

BASICS OF CHEMOMETRICS

Juan Antonio Fernández Pierna
Vincent Baeten
Pierre Dardenne

Walloon Agricultural Research Centre (CRA-W)
Valorisation of Agricultural Products Department
Gembloux, Belgium



INTRODUCTION

- Chemometrics Introduction
 - What is this and why we need it
- Some definitions
- Overview of methods
- Examples

Without equations!



X - METRICS

The use of multivariate analysis in the discipline X:

Statistical, mathematical or graphical technique, considers multiple variables simultaneously

- Biometrics (used in biology)
- Technometrics (used in engineering)
- Psychometrics (used in psychology)
- **Chemometrics** (used in chemistry)



CHEMOMETRICS – INTRODUCTION

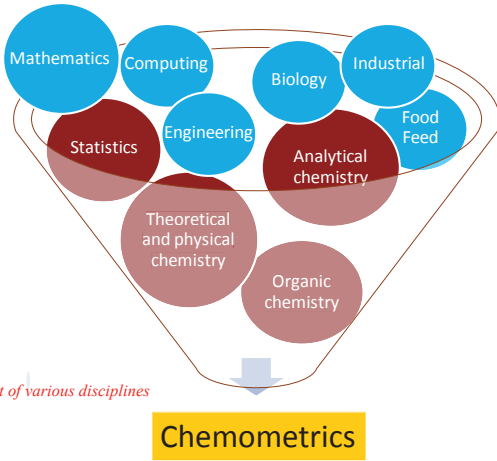


“Chemometrics is the chemical discipline that uses mathematics and statistics to design or select optimal experimental procedures, to provide maximum relevant chemical information by analyzing chemical data, and to obtain knowledge about chemical systems”

D. L. Massart



CHEMOMETRICS – INTRODUCTION



CHEMOMETRICS – INTRODUCTION

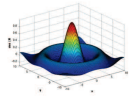
The scientific world today

• The data flood generated by modern analytical instrumentation produces **large quantity of numbers** to understand and quantify phenomena around us.



• The evolution of personal computers allows **faster acquisition, processing and interpretation** of chemical data.

• Every scientist uses **software** related to mathematical methods or to processing of knowledge.



• A deeper understanding of those methods and **tools** for viewing all data simultaneously are needed.



CHEMOMETRICS – INTRODUCTION

Use of mathematical and statistical methods for selecting optimal experiments
Statistical experimental design
Design of Experiments (DoE)...

Extracting maximum amount of information when analysing multivariate (chemical) data
Classification
Process monitoring,
Multivariate calibration...



CHEMOMETRICS – INTRODUCTION

CHEMOMETRICS

Analytical request

Analytical method

Analytical answer

Useful at any point in an analysis, from the **first conception of an experiment until the data is discarded.**



CHEMOMETRICS – APPLICATIONS

Huge growth area in past 15 years

- Process Control and analysis
- Food and feed analysis
- Biology – metabolomics etc
- Environmental monitoring
- Analytical Chemistry



CHEMOMETRICS – DEFINITIONS

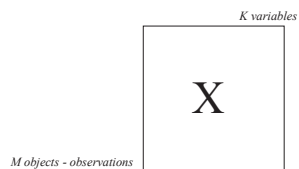
Linear algebra is the language of Chemometrics. One cannot expect to truly understand most chemometric techniques without a basic understanding of linear algebra (Wise and Gallagher, 1998)

Matrix and vector operations



CHEMOMETRICS – DEFINITIONS

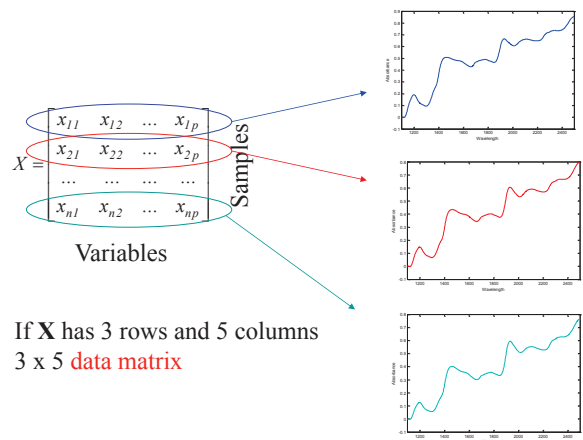
- Samples are referred to as **OBJECTS**
- Measurement results (e.g. concentration, absorbances, ...) are referred to as **VARIABLES**
- A data table of K variables and M objects is referred to as a **DATA MATRIX OF SIZE $M \times K$**



CHEMOMETRICS:
Extract meaningful information about the objects and the variables from data matrices



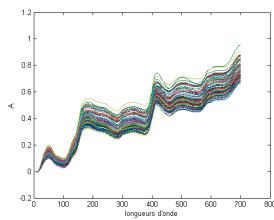
CHEMOMETRICS – DEFINITIONS



CHEMOMETRICS – INTRODUCTION

In summary ...

$$\begin{array}{|c|} \hline K \text{ variables} \\ \hline \mathbf{X} \\ \hline M \text{ objects} \\ \hline \end{array} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

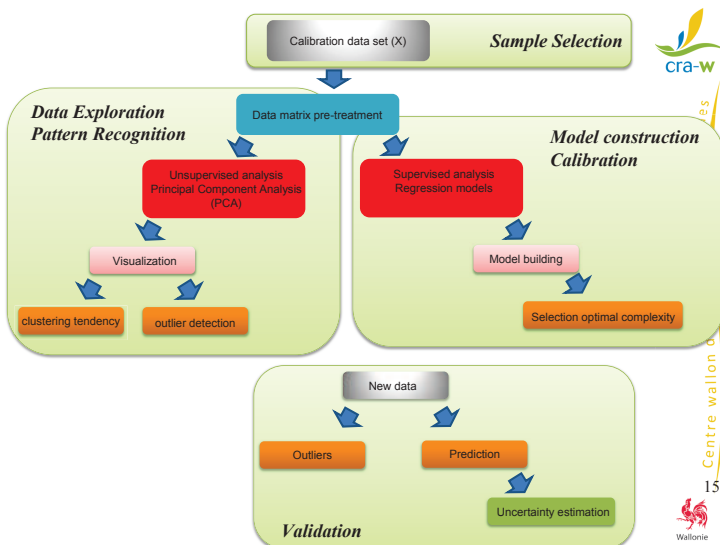


CHEMOMETRICS – IMPORTANT DISCIPLINES

- Sampling, selection of objects and variables
- Clustering
- Multivariate regressions, calibrations and predictions
- Neural Networks
- Validation
- Graphical display and outlier detection



Source: Chemometrics – Introduction - Jens C. Frisvad



SAMPLE SELECTION

INCLUDE IN THE CALIBRATION SET ALL THE FACTORS OF VARIATION:

Measurement factors:

- room temperature
- operator
- instrument setup

Population sources of variation:

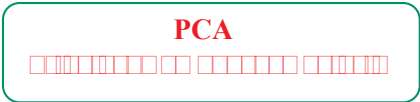
- origins of samples
- processes
- varieties
- storage conditions
- sample preparations (t°, particle size)
- residual moisture



PATTERN RECOGNITION



how similar are products...



...

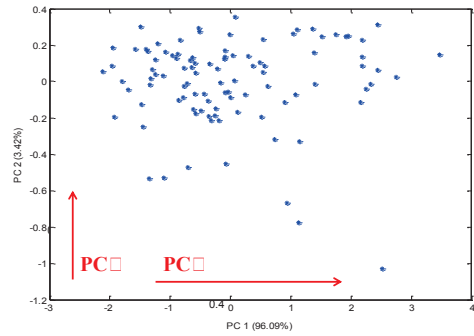
food samples / patients / people / spectra / ...

concentrations / spectral peaks / ...

Centre wallon de Recherches agronomiques



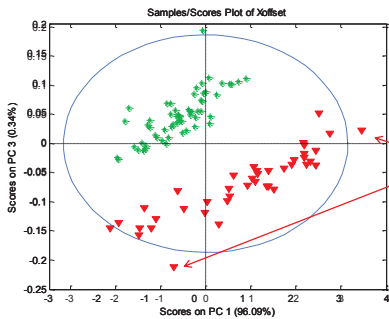
PRINCIPAL COMPONENT ANALYSIS



Centre wallon de Recherches agronomiques



PRINCIPAL COMPONENT ANALYSIS



Clusters

Outliers

Object space

Centre wallon de Recherches agronomiques



CALIBRATION



Centre wallon de Recherches agronomiques



MULTI-VARIABLE CALIBRATION



Centre wallon de Recherches agronomiques



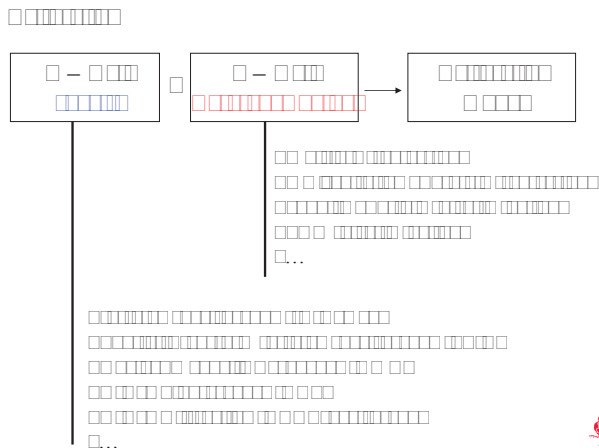
- Calibration of a model using a set of reference data
- Prediction of a new variable using the calibrated model
- Calibration of a model using a set of reference data and a set of new data

Calibration of a model using a set of reference data and a set of new data

MULTI-VARIABLE CALIBRATION



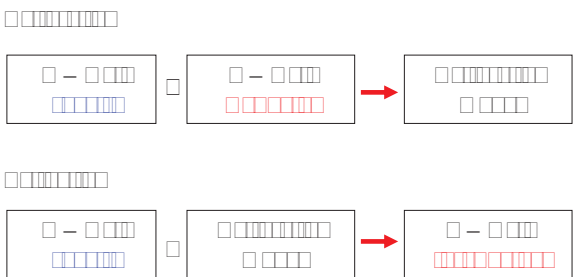
Centre wallon de Recherches agronomiques



MULTI-VARIABLE CALIBRATION - CALIBRATION



Centre wallon de Recherches agronomiques



Calibration uses empirical data and prior knowledge for determining to predict unknown quantitative information from available measurements via some mathematical transfer functions

MULTI-VARIABLE CALIBRATION

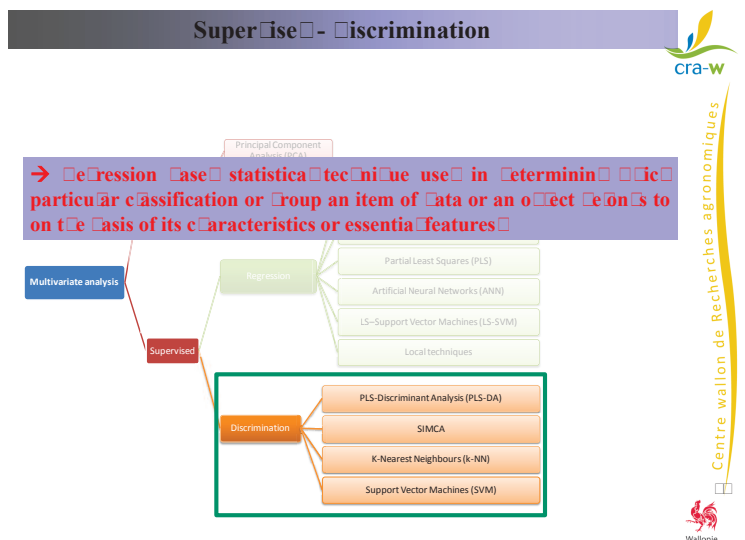
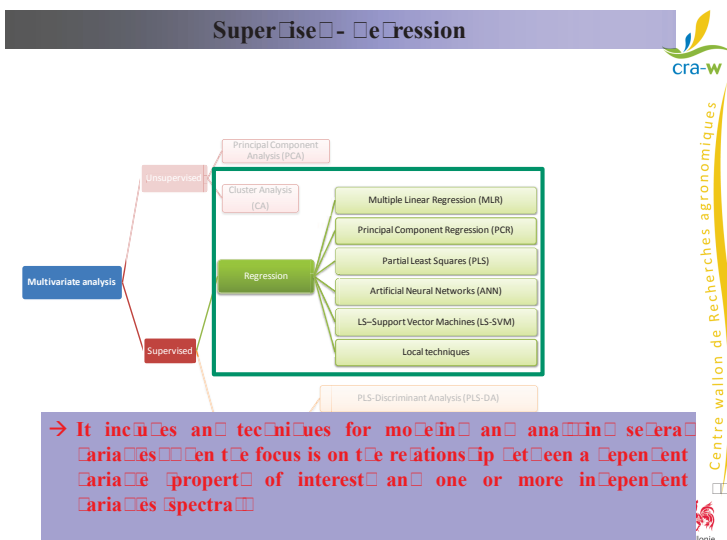
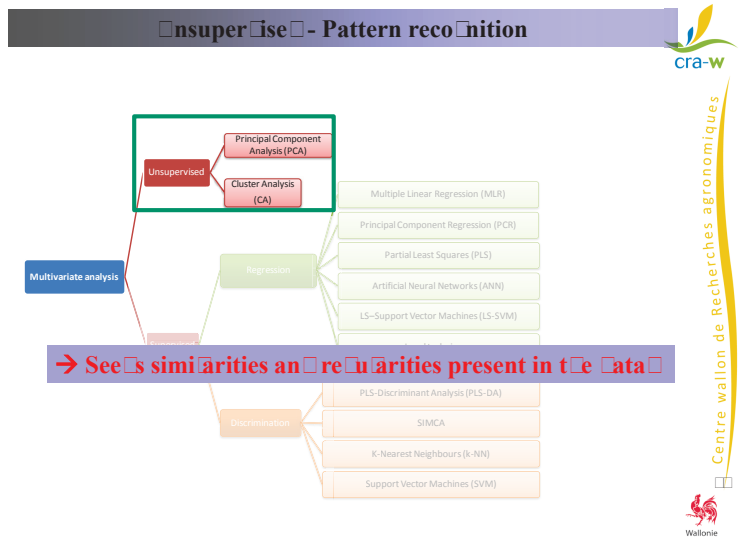
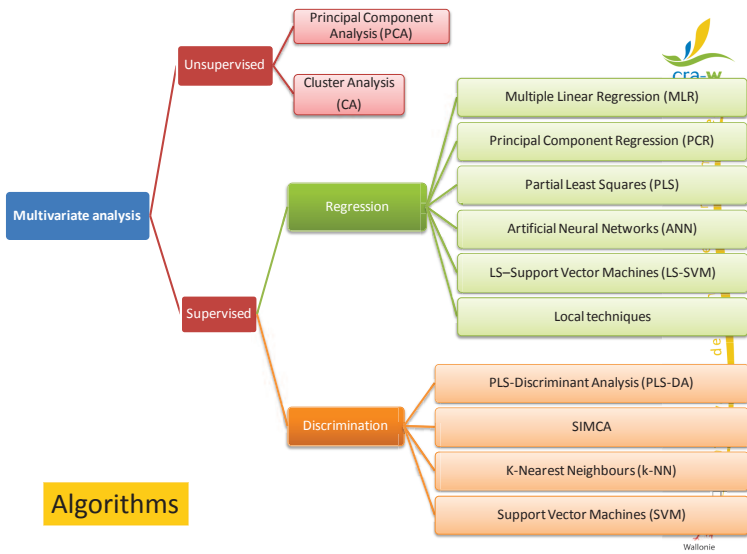


Centre wallon de Recherches agronomiques



X

- X is a matrix containing the reference data
- Y is a matrix containing the spectral data (NIR, MIR, Raman,...)
- B is a matrix containing the regression coefficients
- E is a matrix explaining the model error



AIABLE SELECTION

... ..

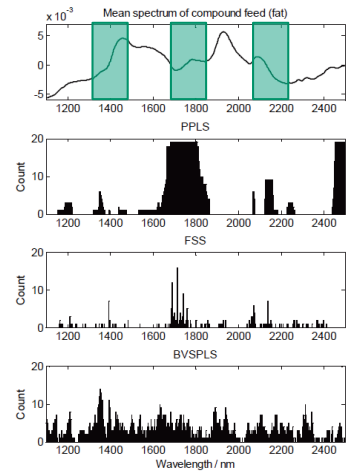
-
-
-



Centre wallon de Recherches agronomiques



AIABLE SELECTION



Centre wallon de Recherches agronomiques



AIABLE SELECTION

...

... ..

...

- '... (BVSPLS)' ...
- '...' ...



Centre wallon de Recherches agronomiques



EMOMETICS – SOFTWARE

TABLE XXVI
Software for Use with Near-Infrared Reflectance Instruments

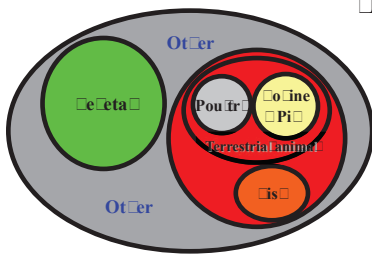
Type	Company	Address	Contact Information
Dedicated			
WINISI	InfraSoft International	Port Matilda, PA, U.S.A.	814/237-0867 (fax)
NSAS	Foss/NIRSystems	Silver Spring, MD, U.S.A.	301/236-0134 or 301/989-1485 (fax)
Vision	Foss/NIRSystems	Silver Spring, MD, U.S.A.	301/236-0134 or 301/989-1485 (fax)
Sesame	Bran+Luebbe	Buffalo Grove, IL, U.S.A.	847/520-0855 (fax)
SpectroMetrix	LT Industries	Rockville, MD, U.S.A.	301/468-2230 (fax)
AnaTec	Buhler Corporation	Uzwil, Switzerland	+41 71 955 3356 (fax)
Delight	Buhler Corporation	Minneapolis, MN, U.S.A.	612/540-9246 (fax); www.buhlerusa.com
	D ² Development	LaGrande, OR, U.S.A.	mayesd@dsquared-dev.com
Generic			
GRAMS 386	Galactic Industries	Salem, NH, U.S.A.	603/898-6228 (fax)
Pirouette	InfoMetrix	Woodville, WA, U.S.A.	206/402-1040 (fax)
Unscrambler	CAMO	Corvallis, OR, U.S.A.	541/757-1402 (phone); www.camo.com
Unscrambler	CAMO AB	Trondheim, Norway	+47 73 514 257 (fax)
MatLab	The MatWorks, Inc.	Matck, MA, U.S.A.	508/647-7001 (fax)



Centre wallon de Recherches agronomiques



EXAMPLE: ANIMAL



ANIMAL



Confusion Matrix	CAL		LOOC	
	classified as	not	classified as	not
is	████	██	████	██
est	██	████	██	████

████ correct classification



Centre wallon de Recherches agronomiques



EXAMPLE: ANIMAL

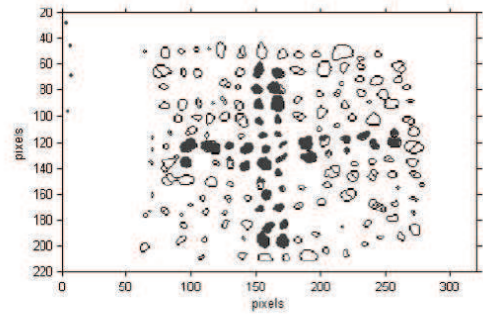


Figure 8. Results of the SVM model for the 'cross' data set using the whole spectra (76 800 pixels). Black dots are classified as animal.



Centre wallon de Recherches agronomiques



EXAMPLE: CEREALS - IMPURITIES

Original image	Background vs rest	Animal contaminant vs rest	Cereals vs rest	Botanical impurities vs rest	Other contaminants vs rest	Cellulose waste vs rest

SVM discrimination models

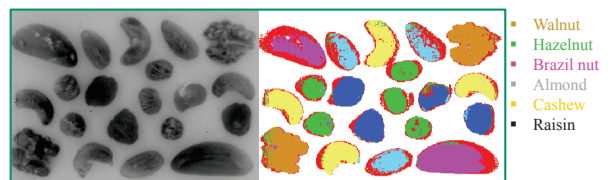
'NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed' J.A. Fernández Pierna, Ph. Vermeulen, O. Amand, A. Tossens, P. Dardenne and V. Baeten. Special issue Chemometrics and Intelligent Laboratory Systems 117 (2012) 233-239



Centre wallon de Recherches agronomiques



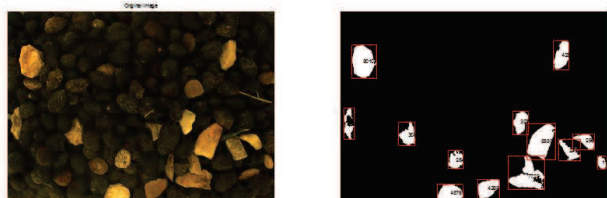
EXAMPLE: NUTS AND DRIED FRUITS



Centre wallon de Recherches agronomiques



EXAMPLE: OLIVES - IMPURITIES



KNN - PCA models

Application of a vision system for on-line determination of quality parameters of olive fruits. J. A. Fernández Pierna, O. Amand, A. Tossens, P. Dardenne and V. Baeten. Food Chemistry 117 (2013) 200-209

Centre wallon de Recherches agronomiques



EXAMPLE: OLIVE OIL - QUALITY PARAMETERS

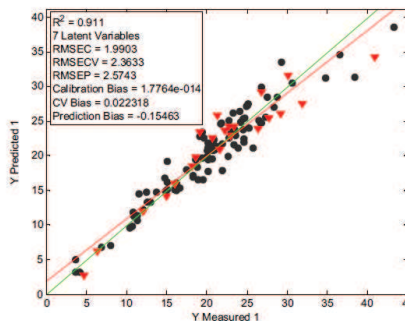


Fig. 3. Scatter plot of the actual Peroxide Value (IP) and predicted IP for calibration (circle) and validation (triangle) sets.

Application of low-resolution Raman spectroscopy for the analysis of oxidized olive oil. J. A. Fernández Pierna, O. Amand, A. Tossens, P. Dardenne and V. Baeten. Food Control 22 (2011) 203-209

Centre wallon de Recherches agronomiques



EXAMPLE: OLIVE OIL - ADULTERATION

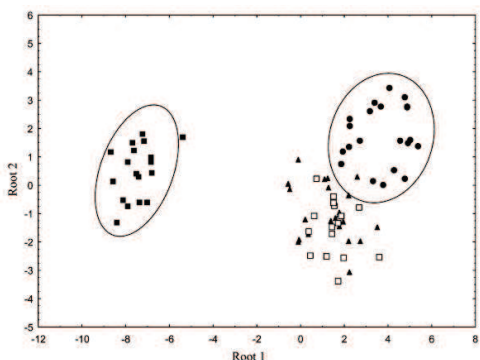


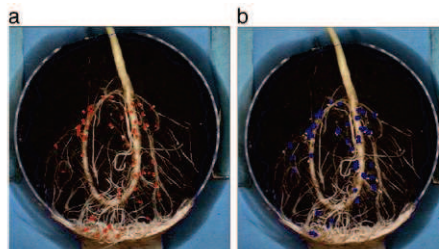
Figure 4. FT-MIR spectra of the unsaponifiable matter of the test samples: classification of genuine olive oil (●) and hazelnut oil (■) and their mixtures (>10% (□) and <10% (▲)) by S-LDA. Genuine edible oils are delimited by ellipses of confidence at $\alpha = 0.90$.

Detection of the presence of adulteration in olive oil by Raman and FT-Spectroscopy. Baeten, V., Fernández Pierna, O., Dardenne, P., Tossens, A., Amand, O. Journal of Agricultural and Food Chemistry 58 (2010) 201-209

Centre wallon de Recherches agronomiques



EXAMPLE: SUGAR BEET PLANTS - COST



Optical microscopy | Reflectance imaging

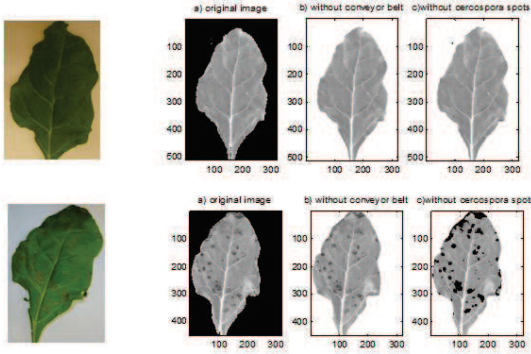
SVM discrimination models

High-resolution Raman spectroscopy and chemometrics for the detection of undesirable substances in food and feed. J.A. Fernández Pierna, Ph. Vermeulen, O. Amand, A. Tossens, P. Dardenne and V. Baeten. Special issue Chemometrics and Intelligent Laboratory Systems 117 (2012) 233-239

Centre wallon de Recherches agronomiques



EXAMPLE: SUGAR BEET PLANTS – CEROSPORA



Centre wallon de Recherches agronomiques



EXAMPLE: FEED PRODUCTS

Multivariate regression method comparison: PLS, ANN and LS-SVM



Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products

J.A. Fernández Pierna^{a,*}, B. Lecler^a, J.P. Conzen^b, A. Niemoeller^b, V. Baeten^a, P. Dardenne^a

^aWallonia Agricultural Research Centre (CRA-W), Unit of Agricultural Products Department, Food and Feed Quality Unit (1015), Research building, Chaussée de Namur 24, 5030 Combloux, Belgium

^bBILKER OPTIK GmbH, NIR & Process Technology, Rudolff-Platz-Str. 27, 74275 Ettlingen, Germany



Centre wallon de Recherches agronomiques



EXAMPLE: FEED PRODUCTS

Feed ----- (28676x700)

Ash, Fat, Fibre, Starch, Protein

Feed Ingredients ----- (26652x700)

Ash, Fat, Fibre, Protein

Fresh Silages ----- (1035x700)

Dry Matter, Fibre, Protein

Soils ----- (1625x700)

CEC, COT_SK, N_Kj

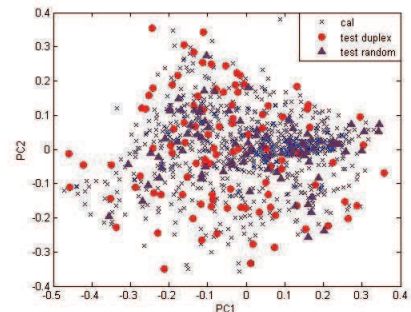
Pre-processing: SNV + detrend + First derivative



Centre wallon de Recherches agronomiques



EXAMPLE: FEED PRODUCTS



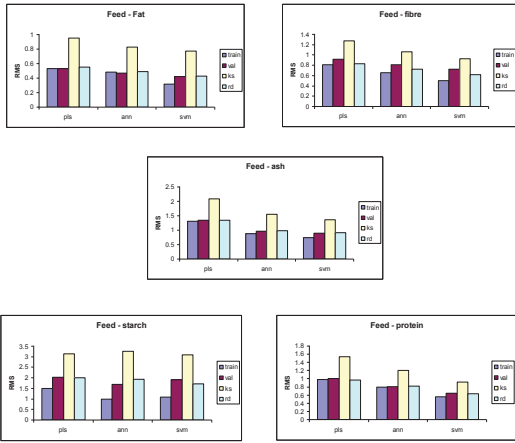
Example of data distribution (Soils N_Kj)



Centre wallon de Recherches agronomiques



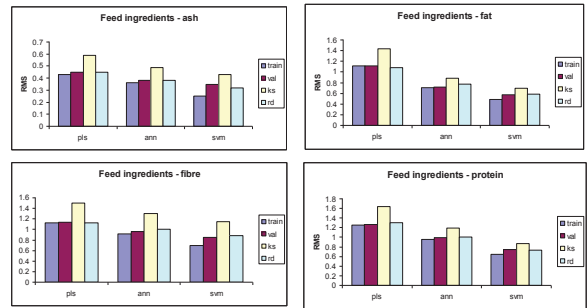
EXAMPLE: FEED PRODUCTS



Centre wallon de Recherches agronomiques



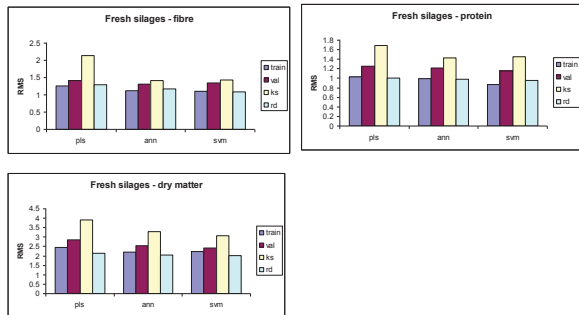
EXAMPLE: FEED PRODUCTS



Centre wallon de Recherches agronomiques



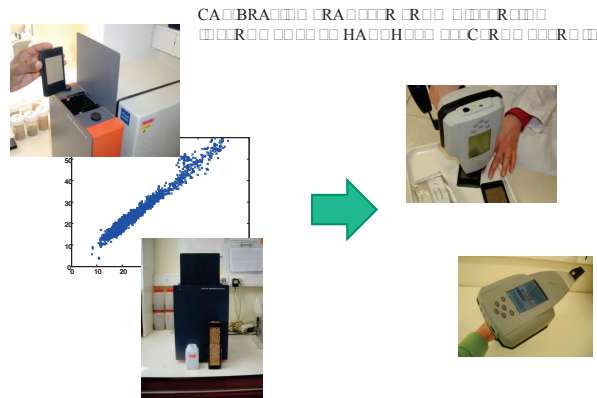
EXAMPLE: FEED PRODUCTS



Centre wallon de Recherches agronomiques



EXAMPLE: TRANSFERT

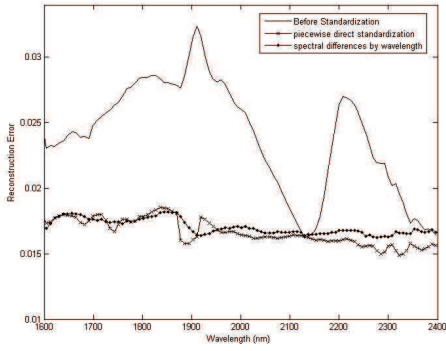


Centre wallon de Recherches agronomiques



'Calibration Transfer from Dispersive Instruments to Handheld Spectrometers', J.A. Fernández Pierna, P. Vermeulen, B. Lecler, V. Baeten, P. Dardenne. Applied Spectroscopy 64 (6) (2010)

EXAMPLE: TRANSFERT



Centre wallon de Recherches agronomiques



ESTIMATION METRICS



Centre wallon de Recherches agronomiques



Vibrational spectroscopy and Chemometrics

Training session
17-21 February 2014



Recherches agronomiques

Vibrational Spectroscopy

NIR/MIR/RAMAN
Theory and applications, experimental/sample plan

NIR microscopy/Hyperspectral imaging/embedded NIR/
MEMs/instrument standardization/networking



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

