 nm (1050 measure points, 2 nm steps)**

✓ **WinISI II spectral acquisition and calibration software**

Altadis

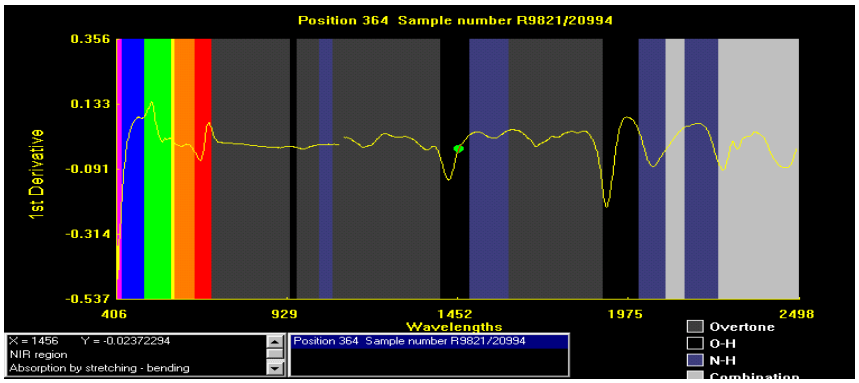
Principle

Tobacco powders : 500 µm

Spectra acquisition

Chemical analyses

log 1/R



λ : 400-2500 nm

variétés	famille	généa	Bloc ligne	N° éch	ALCA	CEN	Nt	MS
BAMM	S3	25-1-3	0101	02-23494	2,42	23,09	3,44	93,48
BAMM	S3	66-2-2	0102	02-23495	2,56	25,52	3,10	93,32
ITB 1000	temoin		0103	02-23496	3,08	20,54	3,64	93,20
BCCM	S3	14-2-2	0104	02-23497	0,70	25,60	3,07	92,89
BCCM	S3	14-2-3	0105	02-23498	1,83	23,28	4,08	94,41
BCCM	S3	20-1-1	0106	02-23499	1,48	23,40	4,14	93,61
BCCM	S3	20-1-6	0107	02-23500	1,57	22,47	4,35	93,68
ITB 1105	temoin		0108	02-23501	2,86	22,54	3,57	93,25
BG1C	S3	14-1-1	0109	02-23502	3,66	21,71	3,74	92,94
BG1C	S3	14-1-2	0110	02-23503	3,23	22,86	3,67	93,01
BG1C	S3	24-1-2	0111	02-23504	2,31	23,00	3,46	93,85
BG1C	S3	24-1-4	0112	02-23505	2,25	22,07	3,69	93,39

Data matrices

Setting up of the predictive models

Altadis

Spectral and chemical data

- ✓ Tobaccos cultivated in ITB from 1995 to 2000
- ✓ Lines and varieties from breeding programs
- ✓ 4 tobacco types :

  - ✓ 489 Flue-cured type 1 *F1*
  - ✓ 765 Flue-cured type 2 *F2*
  - ✓ 507 Dark air-cured *DK*
  - ✓ 659 Burley *BY*

- ➢ 2420 samples

  - spectral variability
  - important chemical component level ranges
  - year effects, varieties, stalk positions … integrated

- ✓ Achieved on the 4 tobacco types separately

Setting up of the predictive models : spectra mathematical treatments

SNV and Detrend 1-4-4-1

⇒ **reduction of particle size and packing density effects**

⇒ **baseline correction**

⇒ **increase of spectral differences between samples (better peak location)**

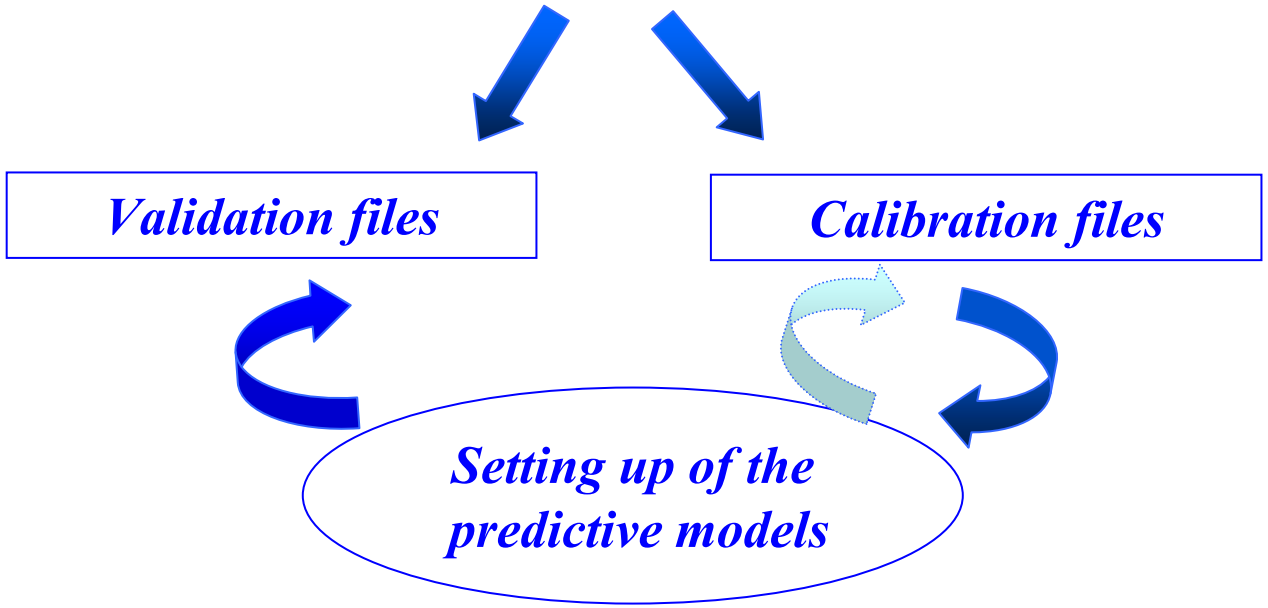
⇒ **noise reduction**

➢ **Models developed with 259 spectral data**

Setting up of the predictive models : spectra selection

- ***Sample classification according to their Mahalanobis H distance from the average spectrum of the file***

- ***« Elimination » of spectra with a H distance less than 0.6 from the neighbor spectra***



Altadis

Setting up of the predictive models

■ Validation and calibration files

		1995	1996	1997	1998	1999	2000	total
F1	calibration	37	79	105	56	40	16	333
	validation	17	41	56	23	11	8	156
F2	calibration	68	105	145	93	104	18	533
	validation	58	64	40	32	30	8	232
DK	calibration	68	84	57	77	69	20	375
	validation	13	21	51	21	18	8	132
BY	calibration	43	105	115	114	115	23	515
	validation	9	18	26	36	50	5	144

Setting up of the predictive models

- **MPLS predictive models (*Modified Partial Least Squares)***

  algorithm patented by InfraSoft International

  for each tobacco type and each chemical component :

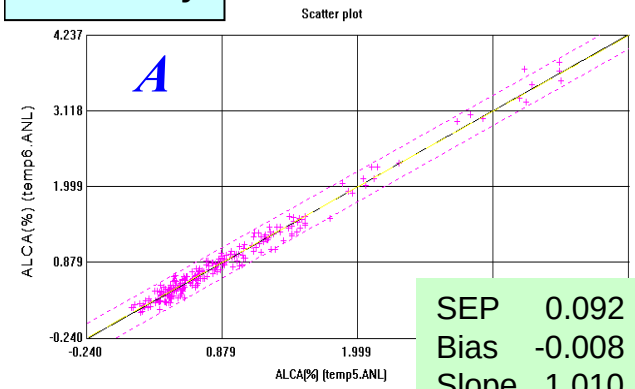
F1 ***F2***	**A, ASH, N, TRS and DM**
DK ***BY***	**A, ASH, N and DM**

➢ **18 predictive models**

Application of the models on validation files

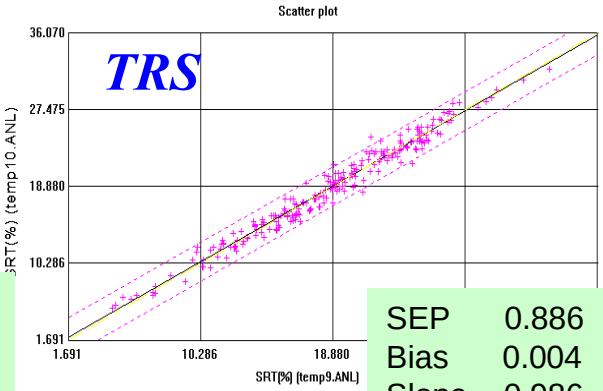
☺ F2 tobaccos

Laboratory

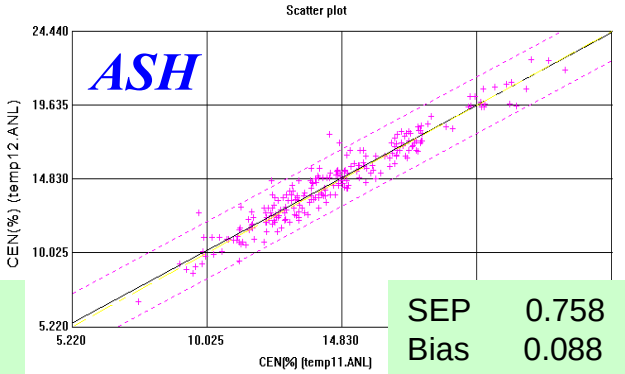


Predicted

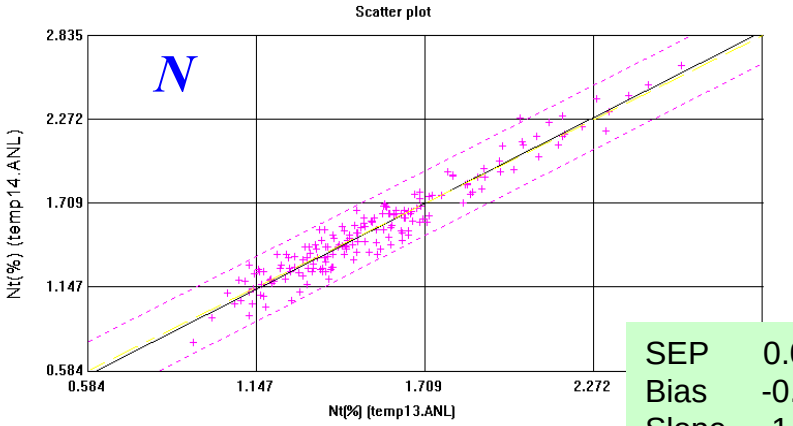
SEP	0.092
Bias	-0.008
Slope	1.010
RSQ	0.982



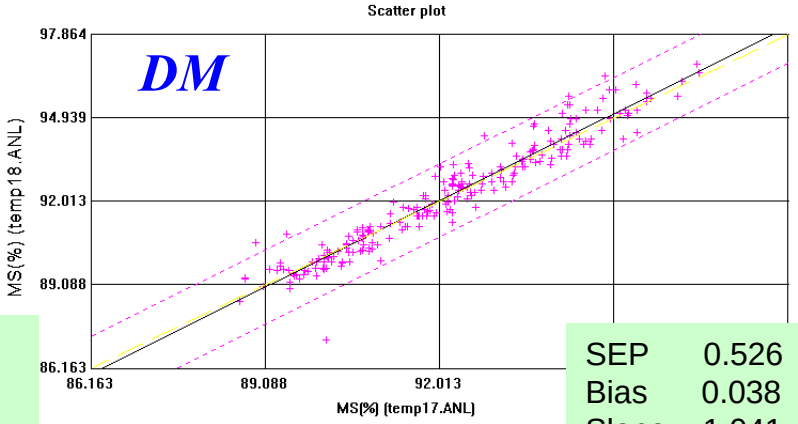
SEP	0.886
Bias	0.004
Slope	0.986
RSQ	0.973



SEP	0.758
Bias	0.088
Slope	0.983
RSQ	0.929



SEP	0.089
Bias	-0.004
Slope	1.026
RSQ	0.927



SEP	0.526
Bias	0.038
Slope	1.041
RSQ	0.929

➢ **Good performances : robustness and accuracy for each tobacco type**

Altadis

Performances of the models

		Spl nb	Average	SD	Est min	Est max	SEC	SECV	SEP	RSQ	1-VR
A (% DM)	F1	323	2,07	0,81	0	4,50	0,096	0,123	0,16	0,986	0,977
TRS (% DM)		319	18,82	6,61	0	38,66	0,996	1,148	1,32	0,977	0,970
ASH (% DM)		319	14,36	3,28	4,51	24,21	0,568	0,680	0,81	0,970	0,957
N (% DM)		215	1,98	0,45	0,62	3,34	0,076	0,094	0,12	0,971	0,958
DM (%)		288	92,00	1,66	87,02	96,98	0,375	0,467	0,58	0,949	0,921
A (% DM)	F2	517	1,01	0,65	0	2,96	0,080	0,094	0,11	0,985	0,979
TRS (% DM)		512	18,26	6,41	0	37,49	0,785	0,939	1,12	0,985	0,979
ASH (% DM)		518	14,59	3,18	5,05	24,12	0,626	0,704	0,79	0,961	0,951
N (% DM)		396	1,62	0,38	0,48	2,75	0,074	0,082	0,09	0,961	0,953
DM (%)		455	92,87	1,89	87,18	98,55	0,388	0,437	0,49	0,958	0,947
A (% DM)	DK	376	2,73	1,04	0	5,86	0,135	0,165	0,20	0,983	0,975
ASH (% DM)		375	22,22	2,66	14,20	30,21	0,694	0,816	0,96	0,932	0,906
N (% DM)		353	3,20	0,65	1,25	5,16	0,112	0,137	0,17	0,970	0,956
DM (%)		380	91,82	1,48	87,38	96,26	0,263	0,344	0,45	0,968	0,946
A (% DM)	BY	488	2,74	1,07	0	5,96	0,144	0,162	0,18	0,982	0,977
ASH (% DM)		497	23,53	2,94	14,70	32,36	0,738	0,830	0,94	0,937	0,920
N (% DM)		467	3,68	0,72	1,53	5,83	0,134	0,163	0,20	0,965	0,948
DM (%)		488	92,49	1,41	88,27	96,71	0,356	0,396	0,44	0,936	0,921

Satisfying Standard Errors of Calibration and of Cross Validation

High degree of performance in comparison with classical chemical analysis

Coefficients of determination > 0.9

Altadis

Annual controls

■ **Annual enrichment is necessary to include maximum variability among samples used for calibrations**

■ **Classical chemical analyses performed on a sample set from new year trials (2000, 2001, 2002)**

➢ ***Confirmation of the performances of the models***

➢ ***No new enrichment necessary with 2001 and 2002 samples :***

- satisfying predictions

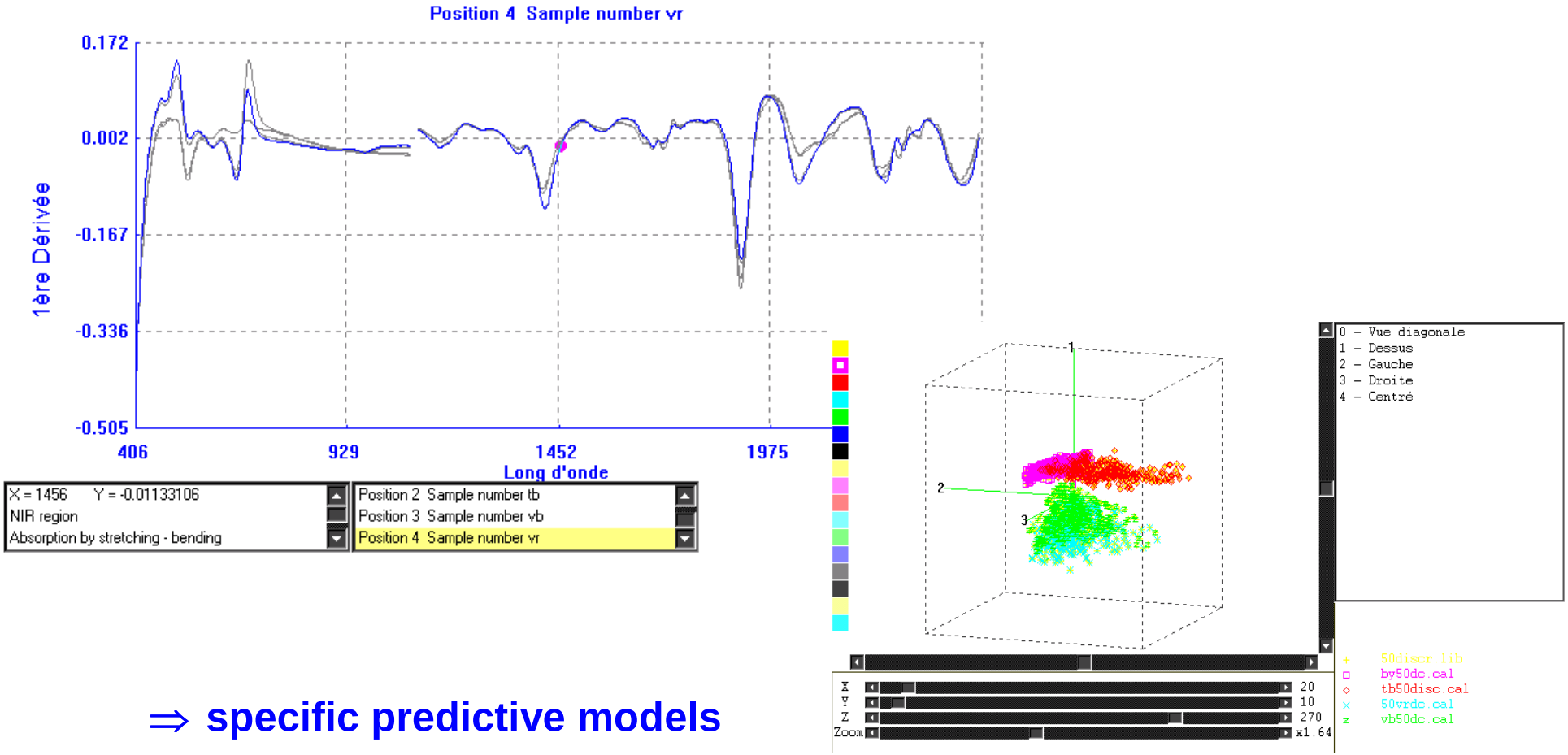
- spectra integrated in the initial spectral population

- predicted values integrated in the range of values observed in the calibration files

➢ ***Annual controls to be continued***

Setting up of the discriminant equation

- **Spectral profiles different according to tobacco types**

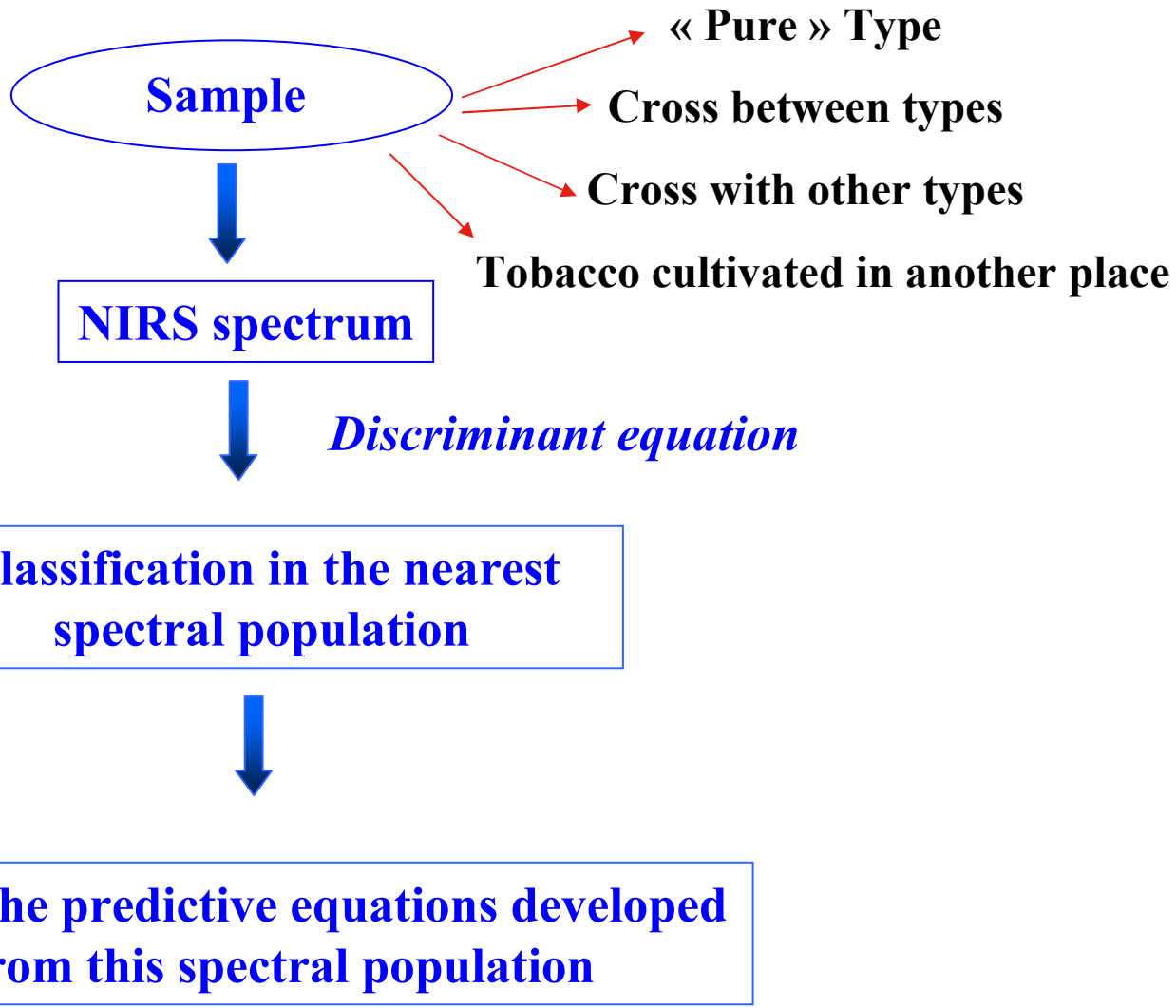


⇒ **specific predictive models**

Altadis

Setting up of the discriminant equation

- Objective



Altadis

Setting up of the discriminant equation

- **Set up**
  - ➔ **Partial Least Squares Discriminant regression (PLSD)**
  - ➔ **from the 4 calibration files : 1756 spectra**

- **Application**
  - ➔ **664 spectra from the validation files**

➢ ***100 % Dark air-cured and Burley correctly classified***

➢ ***91 % Flue-cured type 1***

➢ ***98 % Flue-cured type 2***

« Misclassification » due to inversion between the 2 Flue-cured types

Altadis

Conclusion

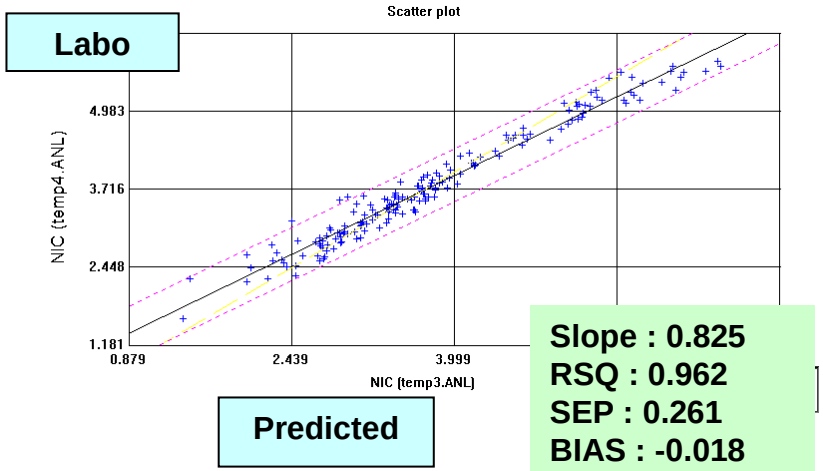
- **Potential of the NIRS technology for the prediction, with an error close to the laboratory error, of total alkaloids, ashes, total nitrogen and total reducing sugars levels in tobacco.**

- **Advantages of the technique :**
  - ➔ **fast (2 mn / scan) and multiplicity of predictions from one sample scan : rapid results from a sample set,**
  - ➔ **simple to perform (sample preparation),**
  - ➔ **non-destructive,**
  - ➔ **safe (no extraction),**
  - ➔ **inexpensive.**

- **Perfectly well-adapted to tobacco breeding in order to improve characteristics implied in tobacco leaf chemical quality.**

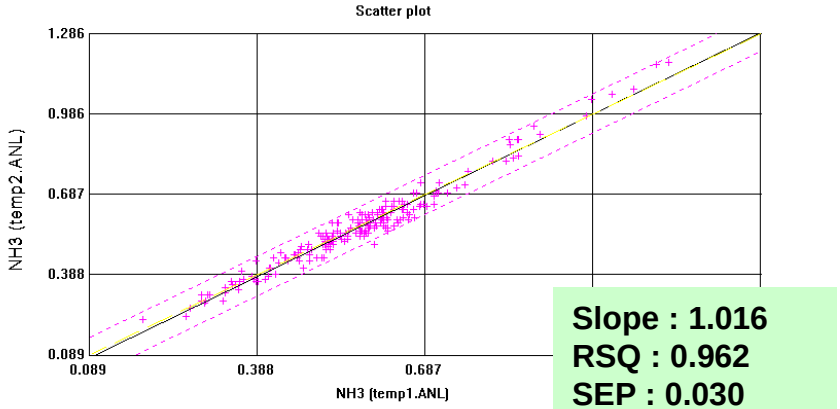
Prospects

- **To increase the number of predicted criteria**
  - ***Other chemical criteria : nicotine, aromatic components, NH3 ...***



Slope : 0.825
RSQ : 0.962
SEP : 0.261
BIAS : -0.018

Nicotine prediction from 2001 Burley samples



Slope : 1.016
RSQ : 0.962
SEP : 0.030
BIAS : -0.005

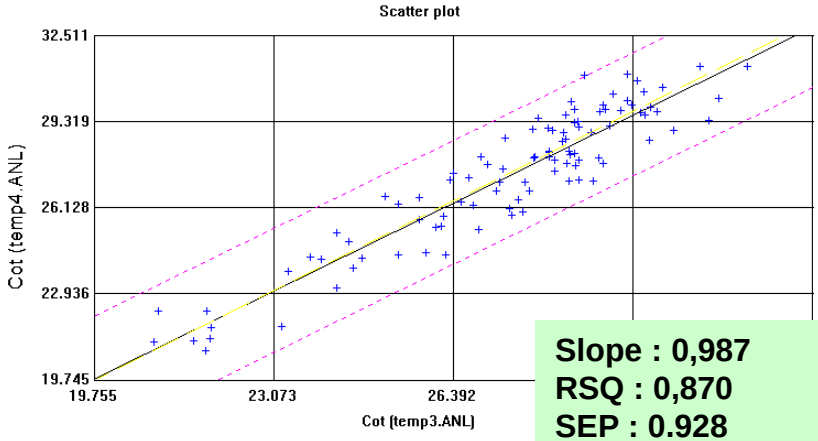
NH_3 prediction from 2000- 2002 Burley samples

Altadis

Prospects

- **To increase the number of predicted criteria**

➢ ***Physical criteria : burning capacity, stem percentage***



Slope : 0,987
RSQ : 0,870
SEP : 0.928
BIAS : -0.076

% stem prediction from 2000- 2001 Burley samples

➢ ***Smoke characteristics : tar, nicotine, biological activity ...***

Altadis