Altadis

Applications of **Near infrared reflectance spectroscopy** to quantitative and qualitative characteristics of tobacco blends

Gilles FIGUÈRES
Altadis, Research Centre, Fleury Les Aubrais

"6th Bergerac Tobacco Scientific Meeting"
1st - 2nd September 2005

Altadis

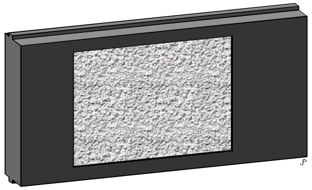
Spectral profile

- Near infrared spectra can be considered as multivariate fingerprints of analysed samples
- The spectrum is affected by:
  - Non-specific light scattering from surface of sample
  - Chemical composition of sample
    - CH, OH, NH molecular bonds with specific vibrational frequency to describe spectral absorption information

A reflectance spectrum = a measurement of total chemical and physical composition of material

Visible and near-infrared diffuse reflectance analysis

- **Dried and ground tobacco samples or “as is” cut tobacco from products scanned by using a nir grating spectrometer (Versatile Foss NIRSystems)**
- **Data collected at 2 nm intervals**
  - wavelength range 400 to 2500 nm
  - raw spectra stored as (log 1/reflectance):
    - 1050 data points collected / sample
    - average of duplicate scans

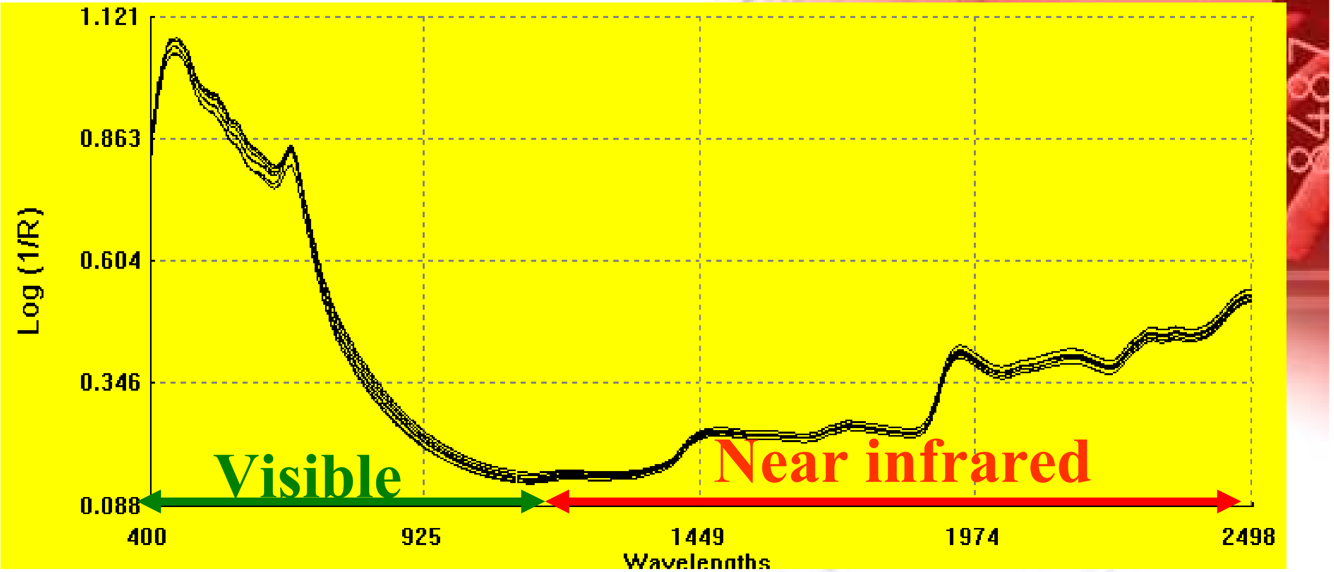


Altadis

Tobacco spectral profile

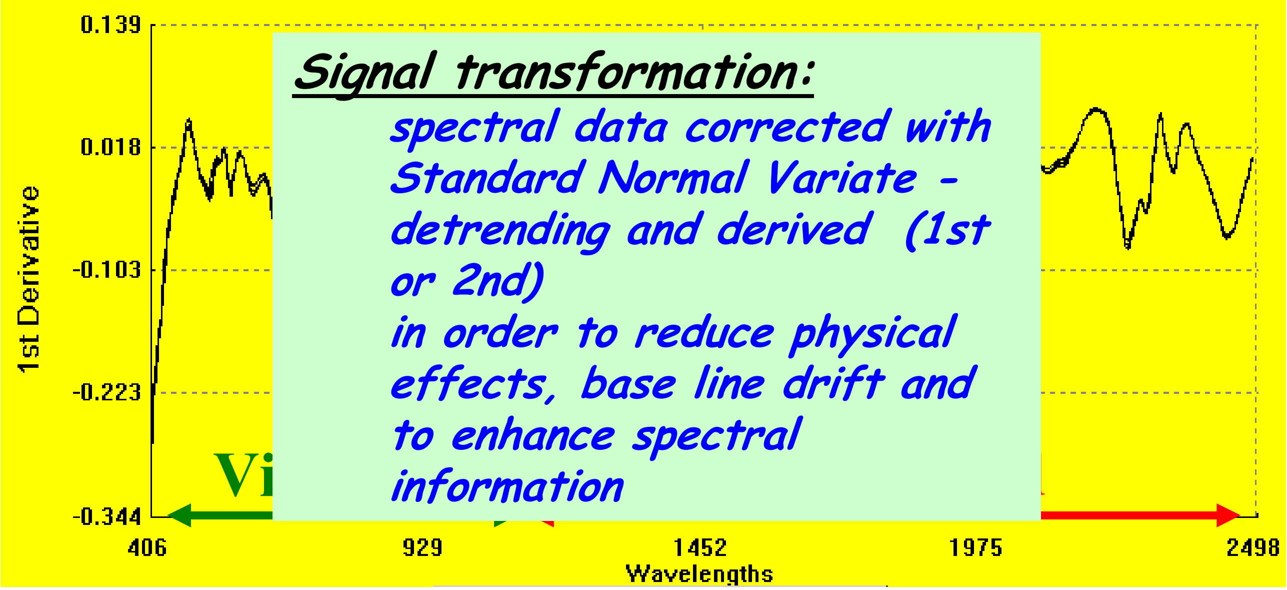
Raw spectra =

Fingerprints of physical properties of tobacco samples (size and distribution of particles)



Pretreated spectra =

Fingerprints of chemical properties of tobacco samples



Signal transformation:

spectral data corrected with Standard Normal Variate - detrending and derived (1st or 2nd) in order to reduce physical effects, base line drift and to enhance spectral information

Altadis

NIRS applications

- **Quantitative approach**: prediction of tobacco chemical composition
  - **Multivariate regression methods:** as Partial Least Squares
    - development of calibration models where spectral response from a set of samples is correlated with known chemical concentration from laboratory analyses in order to try to replace conventional chemical methods (time consuming, laborious and costly procedures)
- **Qualitative approach**: using only tobacco spectral data
  - **Descriptive or comparative method**:
    - Principal Component Analysis (PCA) for ex.
  - **Discriminant method (or pattern recognition method):**
    - Classification of tobacco products: in order to classify an unknown material on the basis of its spectral features without the need for chemical data

QUANTITATIVE APPROACH

First application

Near infrared spectroscopy in prediction of chemical characteristics of raw tobacco

Objective

➔Assessment of potential of near-infrared reflectance spectroscopy as alternative analytical method for simultaneous determination of chemical compound contents in tobacco deliveries :

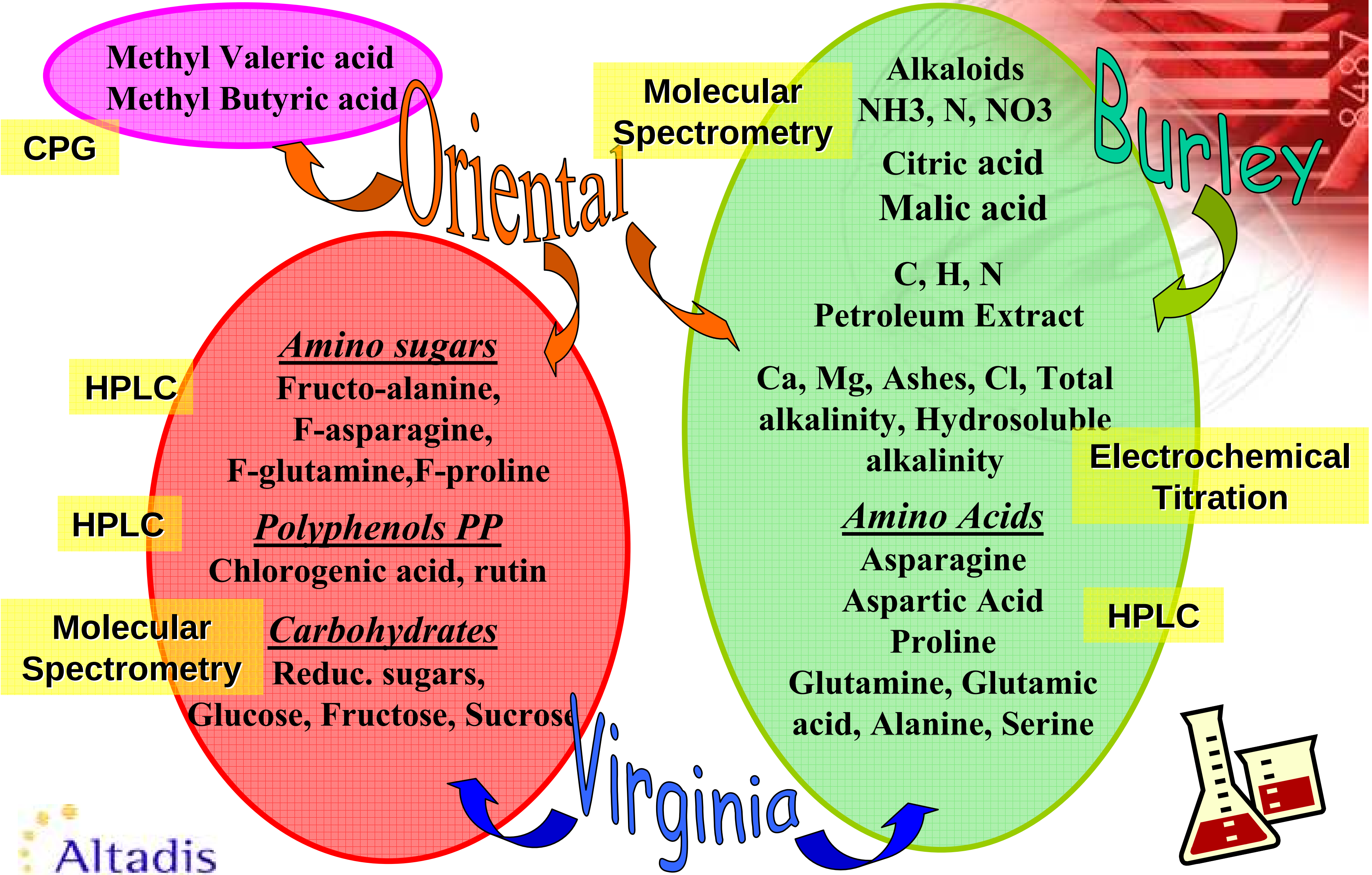
- **develop statistical multivariate calibrations between chemical reference values and spectral data of tobacco**
- **evaluate performances of prediction models**

Studied materials

- Tobacco types - origin grades: data base
  - Virginia : 1100 lots
  - Burley: 560 lots
  - Oriental: 300 lots
  - Dark: 960 lots
  - Recons (70 lots) and Virginia stems (50 lots)

- Seven delivery years of tobacco : 1996 to 2002
- Specific analytical and statistical approach per each tobacco category

Tobacco chemical criteria



Methyl Valeric acid
Methyl Butyric acid

CPG

Molecular Spectrometry

Alkaloids
NH3, N, NO3
Citric acid
Malic acid

C, H, N
Petroleum Extract

Ca, Mg, Ashes, Cl, Total alkalinity, Hydrosoluble alkalinity

Electrochemical Titration

Amino Acids
Asparagine
Aspartic Acid
Proline
Glutamine, Glutamic acid, Alanine, Serine

HPLC

HPLC

Amino sugars
Fructo-alanine,
F-asparagine,
F-glutamine,F-proline

HPLC

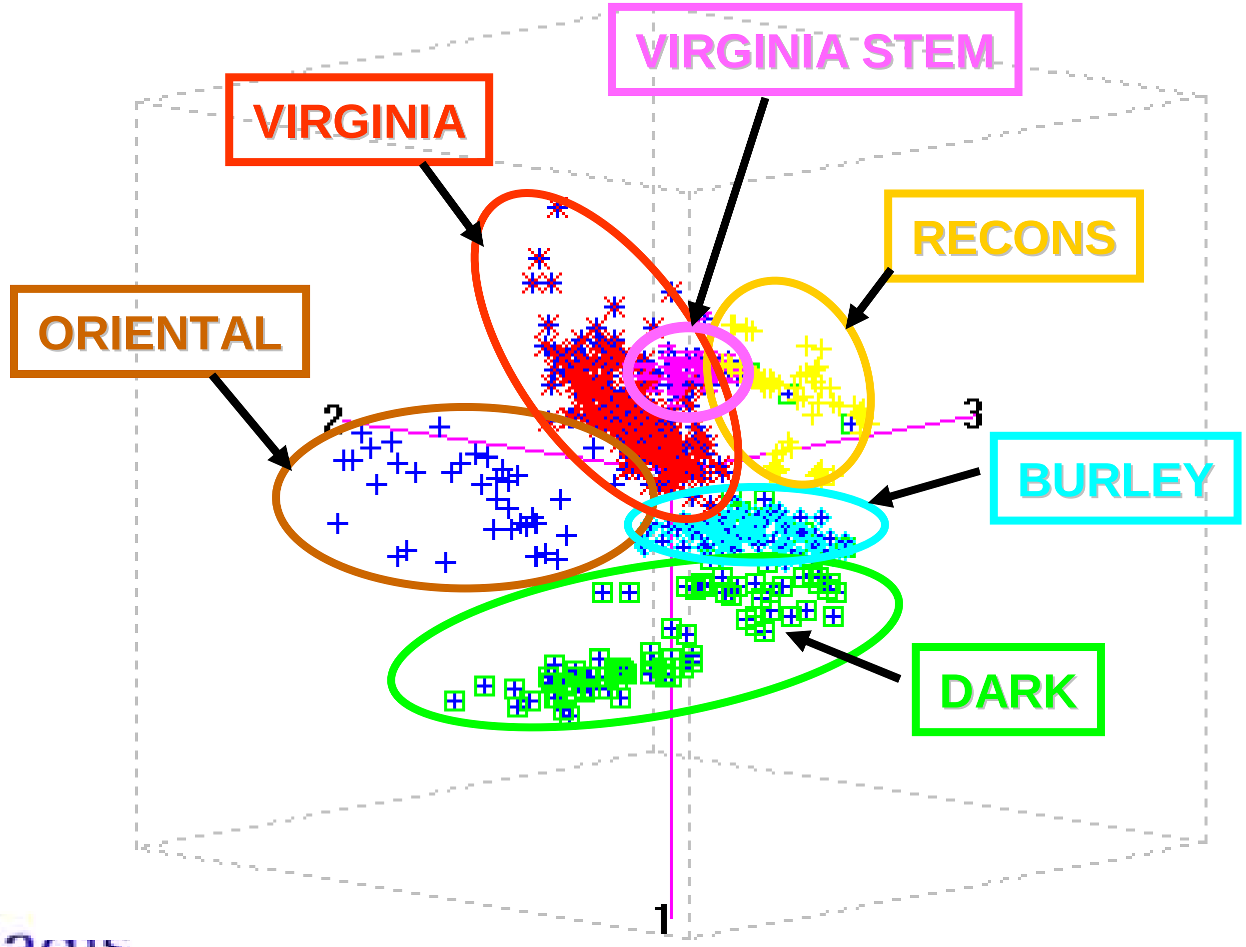
Polyphenols PP
Chlorogenic acid, rutin

Molecular Spectrometry

Carbohydrates
Reduc. sugars,
Glucose, Fructose, Sucrose

Altadis

Score plots of the three first principal components space for spectral tobacco sample types



Mathematical treatments

Spectral signal transformations

+ regression calibrations

		T1	T2	T3	T4	T5
Regression method		MPLS	PLS	PCR	MPLS	MPLS
Maxi nb of MLR terms		16	16	16	5	16
Spectral data pretreatment	derivation-smoothing	1.4.4	1.4.4	1.4.4	1.4.4	2.8.6
	scatter correction	SNV D	SNV D	SNV D	SNV D	SNV D
Omitting samples cycles		2	2	2	2	2
Spectra range	400 - 2500 nm	VIS+NIR	VIS+NIR	VIS+NIR	VIS+NIR	VIS+NIR
	1100 - 2500 nm	NIR	NIR	NIR	NIR	NIR

		T6	T7	T8	T9	T10	T11
Regression method		MPLS	MPLS	MPLS	MPLS	MPLS	MPLS
Maxi nb of MLR terms		16	16	16	16	16	16
Spectral data pretreatment	derivation-smoothing	3.10.10	4.10.10	0.1.1	1.4.4	1.4.4	1.4.4
	scatter correction	SNV D	SNV D	SNV D	SNV	DETREND	MSC
Omitting samples cycles		2	2	2	2	2	2
Spectra range	400 - 2500 nm	VIS+NIR	VIS+NIR	VIS+NIR	VIS+NIR	VIS+NIR	VIS+NIR
	1100 - 2500 nm	NIR	NIR	NIR	NIR	NIR	NIR

Altadis

Comparison of 11 combinations × 2 spectral segments

Performance of calibration models
Test Samples of 2003 Oriental tobacco deliveries

n = 39

				GH = 1.82 NH = 1.18		GH = 1.39 NH = 0.73	
				400 - 2500 nm		1100 - 2500 nm	
	Chemical compound	Content range		RMSEP	R^2	RMSEP	R^2
	Volatile acids ppm						
1	Methyl Valeric	375	3707	241	0.903	228	0.914
2	Methyl Butyric	367	1738	130	0.853	117	0.879
	Carbohydrates %						
3	Glucose	1.45	6.13	0.40	0.904	0.40	0.927
4	Fructose	2.31	8.88	0.28	0.958	0.25	0.968
5	Sucrose	0.00	2.43	0.35	0.717	0.49	0.586
6	GFS	4.11	16.77	0.61	0.957	0.52	0.972
7	Red. Sugars	5.30	18.39	0.67	0.951	0.62	0.961
	Nitrogen compounds %						
8	Alkaloids	0.67	3.02	0.10	0.978	0.09	0.983
9	Nt	1.94	3.70	0.14	0.899	0.15	0.899
10	NH3	0.06	0.42	0.02	0.952	0.02	0.961
11	NO3	0.04	1.00	0.13	0.68	0.20	0.792
	Amino acids %						
12	Aspartic	0.03	0.17	0.03	0.872	0.03	0.875
13	Asparagine	0.18	1.42	0.08	0.931	0.08	0.931
14	Proline	0.84	1.97	0.12	0.735	0.15	0.638
15	Glutamic	0.03	0.12	0.02	0.746	0.02	0.713
16	Glutamine	0.07	0.41	0.04	0.819	0.05	0.746
17	Serine	0.01	0.15	0.01	0.827	0.01	0.768
18	Alanine	0.07	0.14	0.02	0.168	0.02	0.265

Altadis

Performance of calibration models
Test Samples of 2003 Oriental tobacco deliveries

GH = 1.82 NH = 1.18 (400 - 2500 nm)

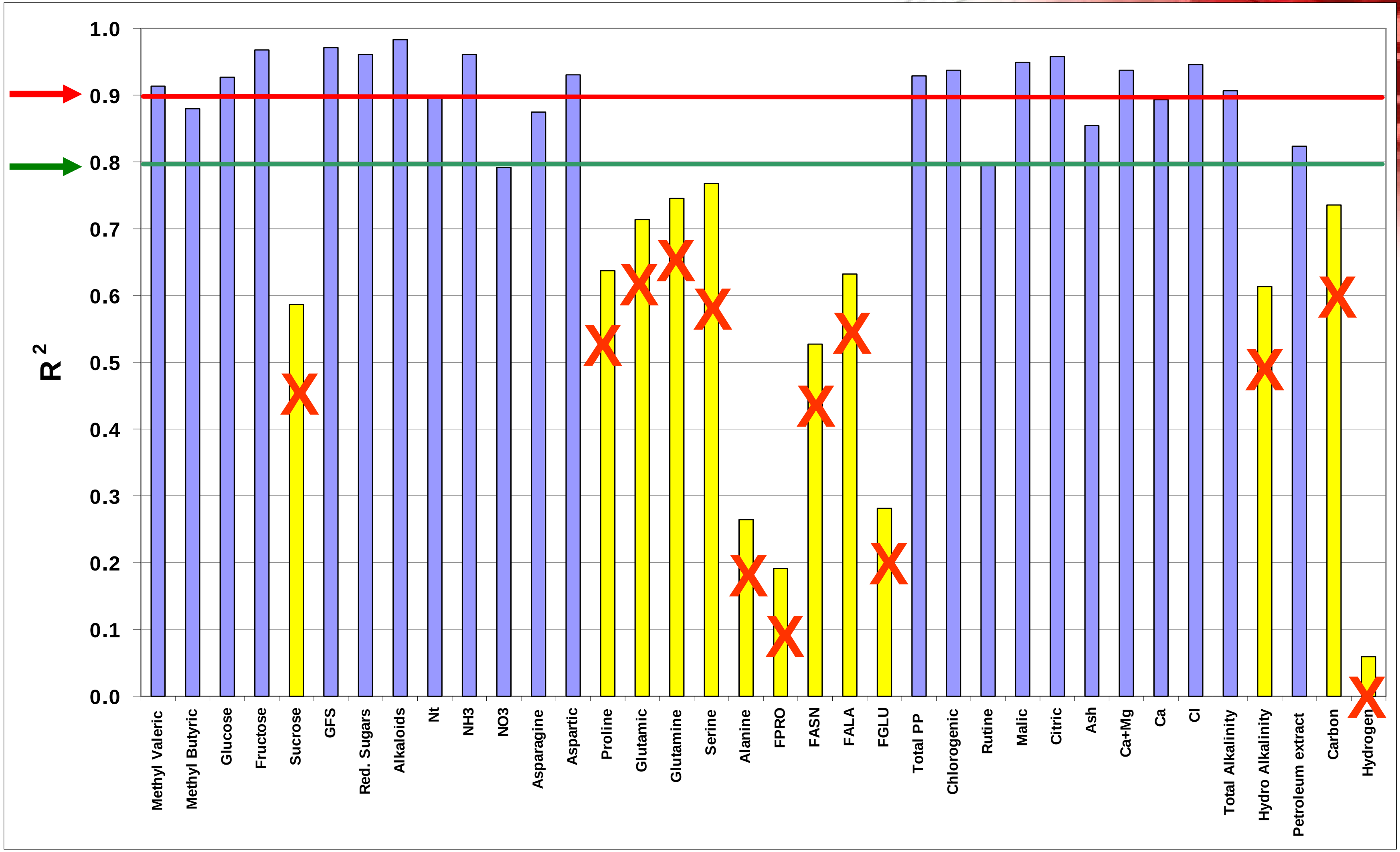
GH = 1.39 NH = 0.73 (1100 - 2500 nm)

n = 39

	Chemical compound	Content range		400 - 2500 nm		1100 - 2500 nm	
				RMSEP	R^2	RMSEP	R^2
	Amino sugars %						
19	FPRO	0.22	0.89	0.23	0.175	0.19	0.192
20	FASN	0.14	0.49	0.05	0.544	0.05	0.527
21	FALA	0.09	0.37	0.09	0.349	0..049	0.633
22	FGLU	0.03	2.37	0.36	0.342	0.37	0.282
	Polyphenols %						
23	Total PP	0.84	2.83	0.13	0.954	0.17	0.928
24	Chlorogenic	0.42	1.60	0.77	0.971	0.11	0.938
25	Rutine	0.32	1.20	0.11	0.748	0.10	0.796
	Organic acids %						
26	Malic	2.64	8.64	0.34	0.952	0.35	0.949
27	Citric	0.53	3.18	0.15	0.958	0.17	0.957
	Inorganics/Organics						
28	Ash %	11.10	21.10	1.03	0.807	0.89	0.854
29	Ca+Mg meq%	137.00	276.00	10.31	0.922	9.53	0.938
30	Ca %	1.83	4.52	0.21	0.886	0.20	0.894
31	Cl %	0.26	1.83	0.08	0.945	0.09	0.946
32	Total Alkalinity meq%	144.0	309.0	13.0	0.892	12.3	0.906
33	Hydro Alkalinity meq%	0.0	42.0	8.8	0.584	8.5	0.614
	Miscellaneous						
34	Petroleum extract %	2.60	8.40	0.53	0.798	0.50	0.823
35	Carbon %	39.65	44.91	0.83	0.697	0.77	0.735
36	Hydrogen %	3.75	5.61	0.39	0.067	0.41	0.06

Altadis

Performance of calibration models
Test Samples of 2003 Oriental tobacco deliveries



QUANTITATIVE APPROACH

Second application

Near infrared spectroscopy in prediction of tobacco blend additives

Example of Humectants

Altadis

Humectant quantification

1.2 propylene glycol (PRGL) and glycerol (GLY)

Principle of Humectant Reference Analysis

A methanol extract of the tobacco sample is prepared and PRGL and GLY are determined in that extract by quantitative gas chromatography.

Results are reported as percent (m/m)

Calibration statistics for humectants

	GROUND SAMPLE				CUT TOBACCO			
	VIS+NIR		NIR		VIS+NIR		NIR	
	GLY	PRGL	GLY	PRGL	GLY	PRGL	GLY	PRGL
	Calibration performance values				Calibration performance values			
N	69	69	69	69	68	68	68	68
Mean%	1.5654	1.1909	1.5654	0.1909	1.5616	1.2071	1.5616	1.2071
SEC%	0.1931	0.1404	0.1862	0.1162	0.2261	0.1275	0.2013	0.1417
R^2 cal	0.9774	0.9824	0.979	0.9879	0.9695	0.9855	0.9758	0.9821
SECV%	0.3334	0.2387	0.2763	0.1842	0.421	0.2701	0.3136	0.2346
R^2CV	0.9331	0.9495	0.954	0.97	0.8948	0.9361	0.9416	0.9518
	Validation performance values				Validation performance values			
Slope	0.969	0.997	0.972	0.998	1.064	1.019	1.044	0.982
R2	0.904	0.968	0.947	0.986	0.901	0.979	0.966	0.975
SEP%	0.276	0.149	0.207	0.1	0.287	0.121	0.175	0.136
SEP©%	0.282	0.15	0.21	0.101	0.292	0.124	0.172	0.137
	average NH = 0.135 average GH = 0.771		average NH = 0.079 average GH = 0.620		average NH = 0.213 average GH =0.754		average NH = 0.140 average GH =0.578	

Standard error of reference method:

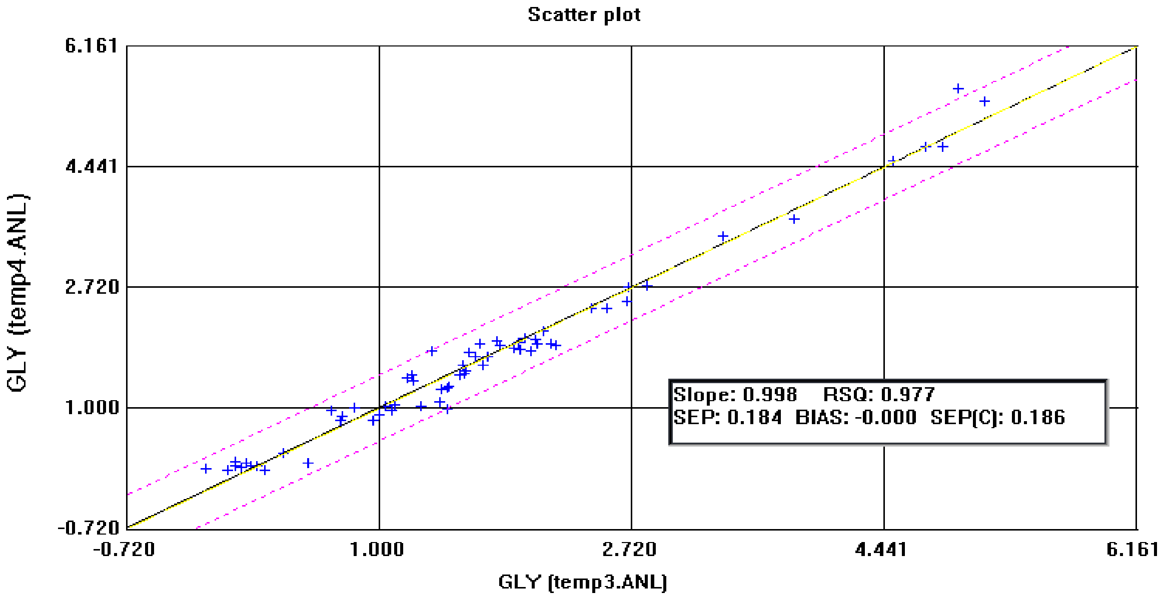
Glycérol = 0.23%, PRGL = 0.08%

Altadis

Chemical values vs. predicted NIRS values

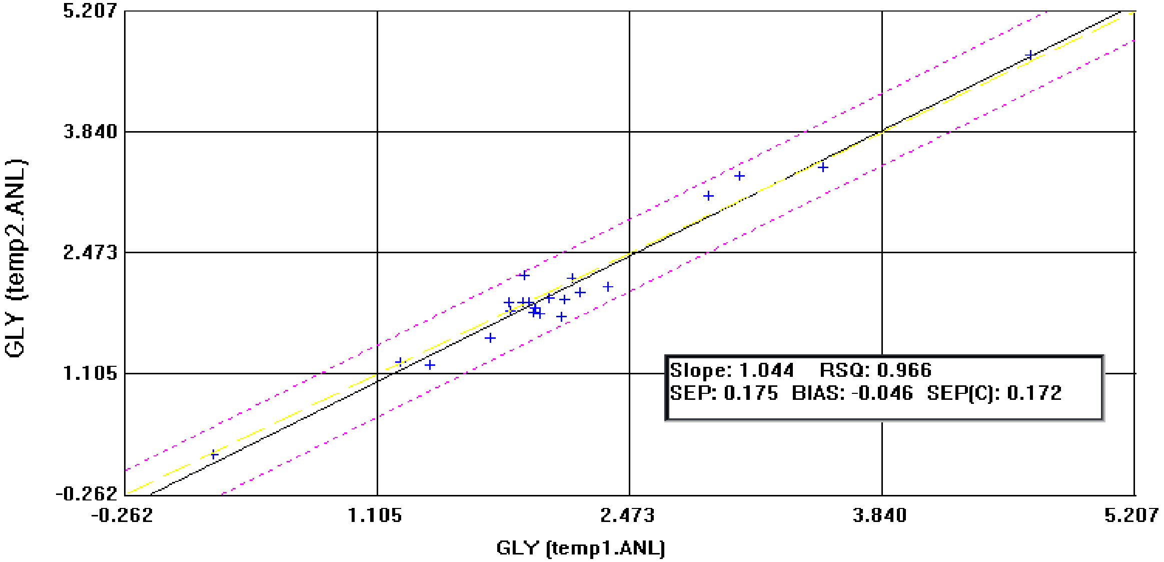
Application to cut tobacco of nir wavelength range model

Calibration set



GLYCEROL

Validation set

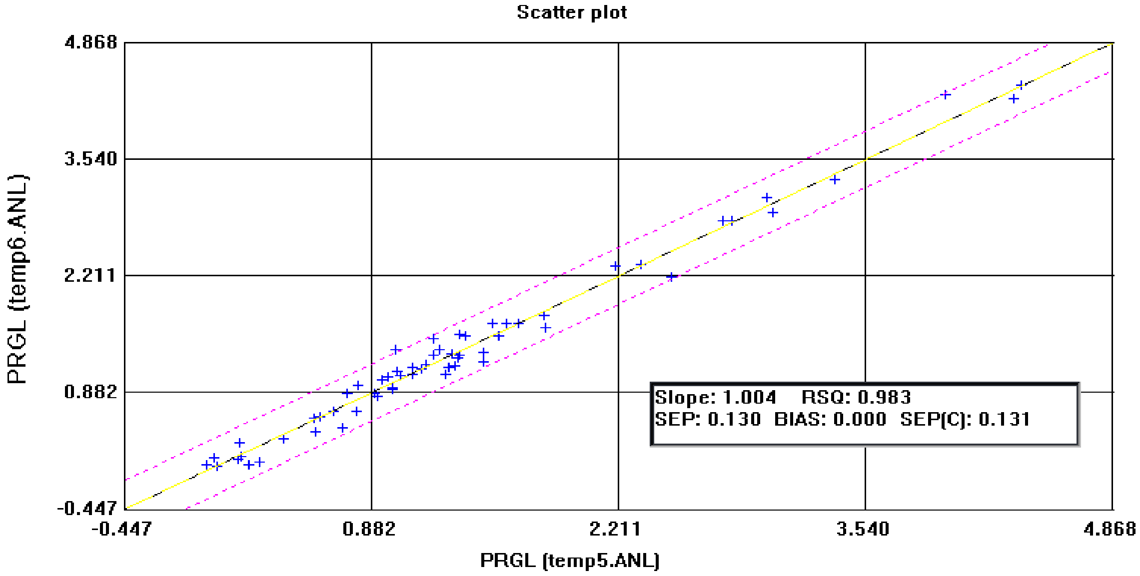


Altadis

Chemical values vs. predicted NIRS values

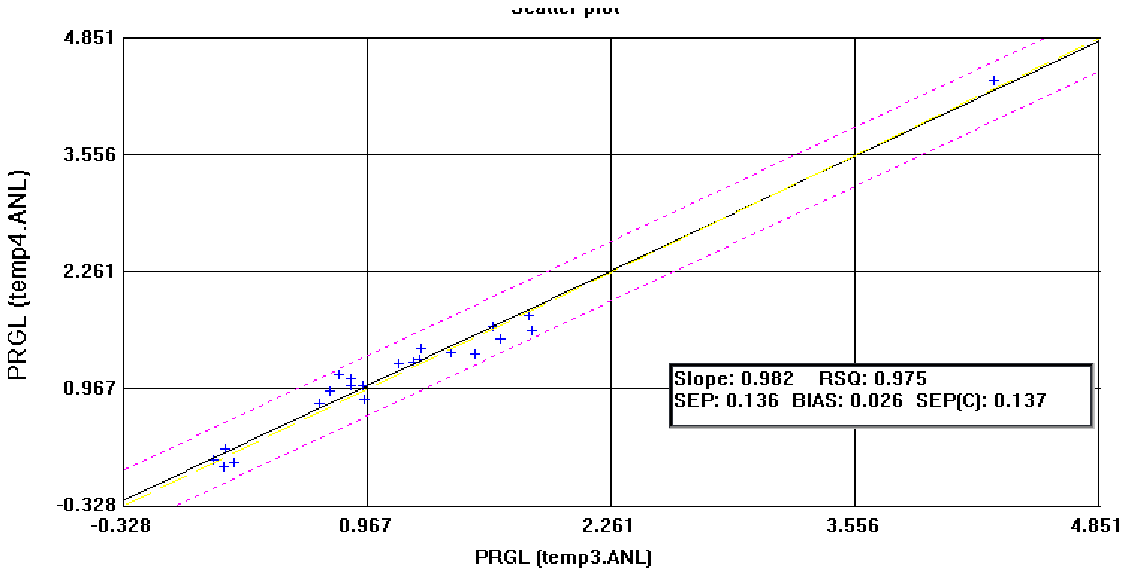
Application to cut tobacco of nir wavelength range model

Calibration set



1.2 PROPYLENE GLYCOL

Validation set



Altadis

QUALITATIVE APPROACH

First application

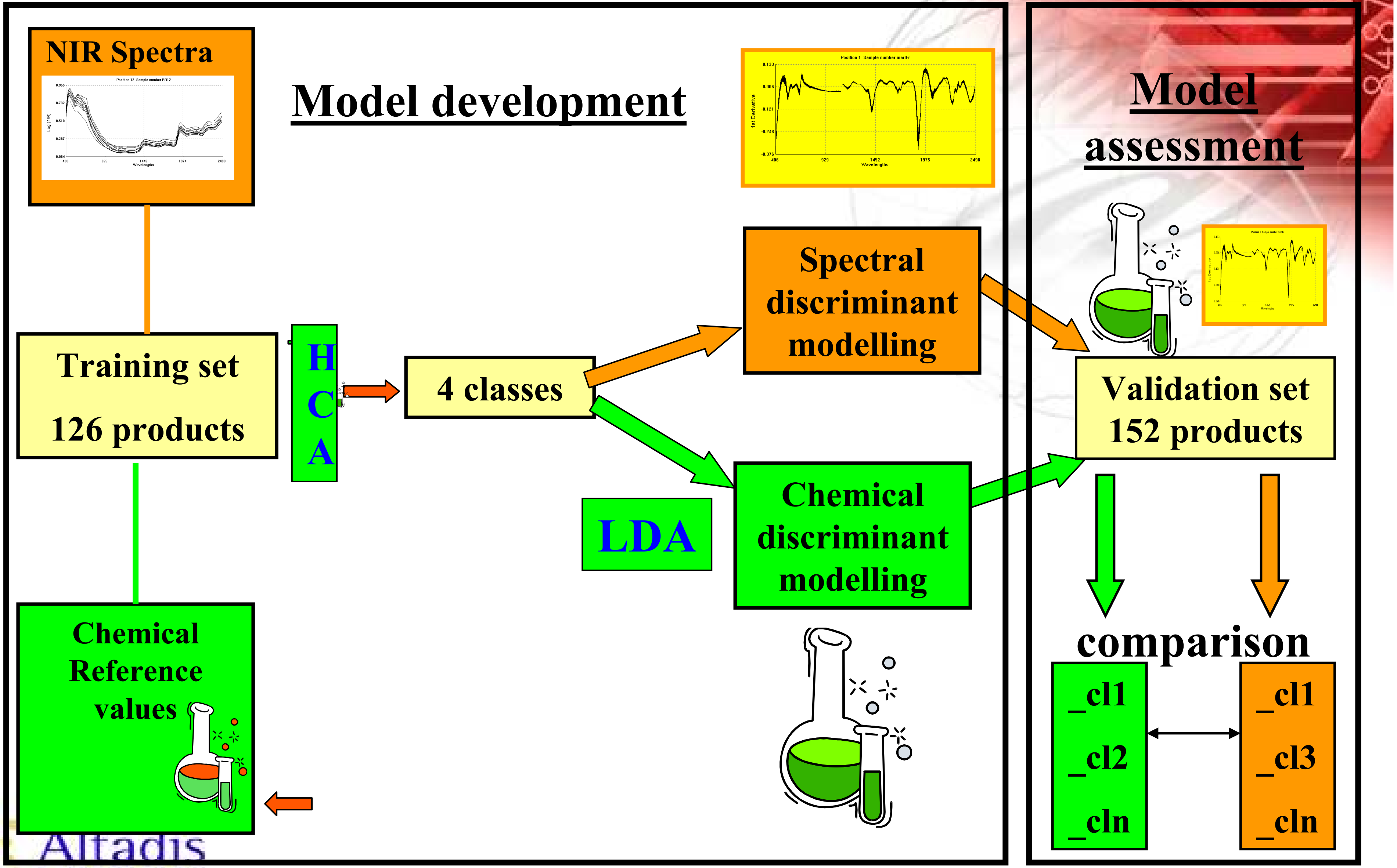
Near infrared spectroscopy and pattern recognition methods for classification of commercial tobacco blends

Objective

- Evaluate potential of near infrared spectroscopy to be a qualitative tool *to classify tobacco blends of commercial products* on basis of spectral features:
  - apply different supervised pattern recognition methods to NIR spectra of finished products (a priori knowledge about category membership of samples)
  - compare the performance of these multivariate data models

Altadis

Chimiometric procedure



NIR Spectra

Model development

Training set

126 products

H
C
A

4 classes

Spectral discriminant modelling

LDA

Chemical discriminant modelling

Chemical Reference values

Model assessment

Validation set
152 products

comparison

_cl1	_cl1
_cl2	_cl3
_cln	_cln

Altadis

Chemical measurements of tobacco blends: 26 variables

Tobacco markers

Burley, FC, Orientals

Amino Acids

ASPARAGINE
ASPARTIC Acid
PROLINE

Alkaloids

Sugars

G, F, S

NH_3 , PO_4 , N

Methyl Valeric ac.
Methyl Butyric ac

CITRIC acid
MALIC acid

Polyphenols

Inorganics

Ca, K
NO_3
Total alkalinity

Additive markers

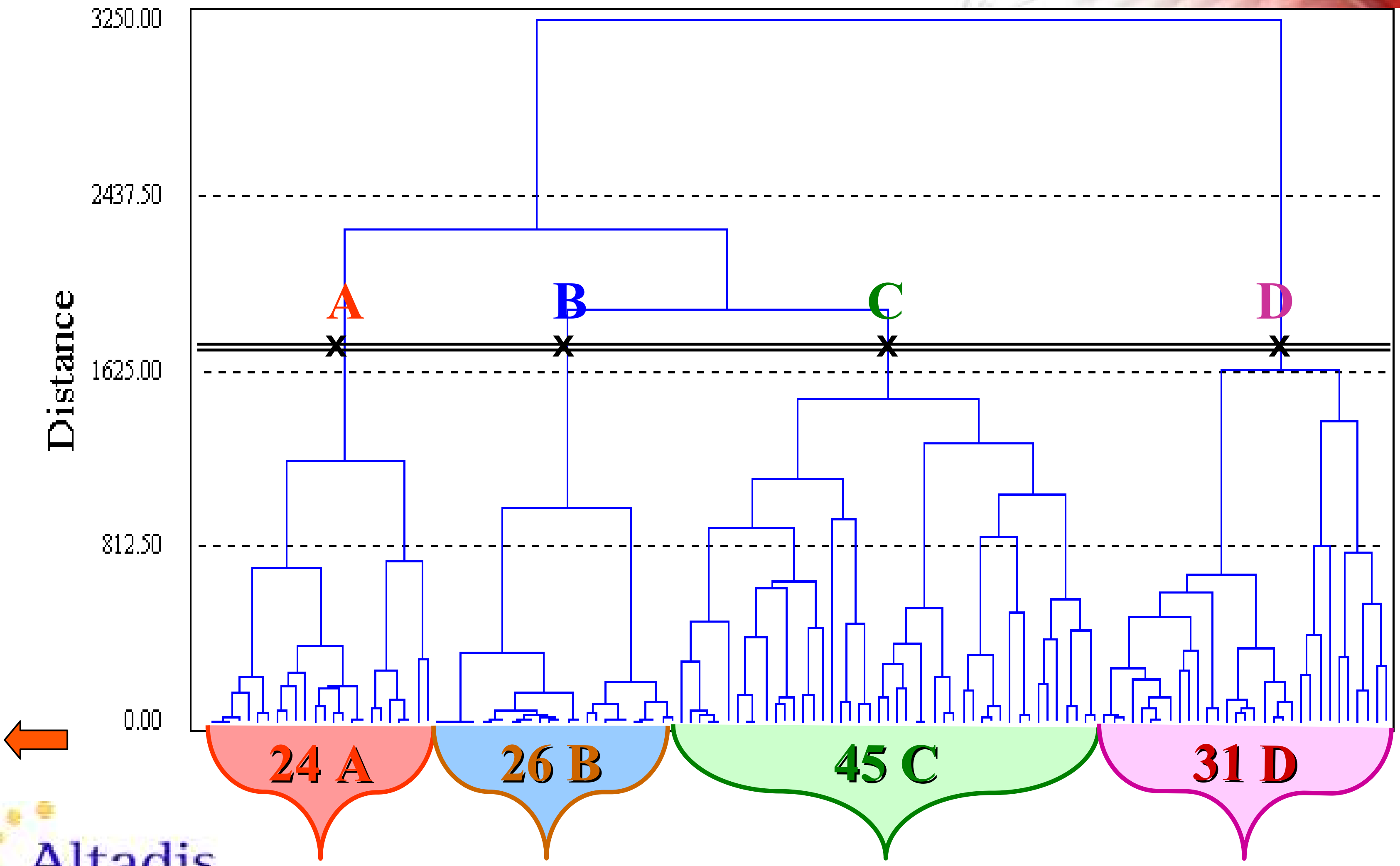
Cacao (theobromin
Liquorice glycyrrhizic ac.
Propy. glycol, Glycerol

Process markers

Deoxyfructosazines
(DFR25, DFR26)
Glucosamine

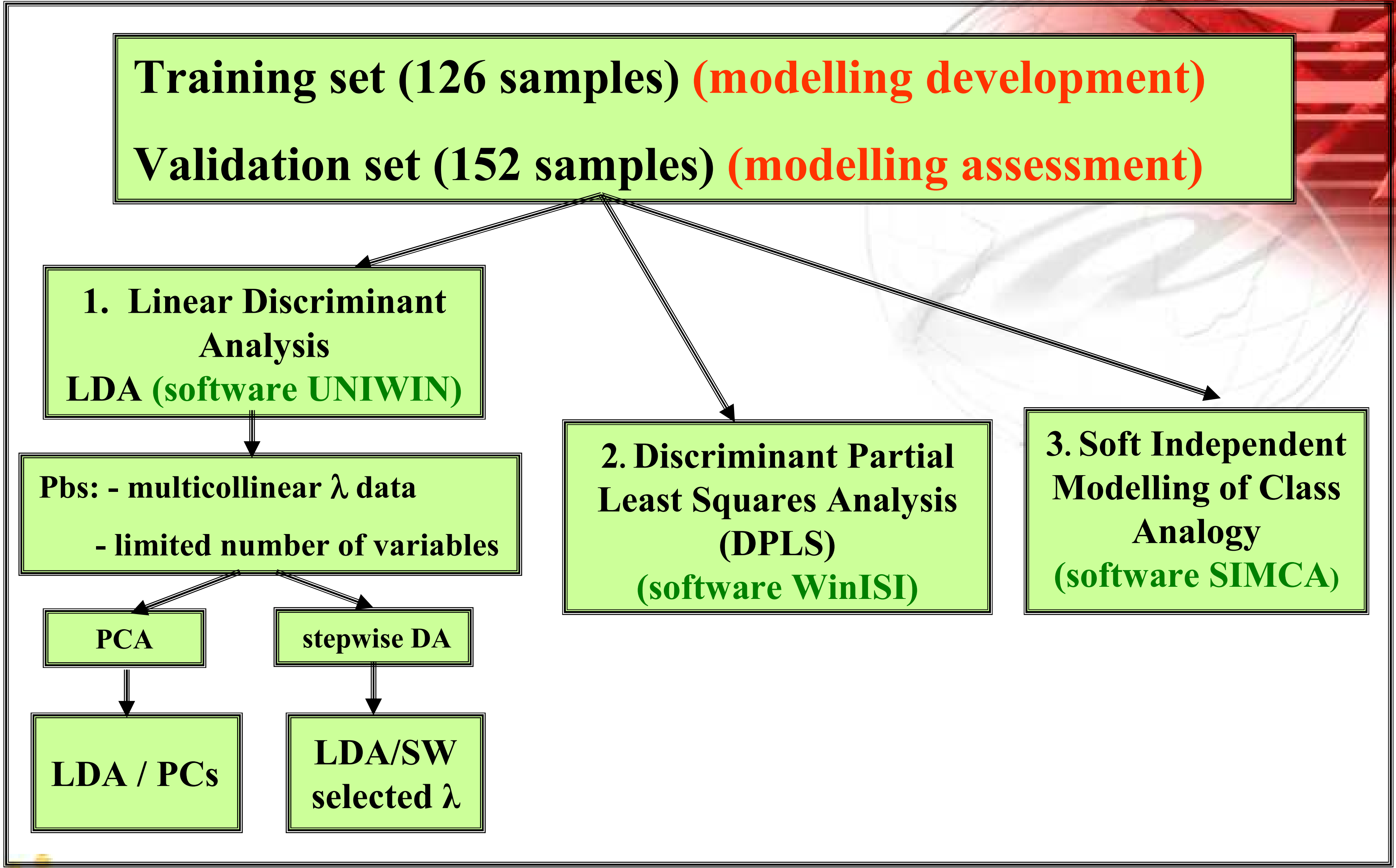
Hierarchical cluster analysis

Training set : 126 tobacco blends / 26 chemical variables



Altadis

Classification methods



Training set (126 samples) (modelling development)

Validation set (152 samples) (modelling assessment)

1. Linear Discriminant Analysis
LDA (software UNIWIN)

Pbs: - multicollinear λ data
- limited number of variables

PCA

stepwise DA

LDA / PCs

LDA/SW
selected λ

2. Discriminant Partial Least Squares Analysis (DPLS)
(software WinISI)

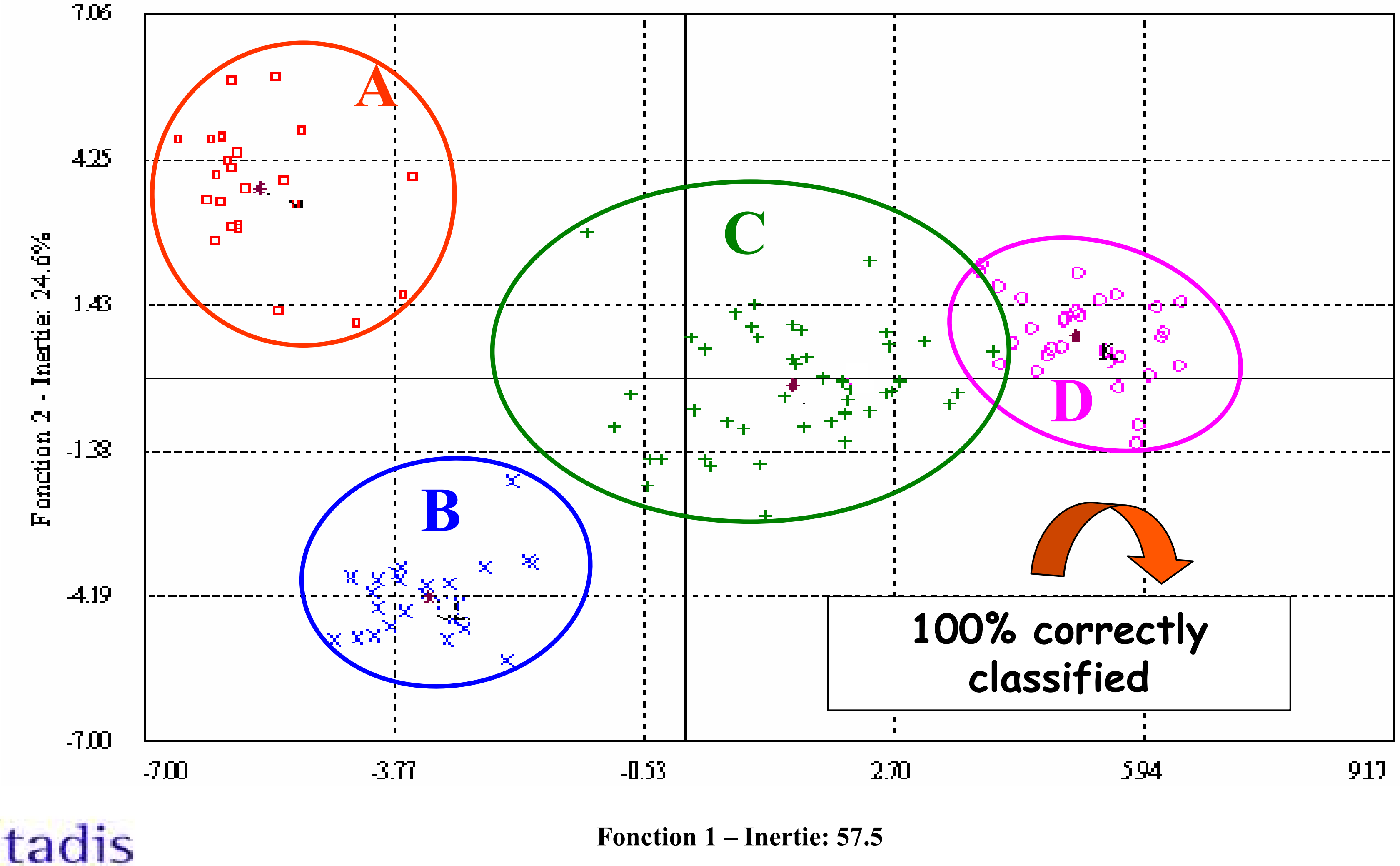
3. Soft Independent Modelling of Class Analogy
(software SIMCA)

Altadis

Linear Discriminant Analysis on principal components (LDA/PCs)

- **Methodology**:
  - data compression step: spectra of training set subjected to a principal component analysis (PCA) to generate orthogonal eigenvectors and sample scores
  - LDA applied to the principal component scores of samples of training set and validation set
    - samples of validation set are considered as supplementary observations
- **Application**: investigation to find the optimal number of PCs in order to obtain an efficient LDA
  - visible + nir : 30, 40, 60, 80 PCs
  - nir : 10, 20, 30 PCs

Training Set LDA / 20 PCs
Factorial Plane 1-2
Spectra SNVD 144, Nir segment

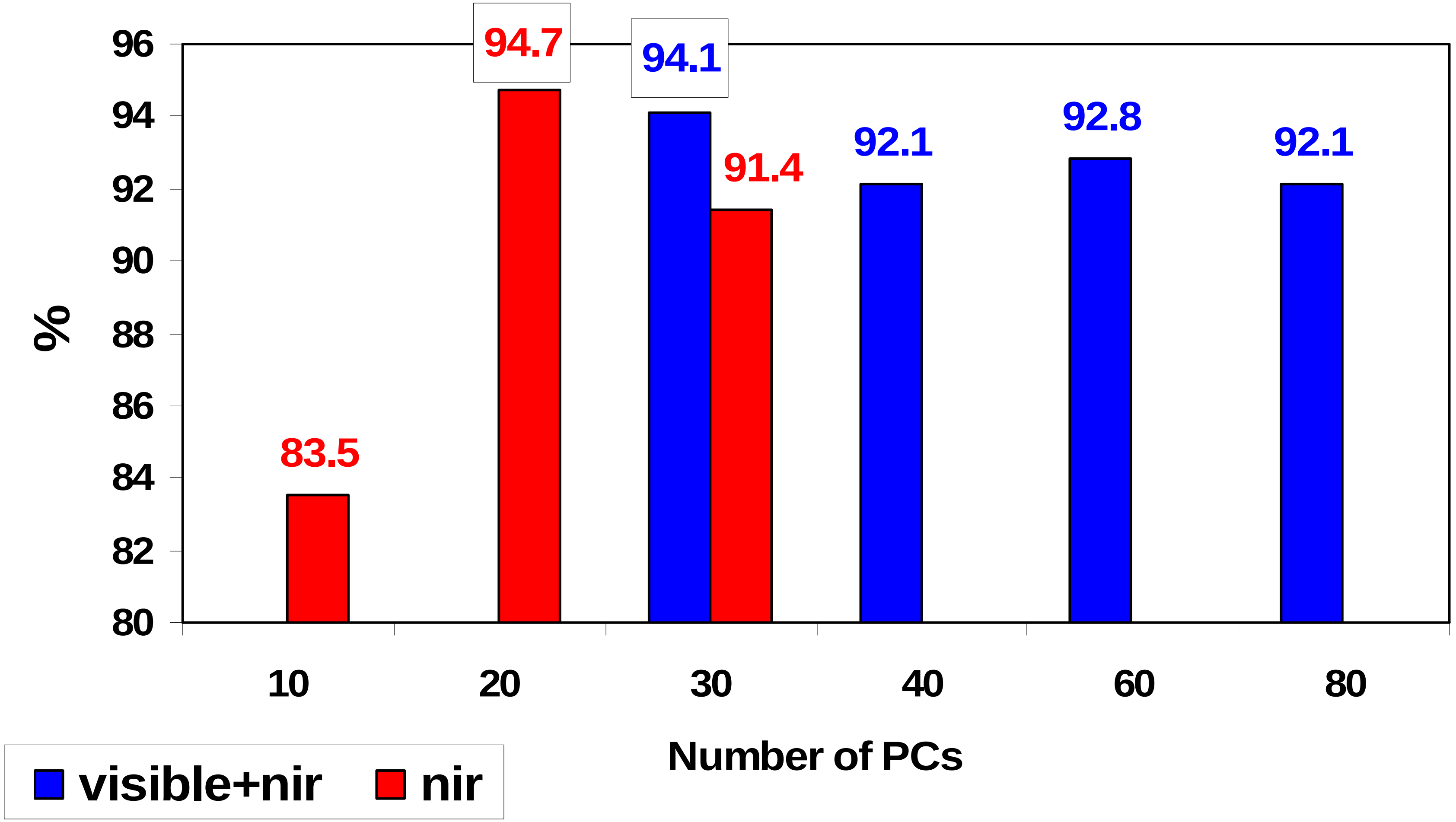


100% correctly classified

Fonction 1 – Inertie: 57.5

Altadis

Validation Set : % well classified according to the number of PCs



Altadis

Main results of the study
Study's main results

			LDA				DPLS		SIMCA
			LDA / PCs		LDA / SW		1st derivative	2nd derivative	
Visible+Nir	% correctly classified	TRAINING SET n = 126	30 PCs	100%	71λ	100%	99.30%	96%	96.90%
		VALIDATION SET n = 152		94.1% (143/152)		92.1% (140/152)	91.4% (139/152)	86.2% (131/152)	35.90%
Nir		TRAINING SET n = 126	20 PCs	100%	57λ	100%	98.60%	97.40%	95.80%
		VALIDATION SET n = 152		94.7% (144/152)		90.1% (137/152)	88.8% (135/152)	82.2% (125/152)	30.70%

LDA > D-PLS

LDA/PCs > LDA/SW

LDA: NIR > VIS + NIR

DPLS: VIS + NIR > NIR

Pb: Simca not suitable for this application

QUALITATIVE APPROACH

Second application

Near infrared spectroscopy and quality control in manufacturing plant

Quality control near infrared measurements of tobacco materials in manufacturing plant

- **Objective**: propose a methodology for the homogeneity Nirs control of processes and finished blends
- Comparison of raw or pretreated sample spectra because a spectrum can be considered as physical or chemical fingerprint of product: homogeneity assessment between samples by comparing spectra
  - Necessary to define a similarity criterion: distance metric
    - Normalized Mahalanobis distance between each product-spectrum and the average product-spectrum in the multidimensional space of PCA scores

→ Spectral Global Index for quality control

Altadis

Mahalanobis Distance in PC Space

Distance between each spectrum i and the mean spectrum μ of the distribution of a set of spectra in the multidimensional space:

$$MD^2 = (Ai - \mu)' C\text{-}1 (Ai - \mu)$$

- Calculation of MD done on PC scores of all product spectra
  (assuming normal distribution of data in the training set)
- MD is a measure of the sample dispersion in a distribution: to sort out products according to their MD to mean spectrum and to detect outlier samples

Altadis

Example of manufacturing follow-up

- In order to decide if the differences between manufactured samples are significant or not:
  - necessary to define threshold limits of Mahalanobis distances calculated in the multidimensional space in order to detect outliers

- Control chart method to check quality of manufactured products and to detect outlier samples: to set up the mean chart ↳ calculation on a learning set of samples:
  - central line (CTR) mean value (=1)
  - calculation upper control limit ($CTR + 3s\sqrt{n}$) and upper warning limit ($CTR + 2s\sqrt{n}$)

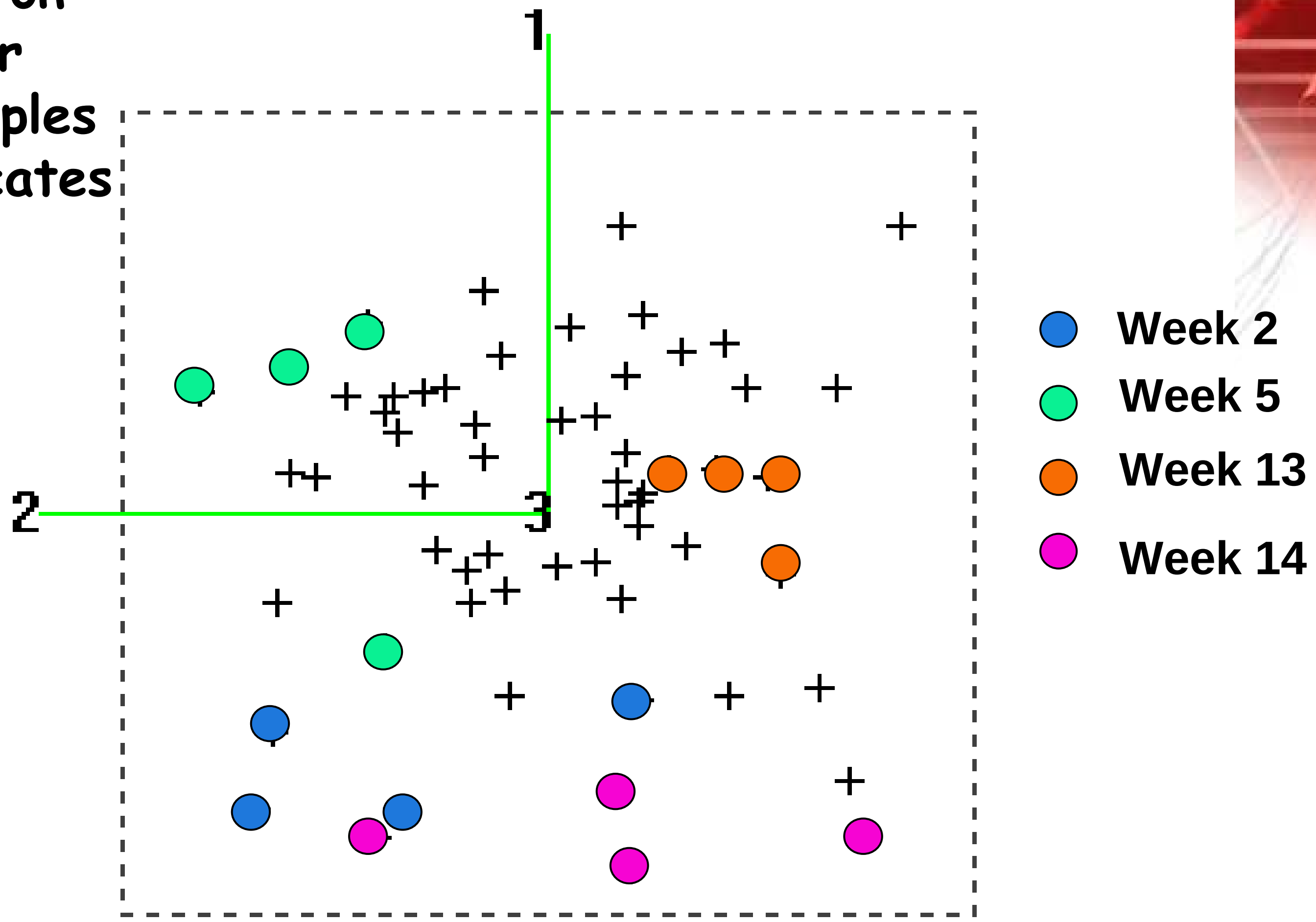
Example of manufacturing follow-up

- ➔ Example of one type of manufactured product blend in a factory: 15 pickings during 2005 productions
- ➔ Protocol:
  - Cut tobacco sample scanned by NIRS in four replicates
  - Calculation of Mahalanobis Distance of each picking to average spectrum from library set previously defined by PCA
    - Infrared range (1100 – 2500 nm)
    - Raw spectra (physical properties)
    - Pretreated spectra : Std Normal Variate Detrend and 1st derivative (chemical properties)

Altadis

Manufacturing follow-up in factory

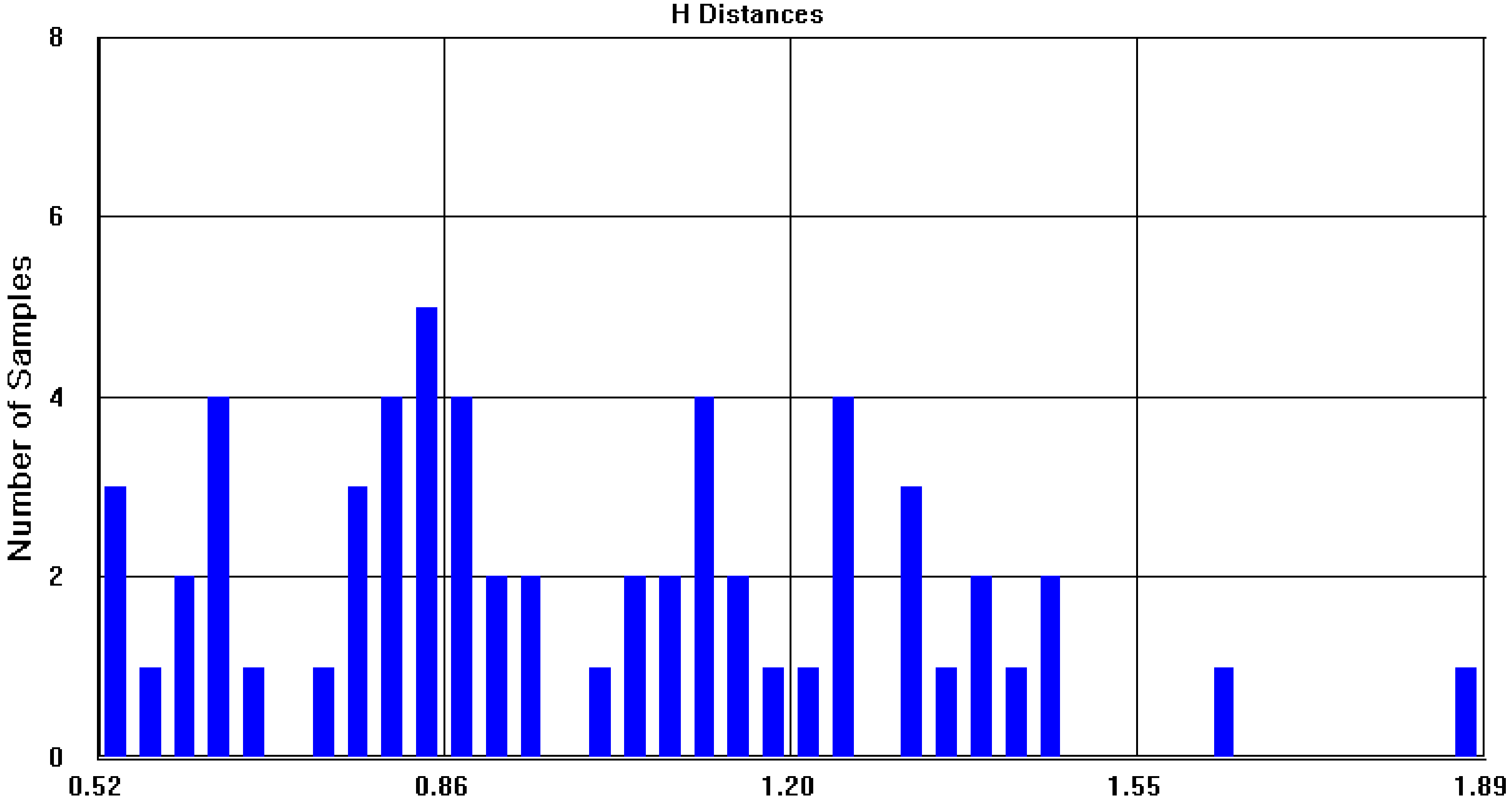
Example : PCA on pretreated nir spectra; 15 samples with 4 nirs replicates



Altadis

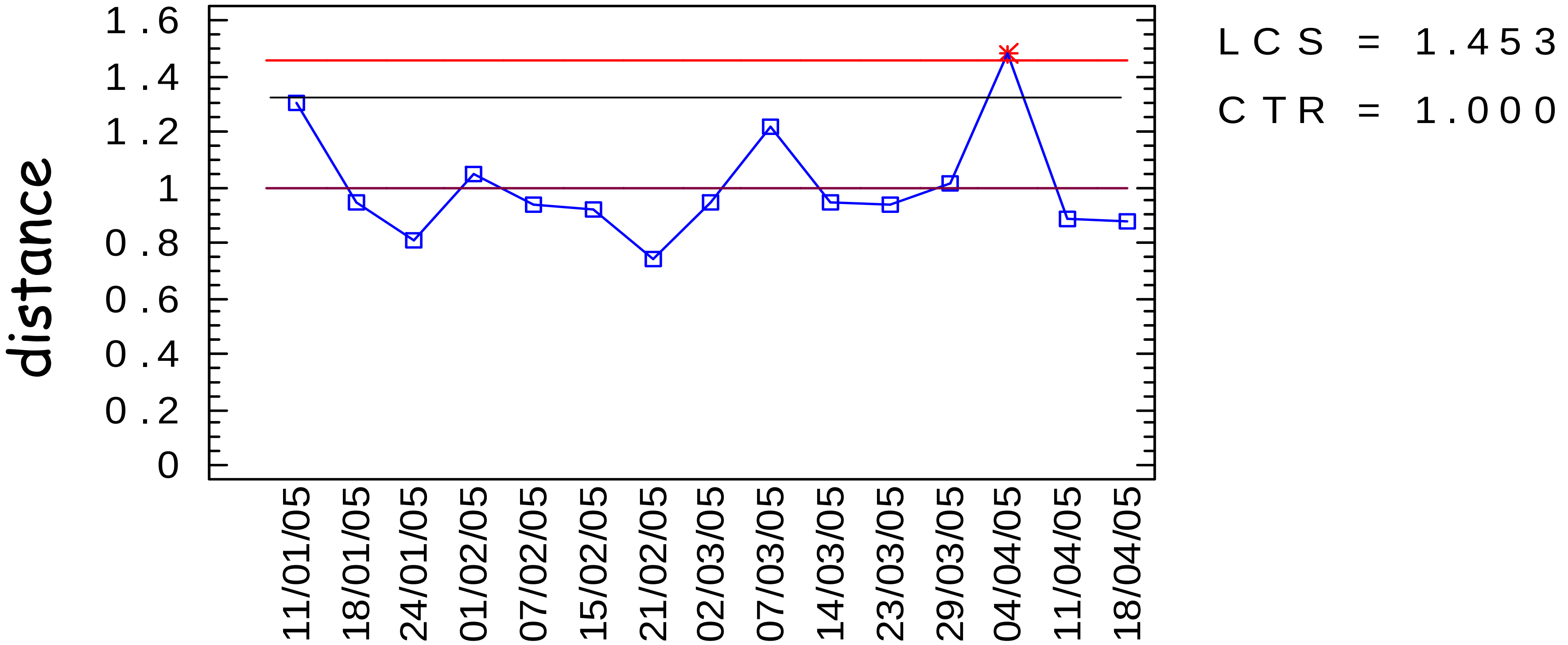
Manufacturing follow-up in factory

Histogram of individual spectral distances from the mean of the library file (pretreated spectra)



Manufacturing follow-up in factory: *mean control chart of distances*

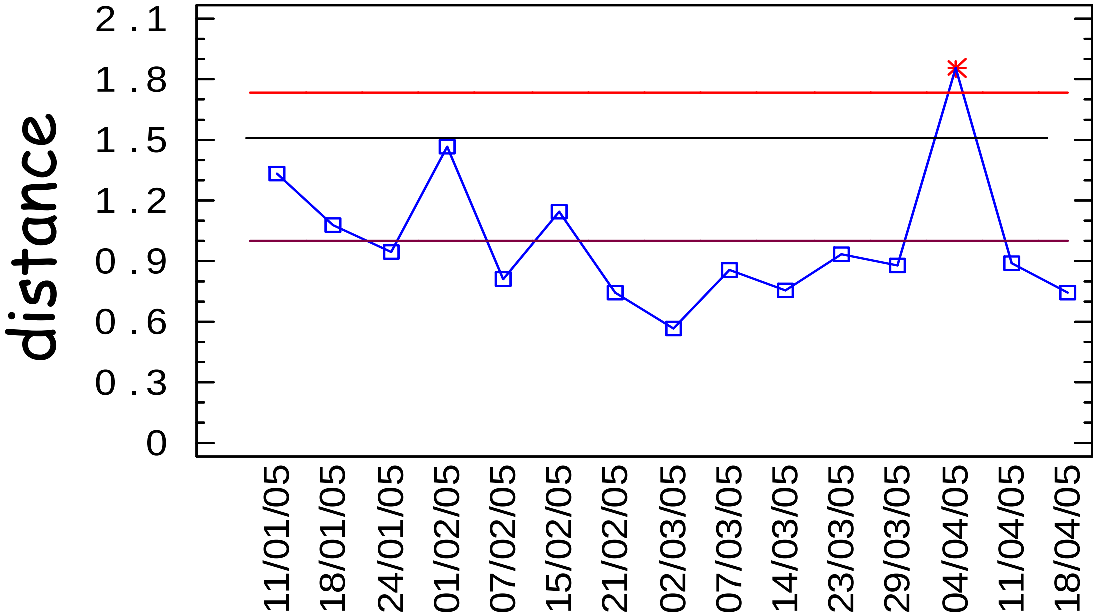
Pretreated spectra



Assessment of homogeneity due to chemical composition

Manufacturing follow-up in factory: *mean control chart of distances*

Raw spectra



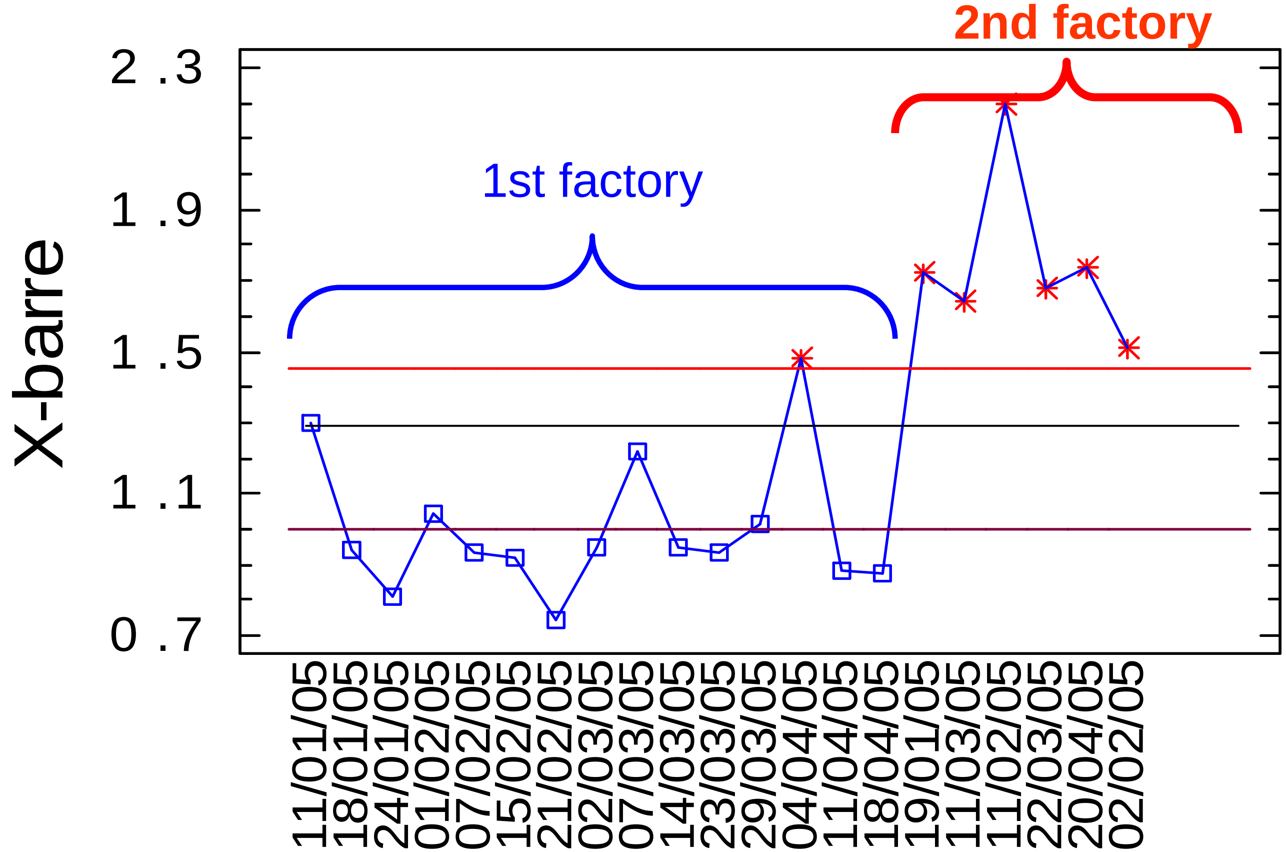
LCS = 1.728

CTR = 1.000

Assessment of homogeneity due to granulometry

Comparison of two blends from two factories

Pretreated spectra



LCS = 1.453

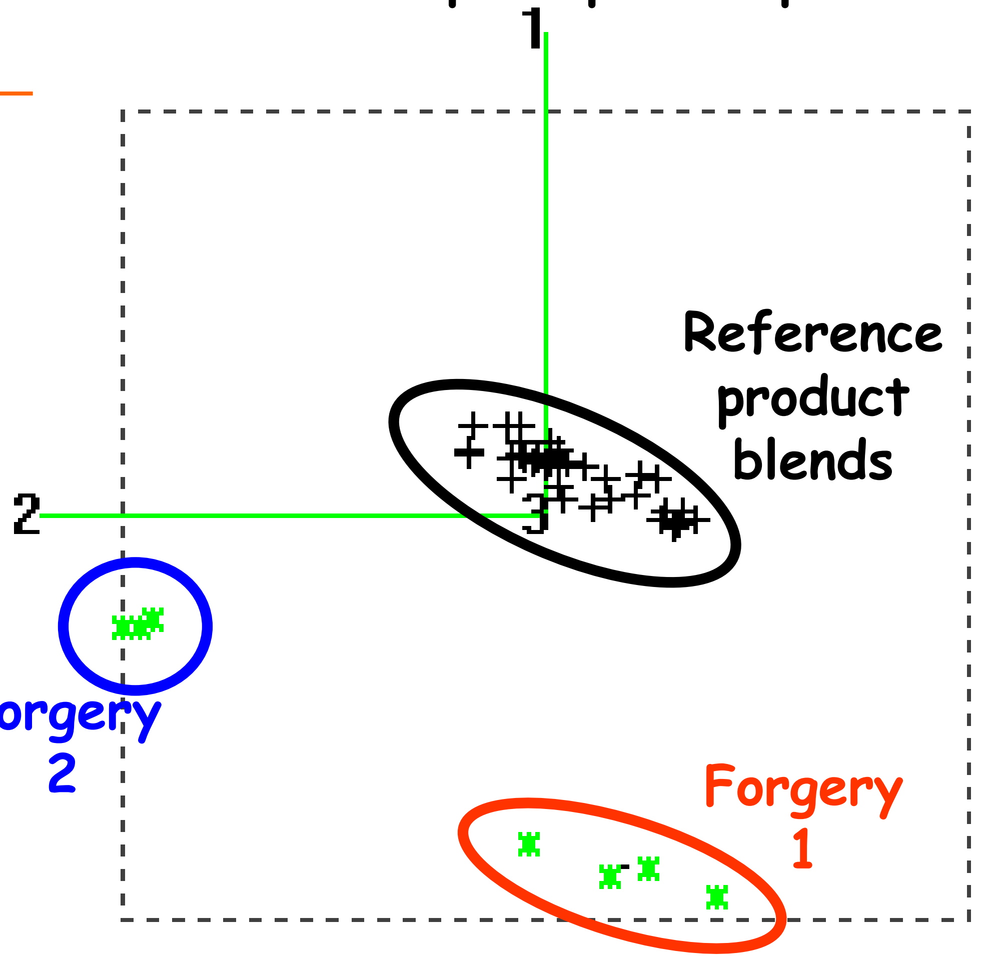
CTR = 1.000

Altadis

Use of NIRS to detect (easily) forgeries

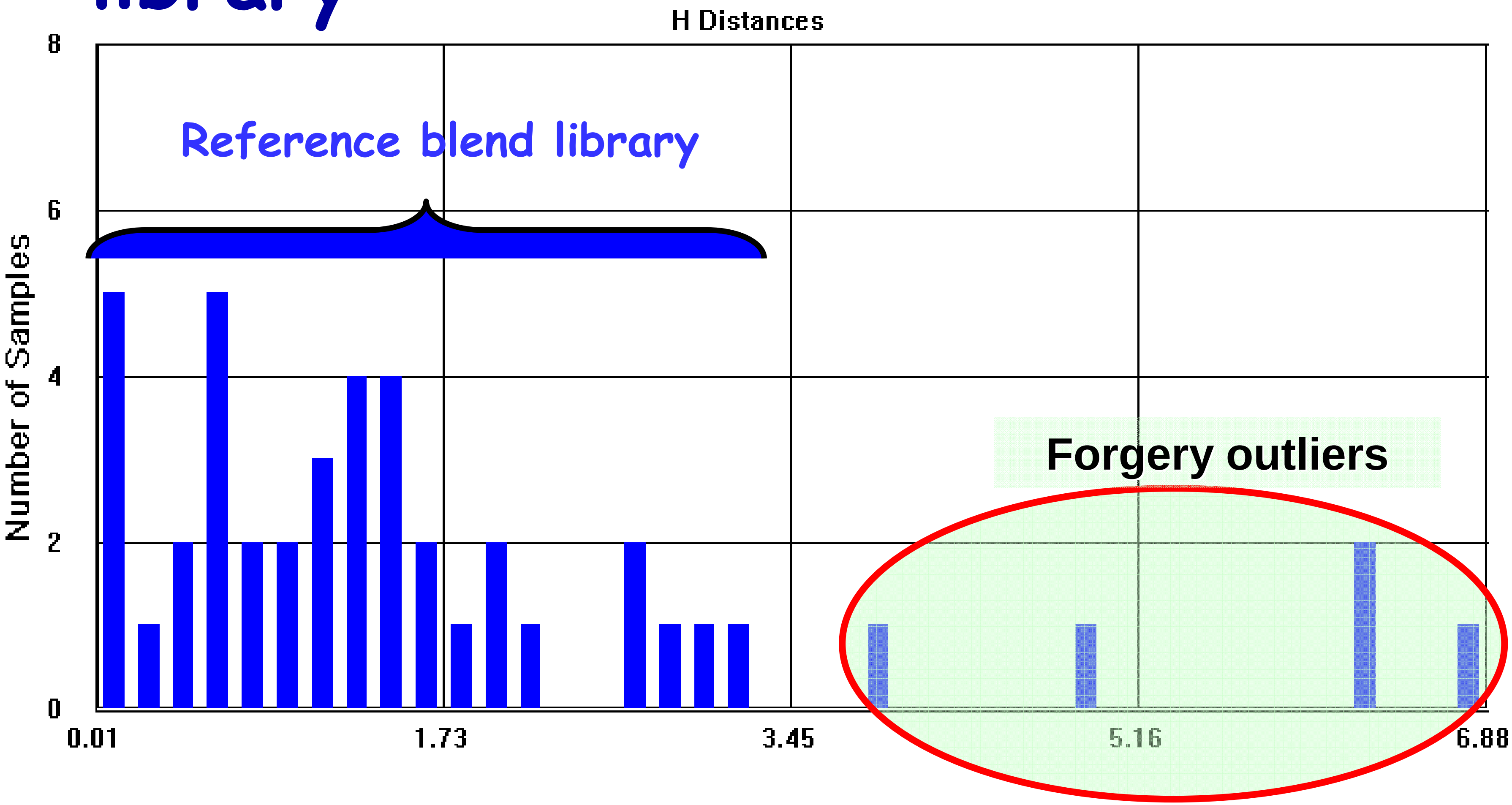
Score plots on the two first principal components space

Pretreated spectra of cut tobacco from 2 forgery blends and reference blends



Altadis

Histogram of distances from the mean of the reference library



Reference blend library

Forgery outliers

Conclusion

- **NIRS diffuse reflectance is highly related to the most of tobacco constituents:**
  - 23 chemical compounds vs 36 have a correlation of prediction higher than 0.8 and a relative prediction error lower than 10%
- **Best multivariate calibrations involve:**
  - signal transformation : scatter and curvature corrections (MSC, SNV; detrending),
  - derivative signal: 1st or 2nd
  - modified partial least squares regression

→ **Models can be applied both on nir and visible + nir segments**

Conclusion

➔NIR reflectance spectrometry can be used to ***predict*** simultaneously chemical compounds contents, rapidly and with a suitable accuracy in a diverse group of tobacco deliveries

. but « predict » is not « determine »

. today NIRS is not a reference validated analytical method

. NIRS is proposed as a « screening » tool

Qualitative Analysis

Conclusion

- NIRS has the potential to be used as **an effective screening tool** on the basis of spectral information, to classify finished products combined with appropriate supervised pattern recognition method:
  - Linear discriminant method is the most appropriated to classify blends from spectral data with high prediction rate (95%) particularly LDA/PCs
  - DPLS algorithm is the most convenient to use

But the use of non supervised techniques (clustering…) is not <u>suitable</u> for NIRS data to classify tobacco blends !

Altadis

Qualitative Analysis

Conclusion

- **Control by NIRS** is suitable to check the quality and the homogeneity of manufactured product blends with easy decision rules (control charts)
- **Application of signal pretreatments** to distinguish heterogeneity due to particle size or to chemical composition
- **Limits of the technique**: no detection of trace contaminants therefore only heterogeneity of tobacco blends due to the major chemical constituents will be characterised by Near Infrared Spectrometry

Altadis

Altadis