UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

NIR & Chemometrics: how to handle NIR signals, recover information and build models

Marina Cocchi

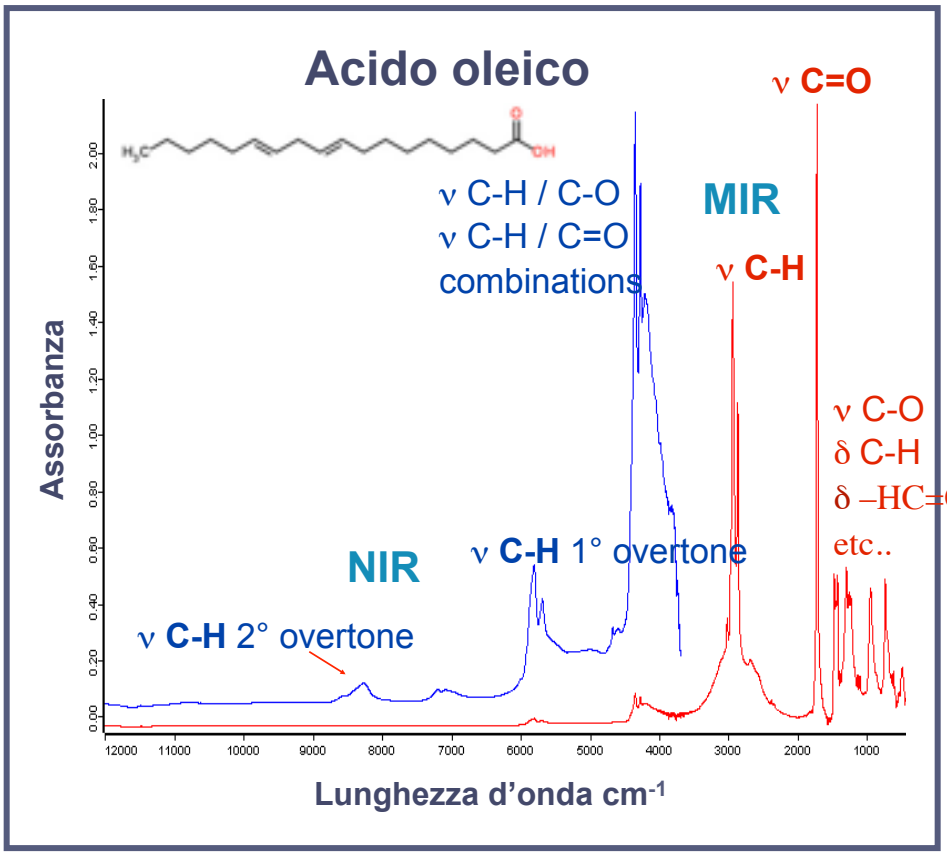
marina.cocchi@unimore.it

Wednesday, May 25, 2016

- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

    1. Handle variability (preprocessing)
    2. Monitoring Quality (raw materials)
    3. Process evolving with time
    4. Treatment effects
    5. Multispectral/Hyperspectral imaging

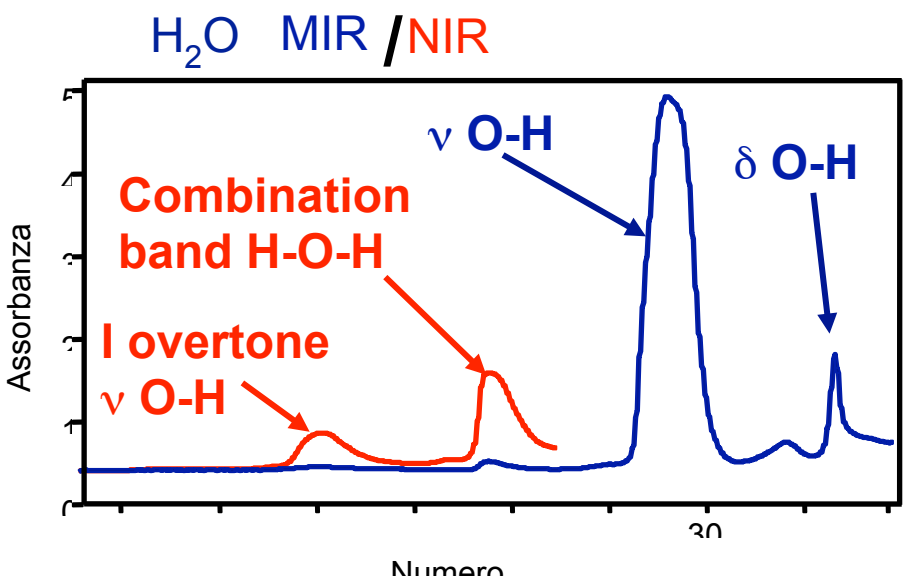
NIR Peculiarities/Issues



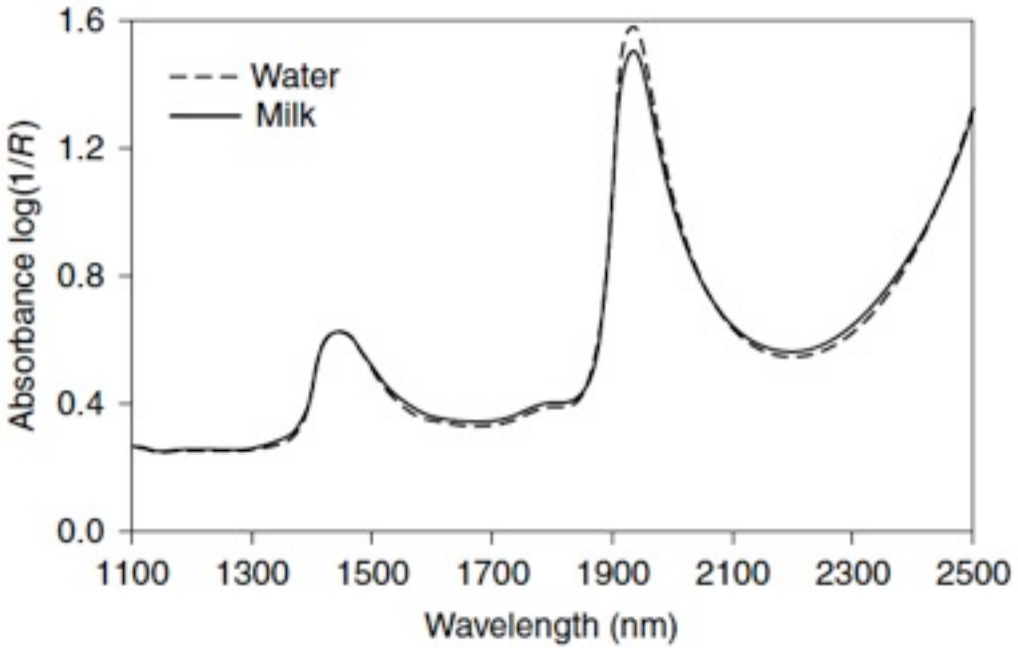
Strong Bands Overlap
due to sum of combinations and overtones of bonds such as:
C-H, N-H, O-H, S-H, C=O, C=C

Absorption Coefficients from 20 to 100 times lower than in the mid-Infrared range

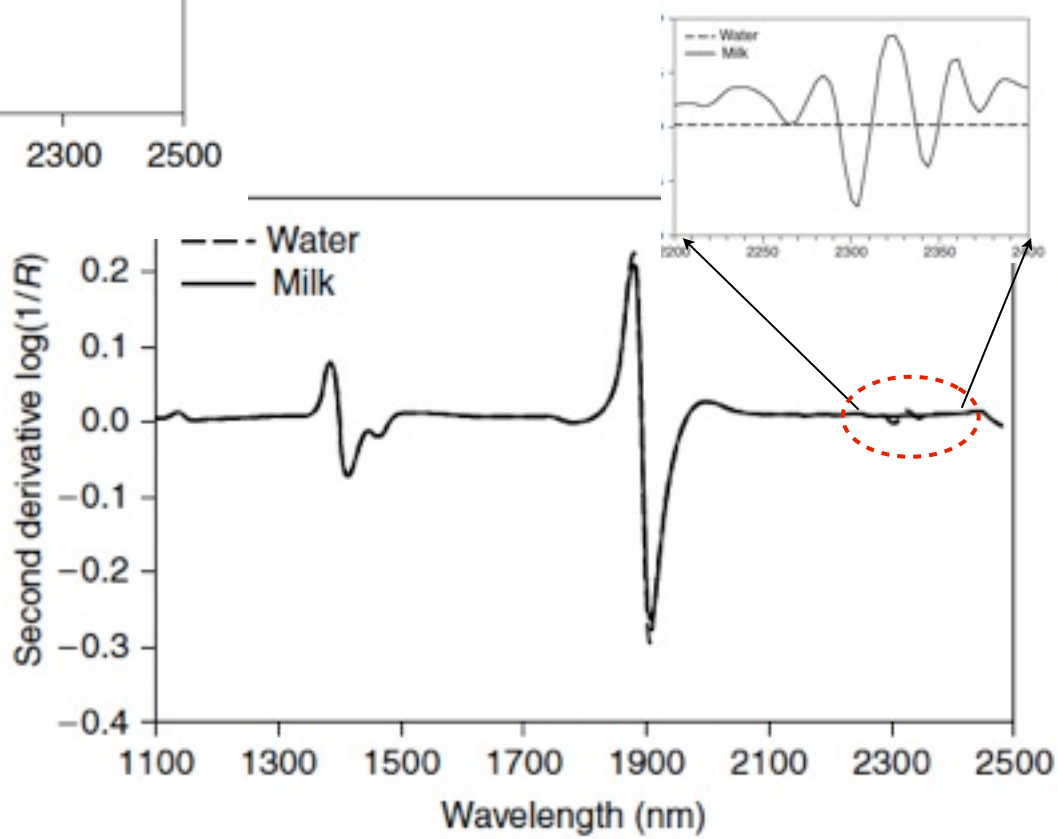
How to extract relevant information ?



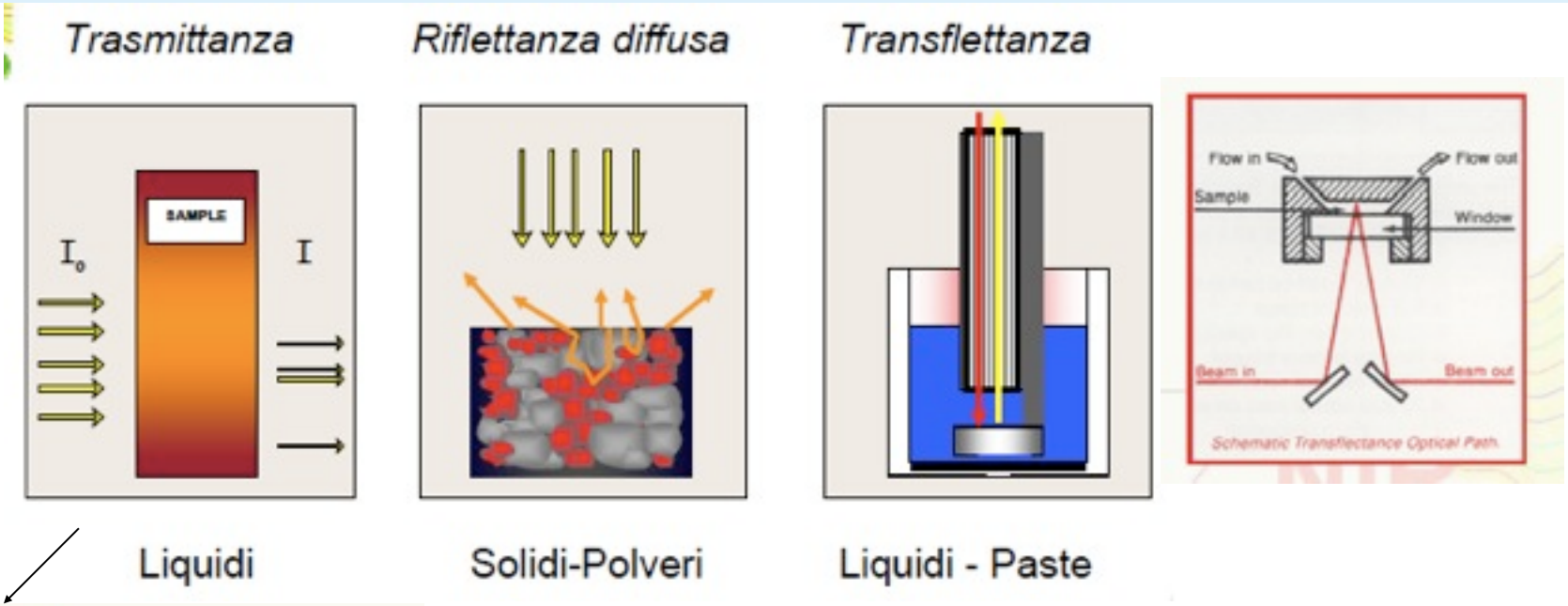
how to extract relevant information?

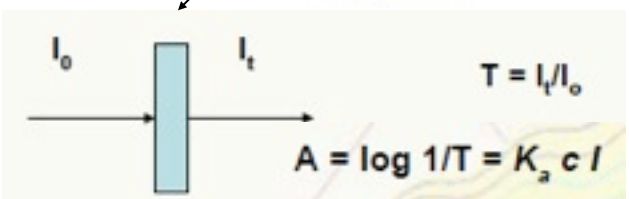


water absorption masks other constituents contribution



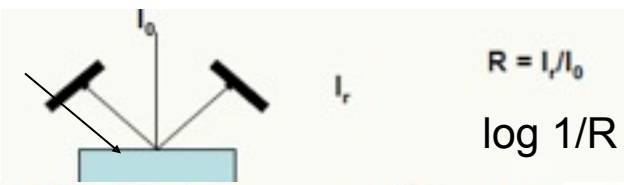
Quantitative Information ?





Linear with concentration

But also optical fibre, measuring head



Legge di Kubelka-Munk

$$f(R_\infty) = K_r / s \propto \frac{ac}{S}$$

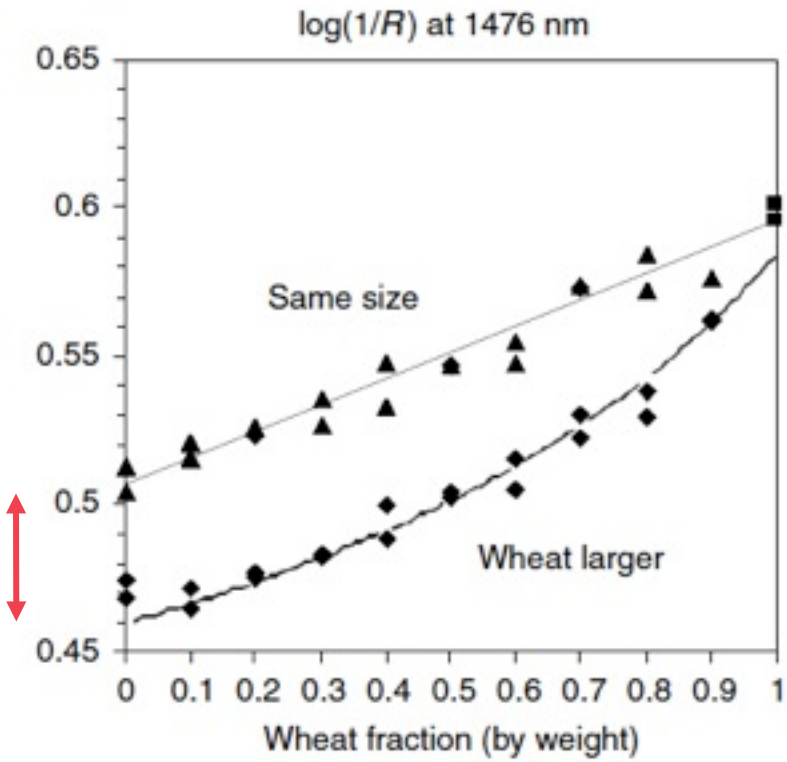
Departure from Linearity easily to occur: anistropy, etc..

NIR variability

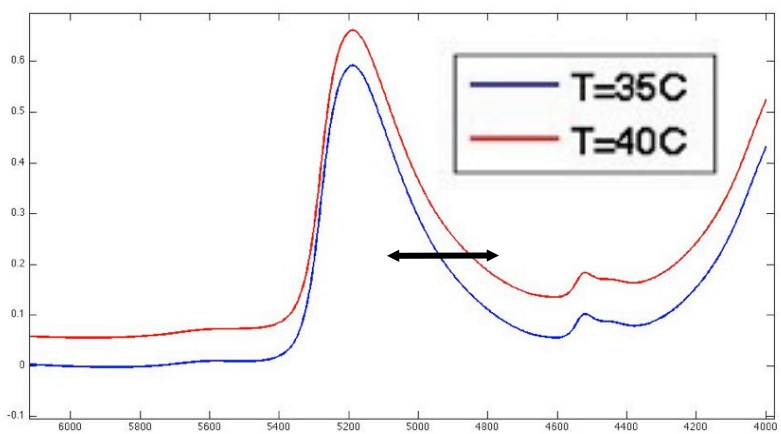
Sources of variation in NIR (other than chemical composition)

- Difference in particles Size

(multiplicative scattering)

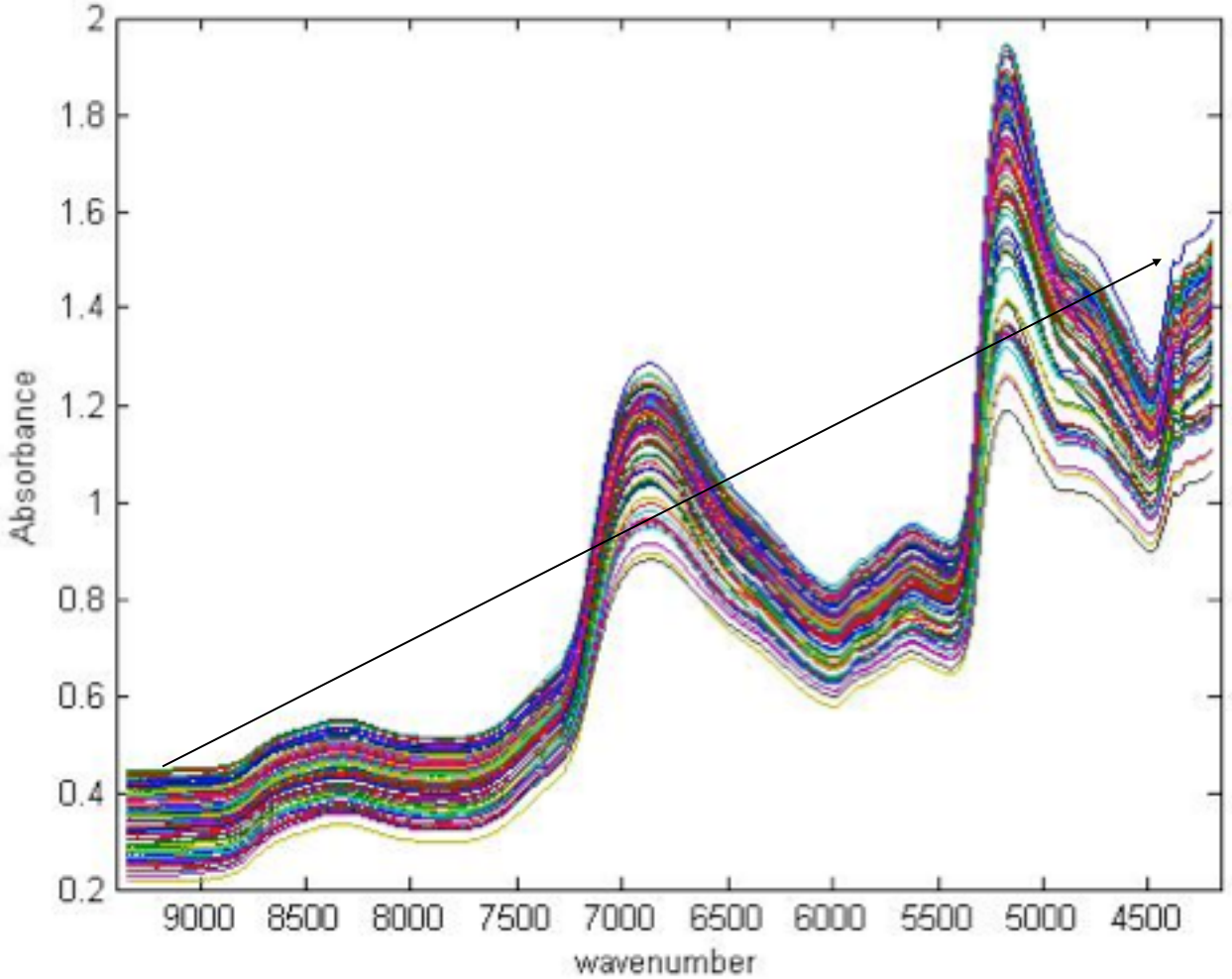


Temperature
humidity
turbidity
…

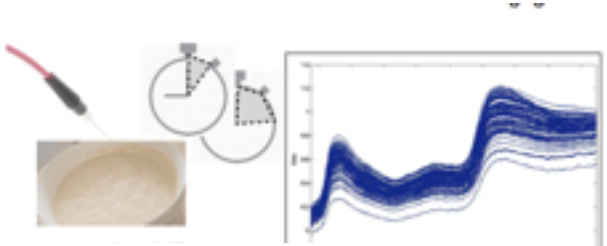


NIR variability

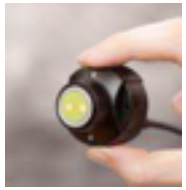
Moreover absorption increase with wavelengths



NIR in situ add further data dimensions

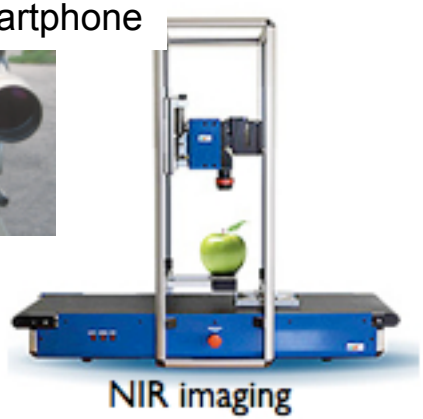


fiber optic NIR probe



portable NIR device

NIR filter on a smartphone



NIR imaging

In situ monitoring allows

- Non destructive characterization
- Process Monitoring
- timely quality assessment

arising Issues

- high dimensional data
- to highlight (or not to) time trends
- not always good resolution

- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

1. Handle variability (preprocessing)
2. Monitoring raw materials
3. Process evolving with time
4. Treatments Effect
5. Multispectral/Hyperspectral imaging

Chemometrics to extract relevant information

..allows the study and exploration of data sets characterized by a huge number of variables, through a significant help in the extraction of useful information, enhanced ways of visualization and a general simplification in the interpretation of results.

Sampling

experimental design

Data Acquisition

optimization of measurement procedure

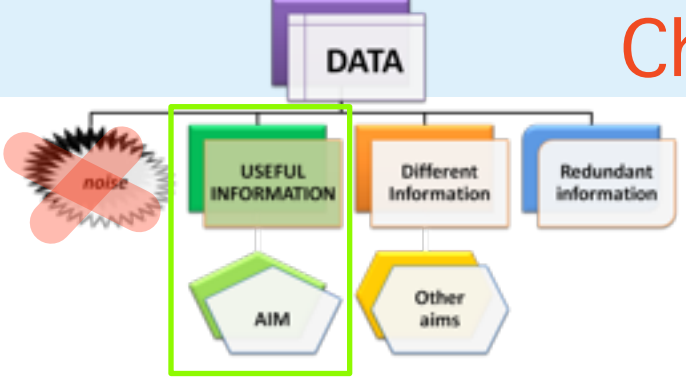
Gather Information:

- Extract features/ time trend /..
- Categories assessment
- Properties prediction
- Chemical Resolution

Data Analysis

- Explorative Data Analysis
- Preprocessing
- Modeling
- Validation

Chemometrics Tools



To address variability:

- Sampling / Design
- Preprocessing
- Explorative Multivariate Data Analysis (PCA)

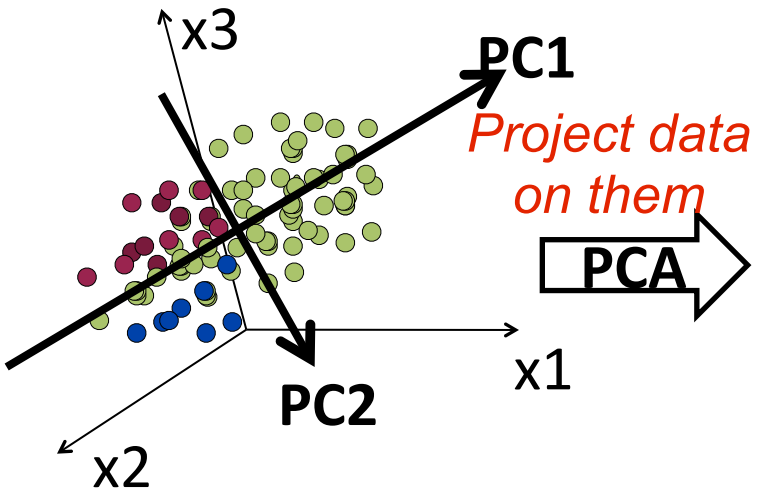
To extract information:

- Explorative Multivariate Data Analysis (PCA)
- Resolve components (MCR)
- Study trends, e.g., time, location,.. (MCR, Multiway)
- Process Monitoring (Multivariate Control Charts)
- Quantify (Multivariate Calibration)
- Assess categories (Class Modeling , Classification)

Unsupervised

Principal Components Analysis (PCA)

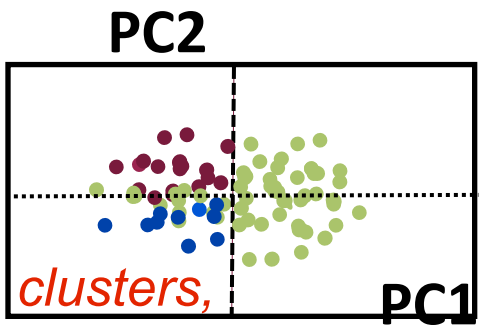
find directions of max variability



Project data on them

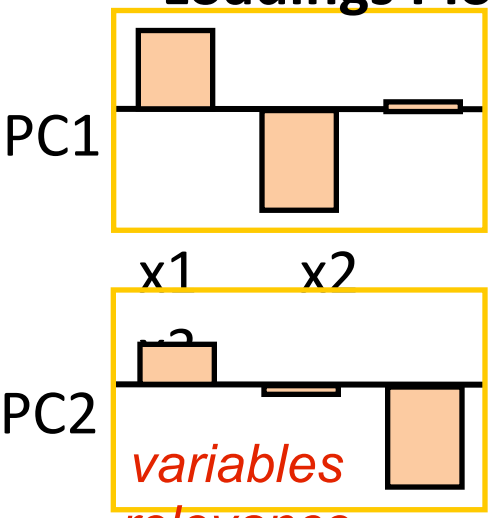
PCA

Scores Plot

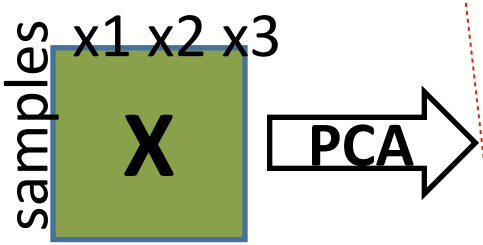


clusters, trend, ...

Loadings Plot

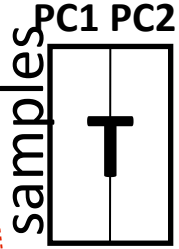


variables relevance



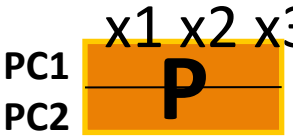
PCA

decompose



Scores

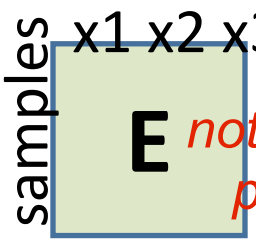
×



Loadings

model of **X**

+

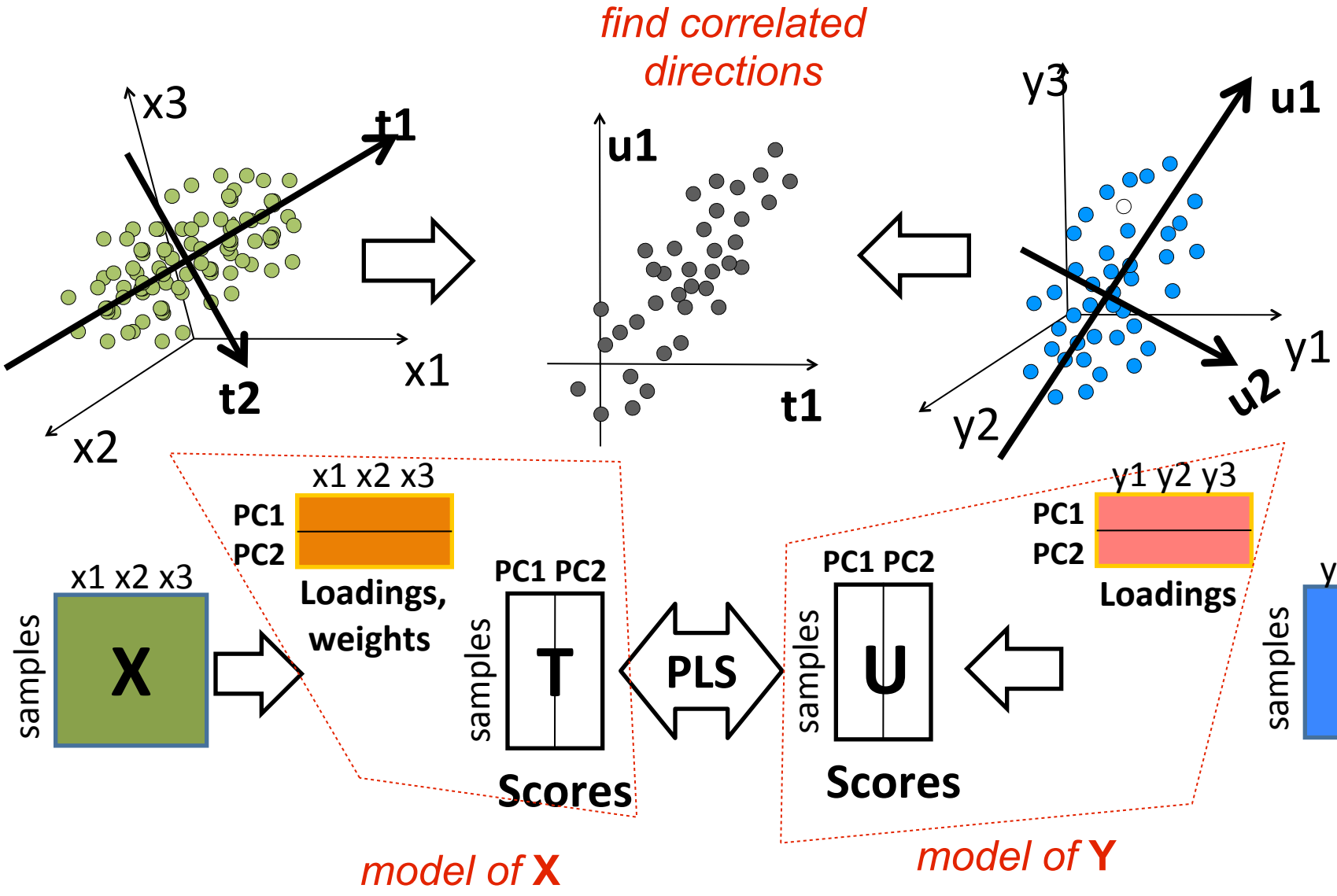


not modeled part of **X**

Residuals

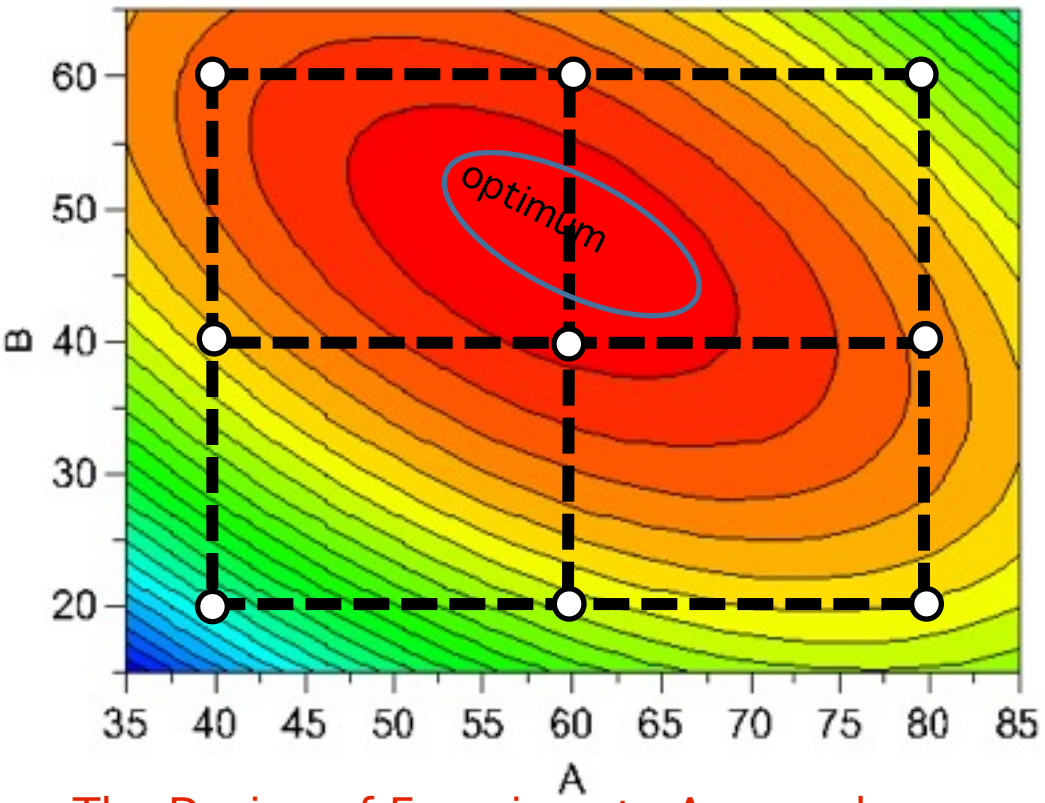
Projection to Latent Structures (PLS)

Supervised



Sampling/Planning : Experimental Design

Planning experiments



Sampling from an existing domain

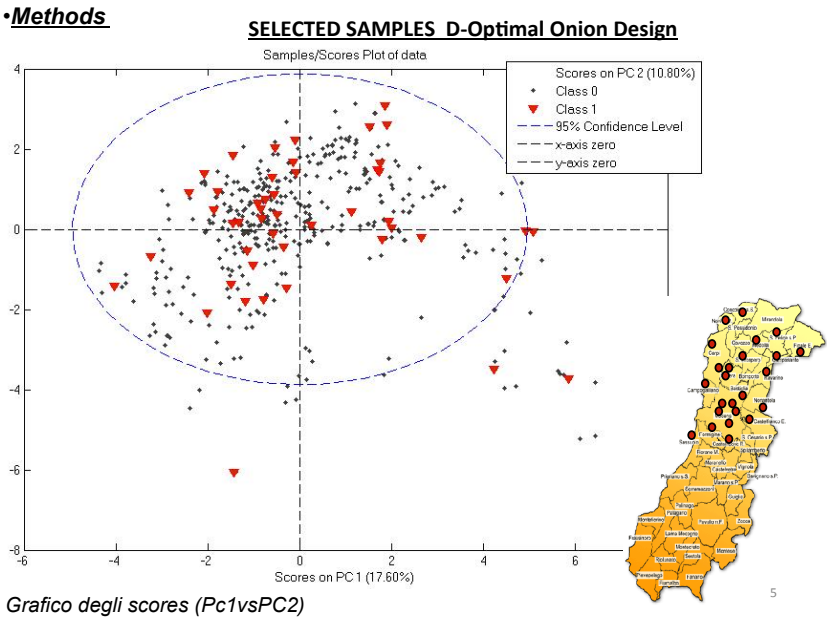


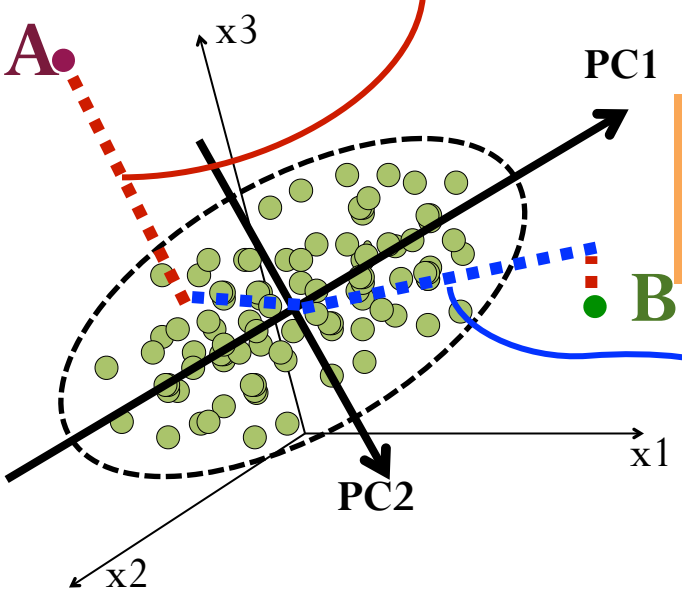
Grafico degli scores (Pc1vsPC2)

The Design of Experiments Approach:

1. Define the experimental domain as variability ranges for A and B
2. Plan the minimum number of experiments to explore most of the domain
3. Obtain a response surface Y=f(A,B,AB,...) to interpolate Y in the whole domain

Principal Components Analysis (PCA)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \sum_{a=1}^{A} t_a p_a^T + \mathbf{E}$$



Q-distance

Distance of an object from the model

$$Q_i = \mathbf{e}_i \mathbf{e}_i{}^{\mathrm{T}}$$

High residuals → high not-explained variability

Sample distances from the model space and within the model are indexes of its peculiarity in a multivariate way
→ they can be used to build Control Charts

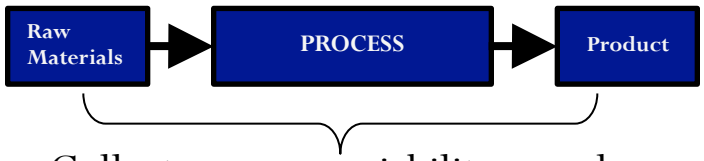
T^2-distance

Distance of an object within the model

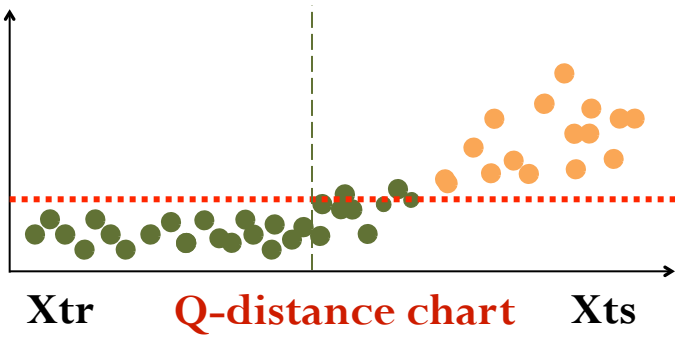
$$T^2{}_i = \mathbf{t}_i \lambda^{-1} \mathbf{t}_i{}^{\mathrm{T}}$$

High T^2 → extreme behavior in the model

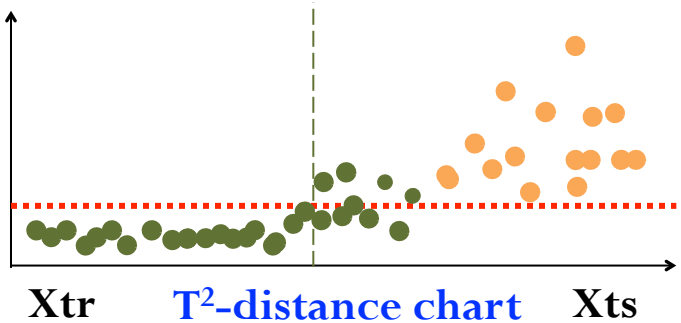
Chemometrics Tools: Multivariate Control Charts

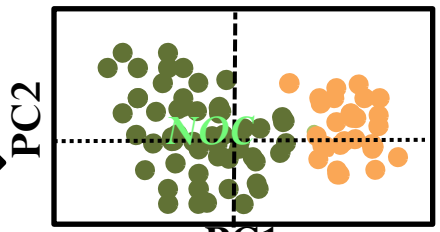


Collect process variability samples (historical data, DoE, …)



future routine samples distances to the model are used to access their compliance to NOC





Identify **Normal Operative Conditions** by means of Multivariate Analysis

Xtr

Representative set of NOC samples

Xts

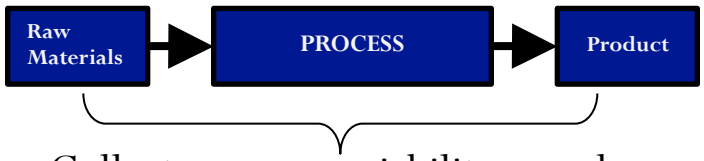
Set of both NOC and fault samples

PCA model on Xtr

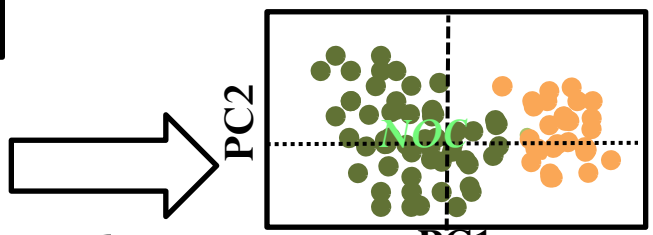
build Multivariate Control Charts based on Q and T^2 distances to the model and obtain critical limits

VALIDATE by projecting Xts and checking the behaviour of NOC samples

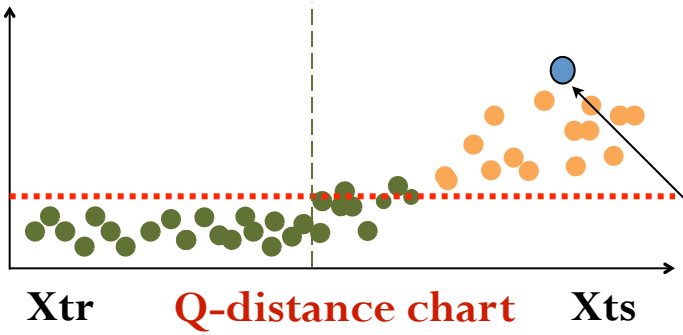
Chemometrics Tools: Multivariate Control Charts



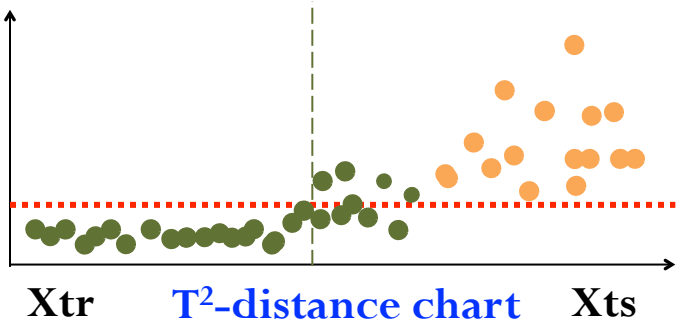
Collect process variability samples (historical data, DoE, …)



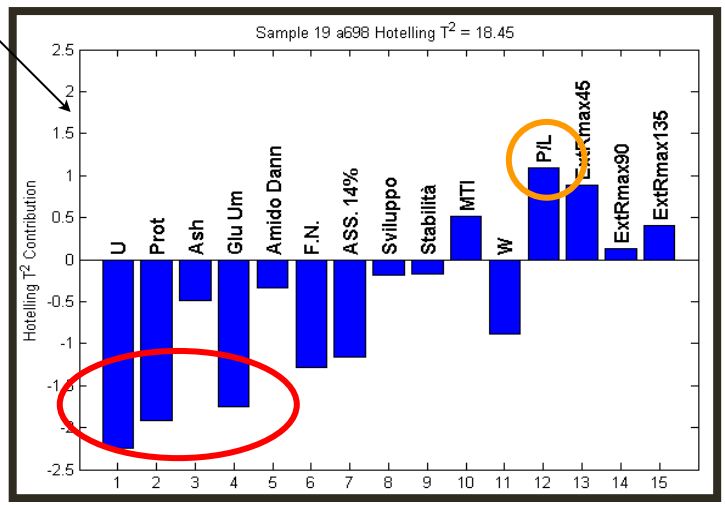
Identify **Normal Operative Conditions** by means of Multivariate Analysis



future routine samples distances to the model are used to access their compliance to NOC



Contribution Plot



Explorative Data Analysis (PCA) Unsupervised

Mind: pretreatments

Supervised

Multivariate Calibration, Regression (MLR, PLS, ...*ANN*)

Classification (SIMCA, LDA, PLS-DA..... *ANN*)

Mind: n° samples / n° parameters to estimate

Multivariate Control Charts

Mind: sampling / reference distributions

VALIDATION !!!!

- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

    1. Handle variability (preprocessing)
    2. Monitoring raw materials
    3. Process evolving with time
    4. Treatments Effect
    5. Multispectral/Hyperspectral imaging

Handle variability

Two approaches (also complementary) to solve the problem

Preprocessing

- Smoothing
- Background / baseline
- vertical shift
- bands deconvolution

Feauture selection

- single variables (stepwise-MLR, UVE-PLS, GA, ….)
- intervals (I-PLS, …)
- features selection in a different domain (FT, WT)

Aim:

To correct for/ remove the contribution of undesired phenomena from stochastic measurement noise to various sources of systematic errors: nonlinear instrument responses, shift problems, and interfering effects of undesired chemical and physical variations.

Preprocessing Issues

➢Preprocessing the samples (the spectra)

✓ **Smoothing**: removing uninformative high frequency variation, such as noise (moving average, polynomial, **SG-filter**, *FT-filter*, WT-filter)

✓ **Normalizing** (vertical shift): put on the same scale
(to unit area, to unit length, to Imax, SNV) ! Mind: introduce closure

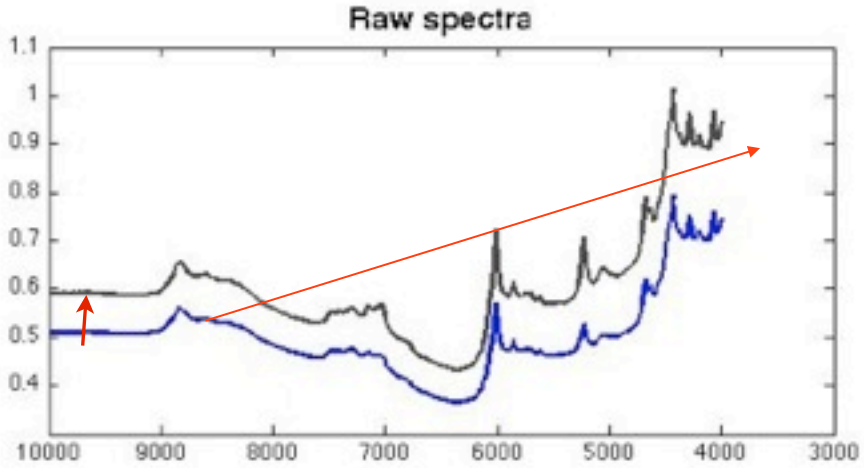
✓ **Baseline/Background correction**: removing uninformative low frequency variation, *removing scattering*
(explicit modeling, **derivatives, MSC, EMSC**, OSC, OPLS**..**)

✓ Deconvolve bands: find hidden components
(derivatives, Transform Methods (FT, WT), MCR, OPLS..)

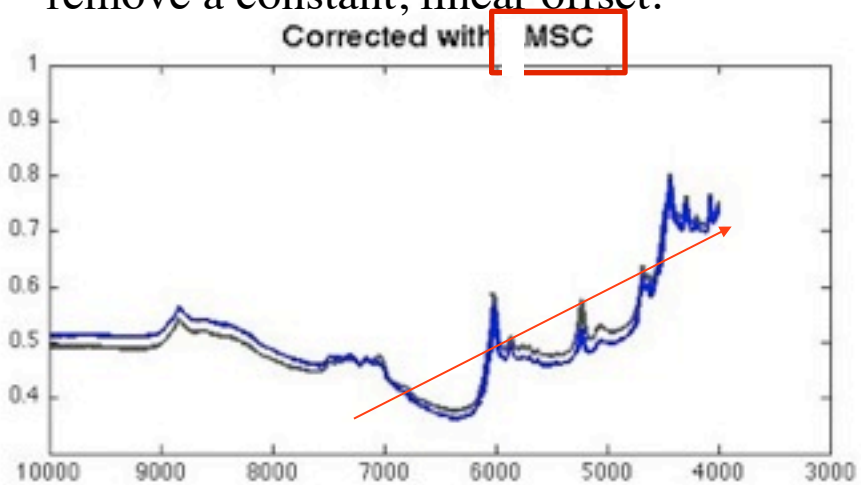
➢ Preprocessing/Pretreatment the variables, i.e. columns centering

Baseline/background correction

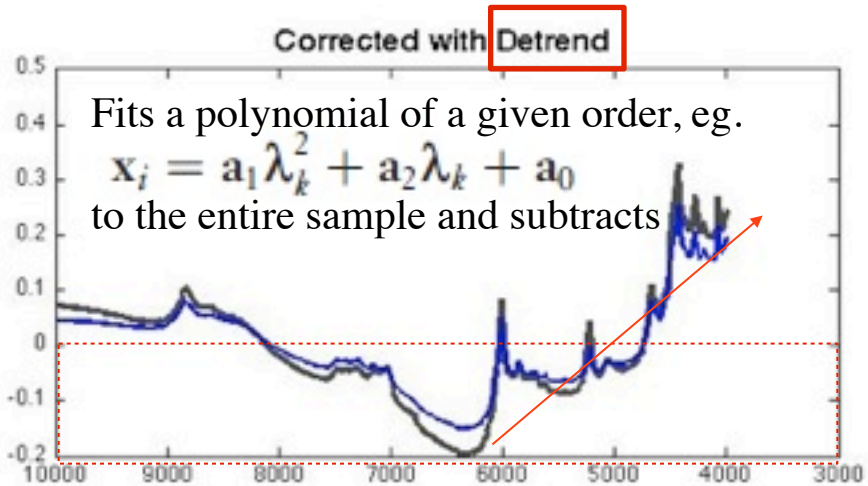
-- = average spectrum
-- = sample 42

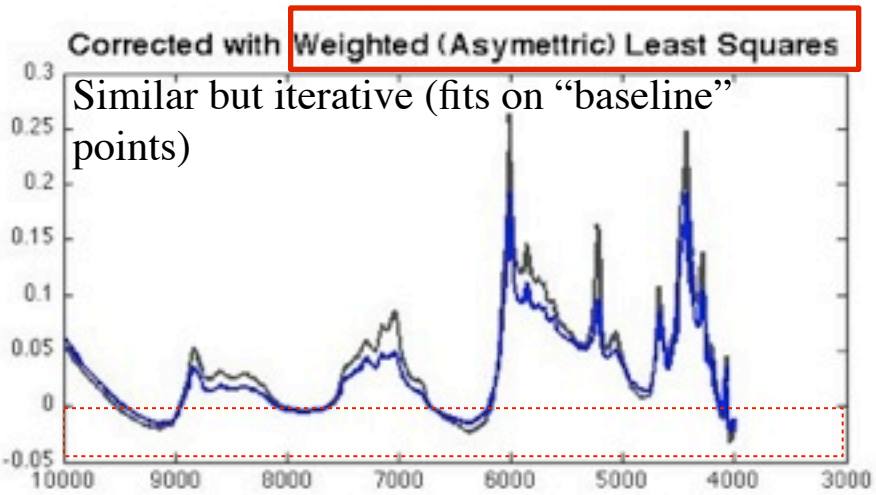


remove a constant, linear offset.



remove a constant, linear, or curved offset.

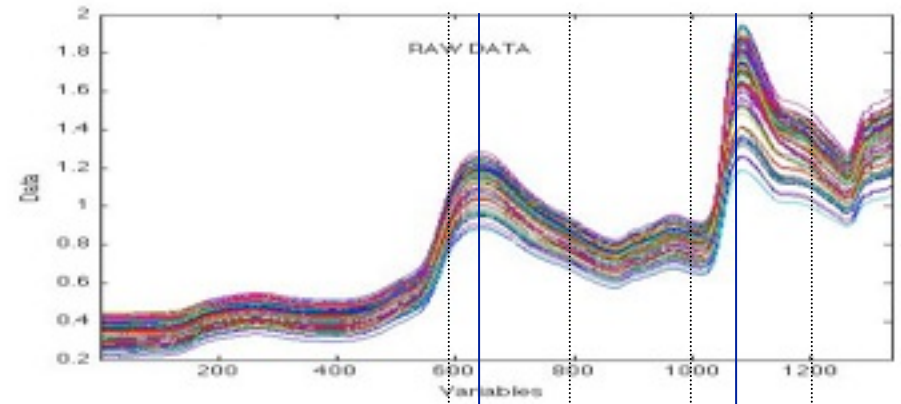




Derivative (Baseline/background correction)

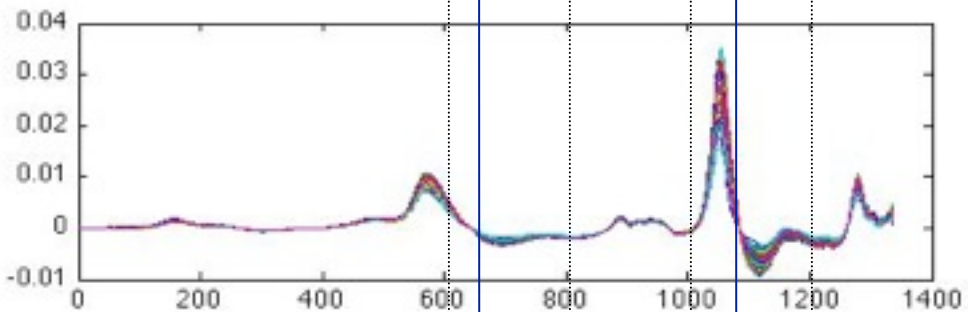
The first derivative allows the additive constant background effects to be removed; the second derivative removes the baseline linear slope variations and additive effects.

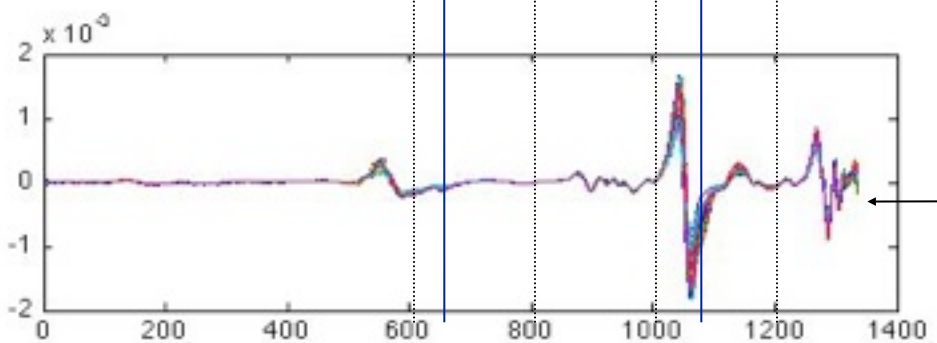
1st 2 nd derivative are usually coupled to smoothing (e.g. Savitzky Golay)



An example with NIR signals of dough

Loosing maxima localization





May be introducing noise

SNV & MSC

When normalization does not work …

i.e. imposing a closure condition (sum of all components is the the same in each sample) in data that do not have

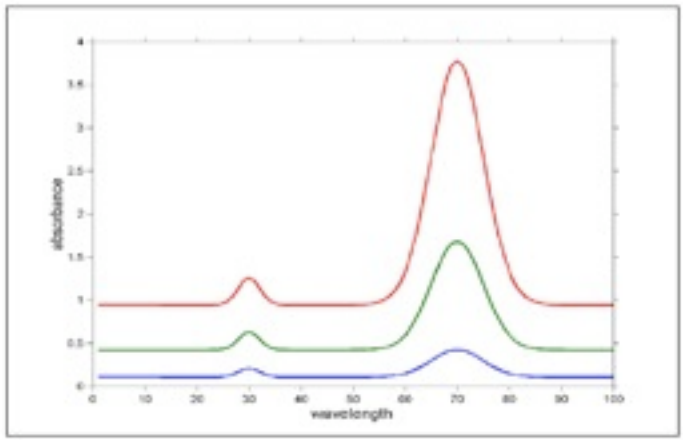


Figure 2. The spectra of Figure 1 with the addition of scatter effects.

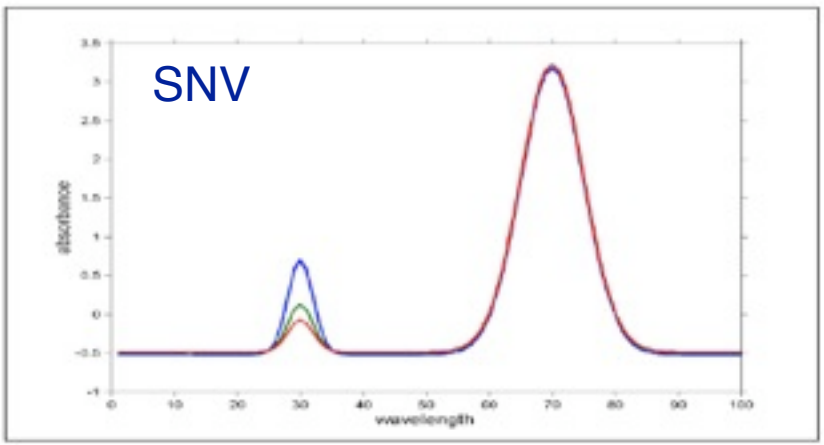


Figure 3. The spectra of Figure 2 after treatment with SNV.

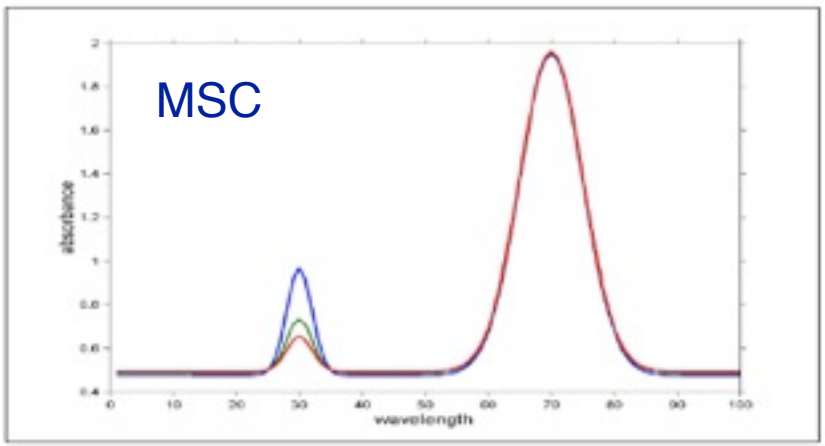


Figure 4. The spectra of Figure 2 after treatment with MSC.

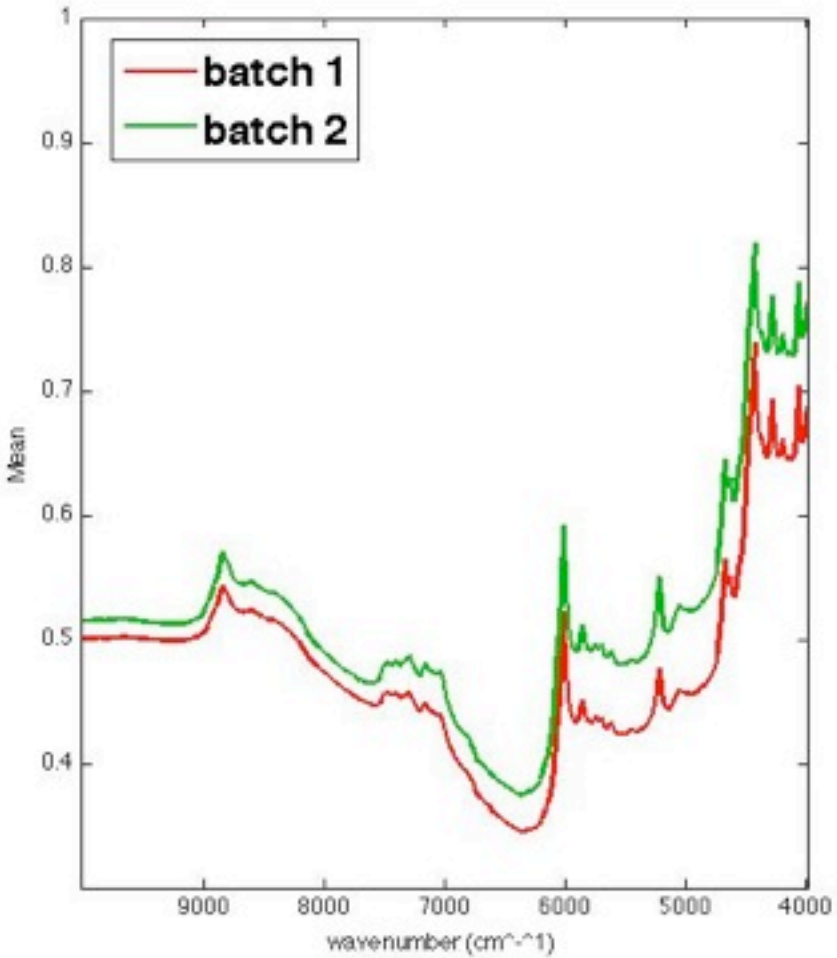
Since there is a peak much larger than baseline……

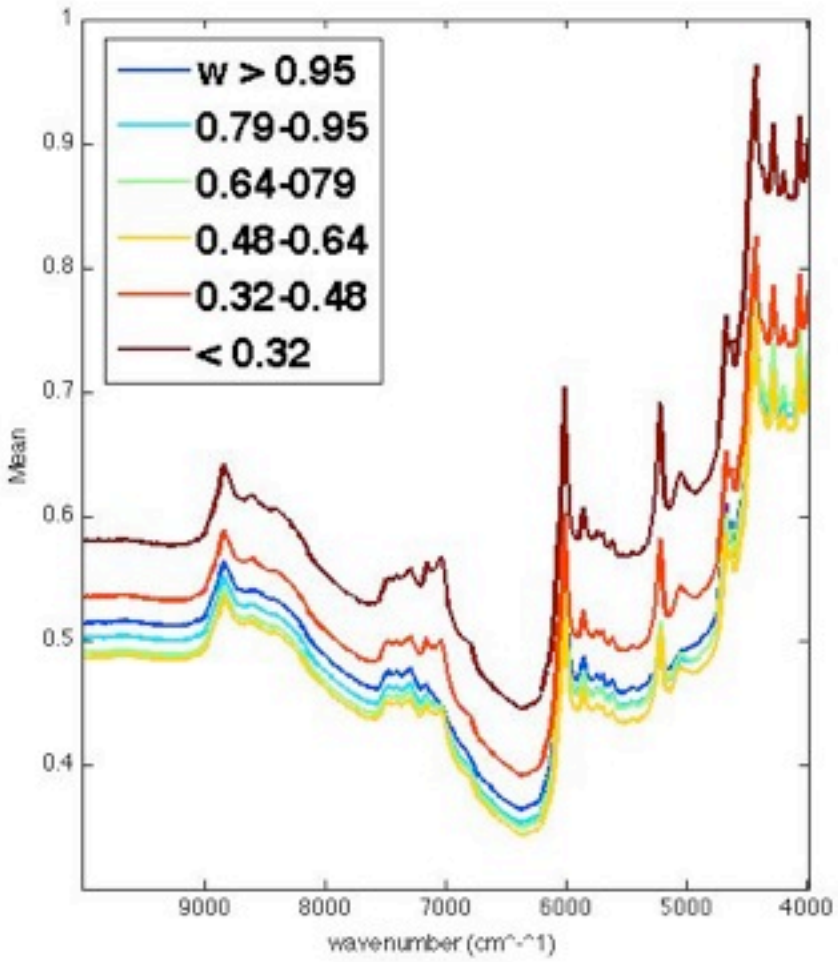
The largest peaks are scaled so that they tend to coincide. MSC scales spectra to make them similar, SNV tries to make vertical range the same for all spectra

how to choose preprocessing

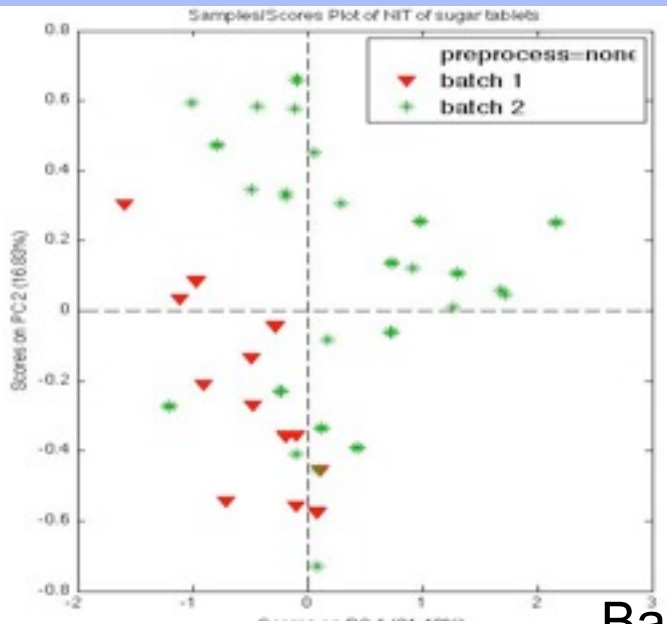
a NIT example: water calibration

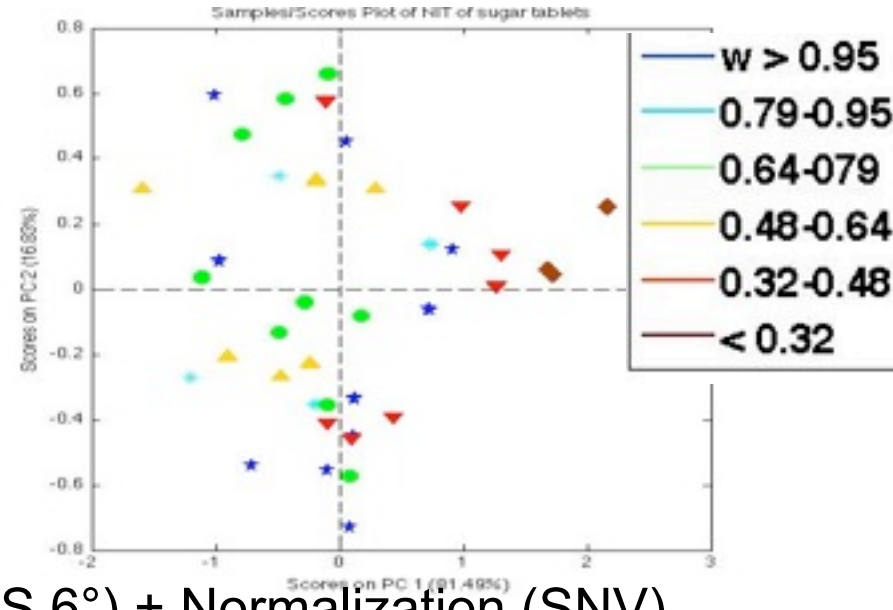
47 compresse sono state monitorate a campione su 2 batch successivi di produzione.
Il contenuto d'acqua (w) è stato determinato per ogni complessa con metodo indipendente.
NIR operante in trasmittanza, equipaggiato con un dispositivo di campionamento per solidi (compresse).
Il range spettrale coperto è da 3996 cm-1 a 10000 cm-1



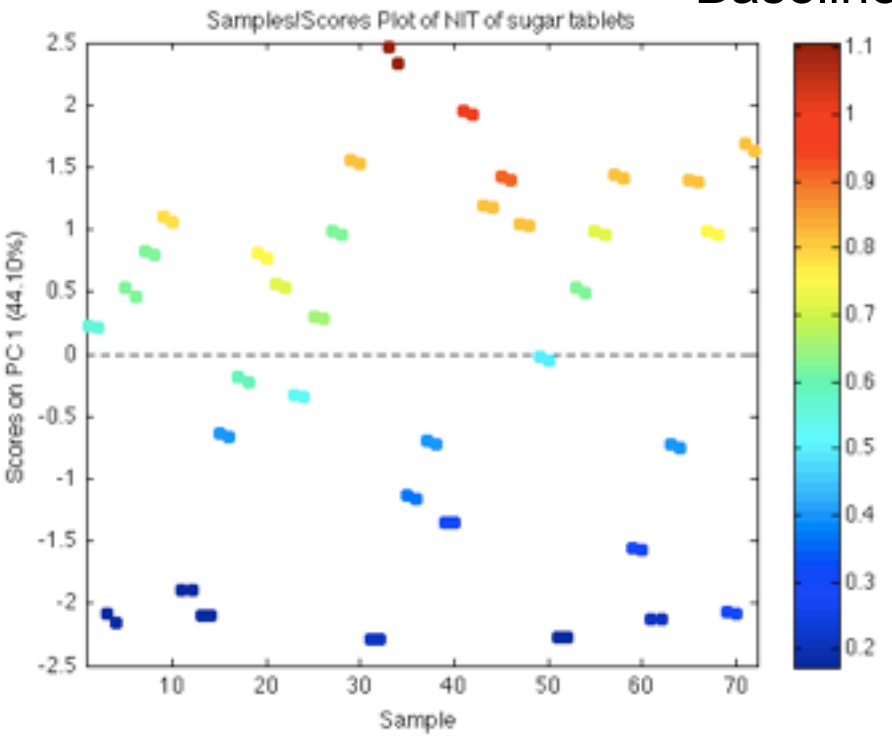


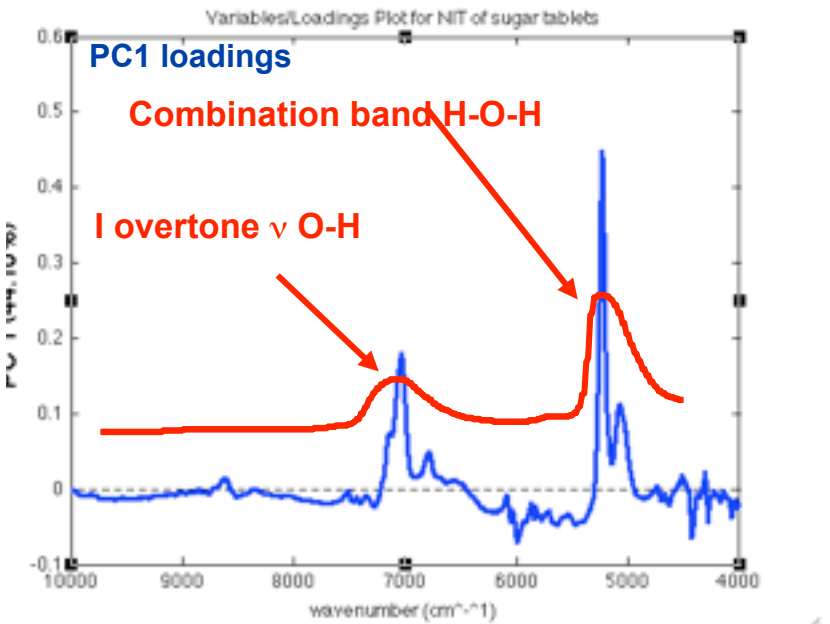
a NIT example: preprocessing solve batches variability issue





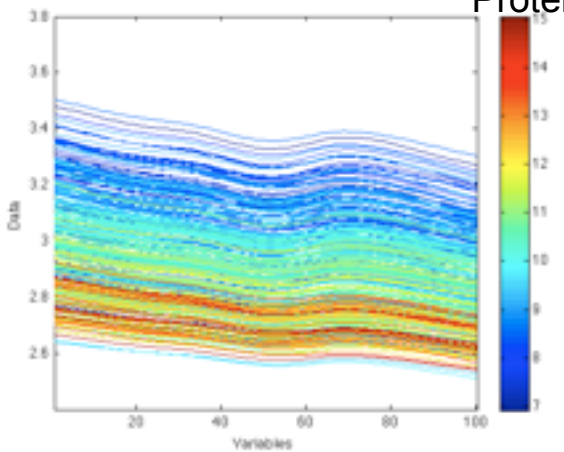
Baseline (WLS 6°) + Normalization (SNV)



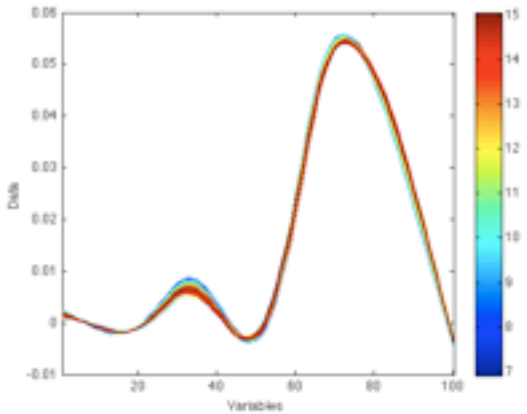


a NIR example: preprocessing solve variability issue

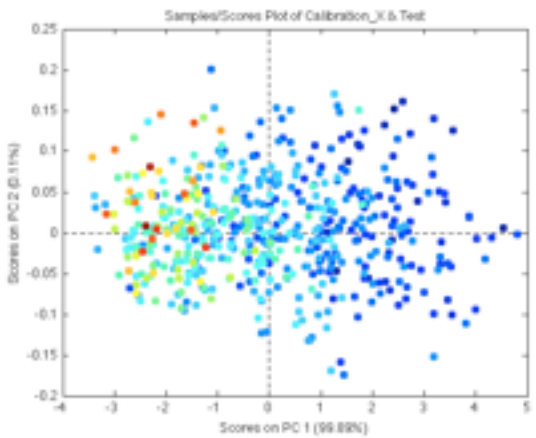
Raw: wheat seed sample (43 varieties 2 location)
validation set projected on
Protein content

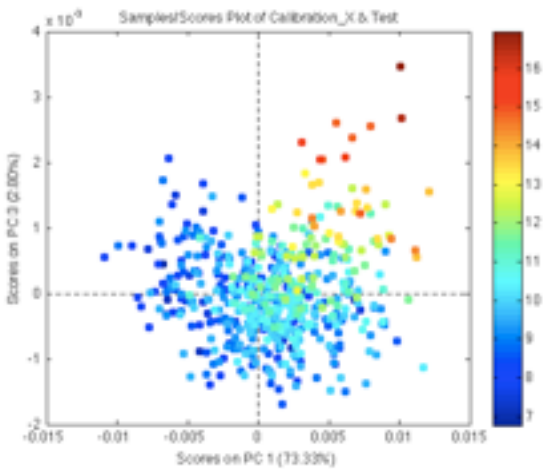


Baseline and normalized (wls + MSC)



Explorative Analysis PCA





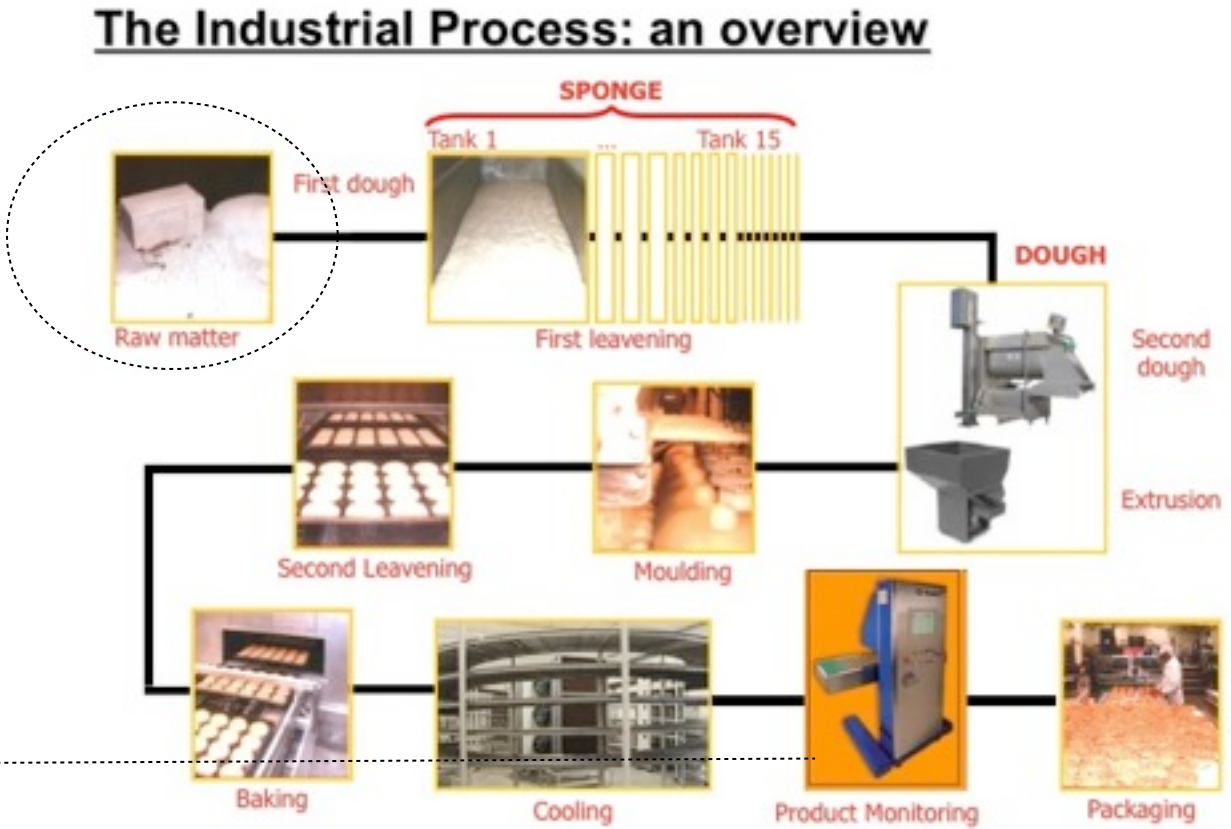
- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

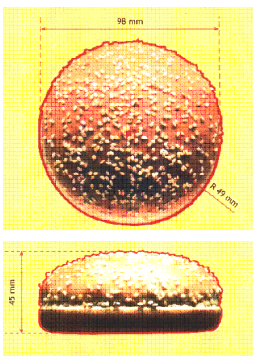
    1. Handle variability (preprocessing)
    2. Monitoring raw materials
    3. Process evolving with time
    4. Treatments Effect
    5. Multispectral/Hyperspectral imaging

Raw material

1. MONITORING RAW MATERIAL

Industrial bread production





size, color

M. LI VIGNI, C. DURANTE, G. FOCA, A. MARCHETTI, A. ULRICI AND M. COCCHI, Near Infrared Spectroscopy and multivariate analysis methods for monitoring flour performance in an industrial bread-making process ANAL. CHIM. ACTA, 642 (2009), 69-76.

Monitoring of raw material quality: Industrial Bread Production

The Aim

Build control chart to monitor incoming flours

In this way it could be known in advance if Flour "quality" it is different from desired and why

Could NIR be suitable to this aim ?

The main challenge: NOC ?!!

- to be representative

use all available sampling periods/ wheat mixture

- **well behaving**

???????????

The "historical" data

chemical/technological flour properties;

NIR spectra (not for all batches); Bread quality

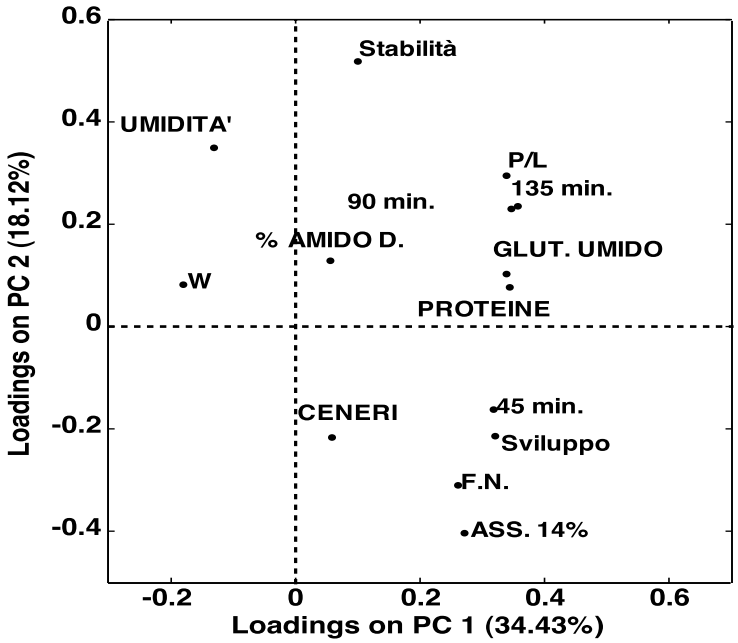
Harvest	Date	Deliveries	Batches
2007	20/03/07 – 19/05/07	16	4
	12/05/07 – 11/09/07	53	13
	28/07/07 – 09/04/08	91	28
	06/06/08 – 30/07/08	16	7
2008	22/07/08 – 14/08/08	6	2
	08/08/08 – 02/09/08	9	3
	04/09/08 – 17/09/08	7	2
	18/09/08 – 30/12/08	41	10
2009	03/01/09 – 07/03/09	30	6
	Total	269	75

Raw material

Monitoring of raw material quality: Industrial Bread Production

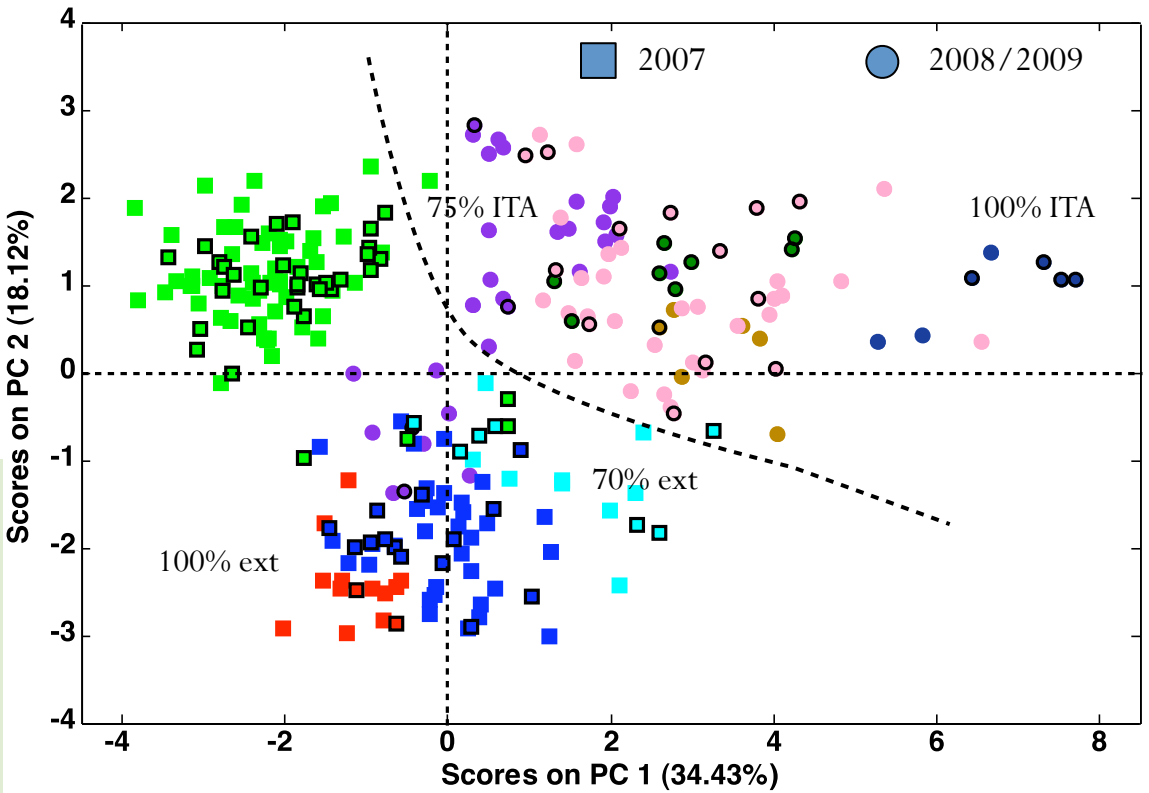
The results

Exploratory Data Analysis
PCA on chosen NOC (test set projected)



flour batches behave in different ways depending on harvesting year and wheat mixtures
PCA shows that both NOC and test set batches (symbols with black boundaries) cover all the observed variability

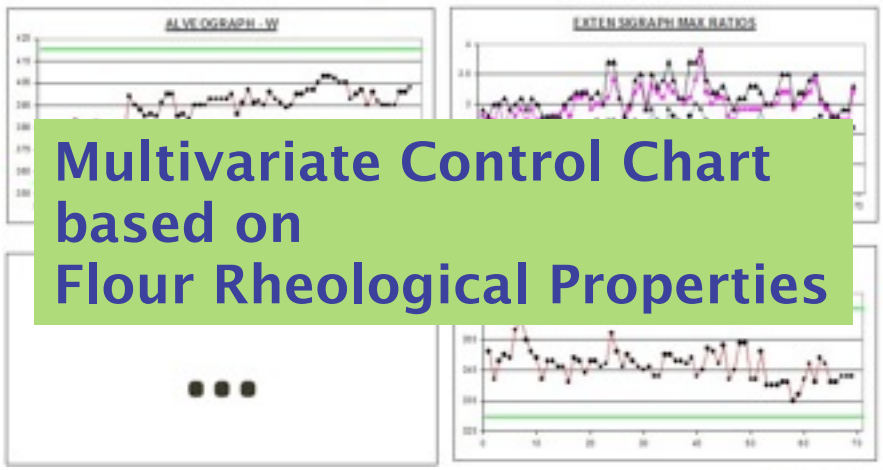
Loadings variables space



Monitoring of raw material quality: Industrial Bread Production

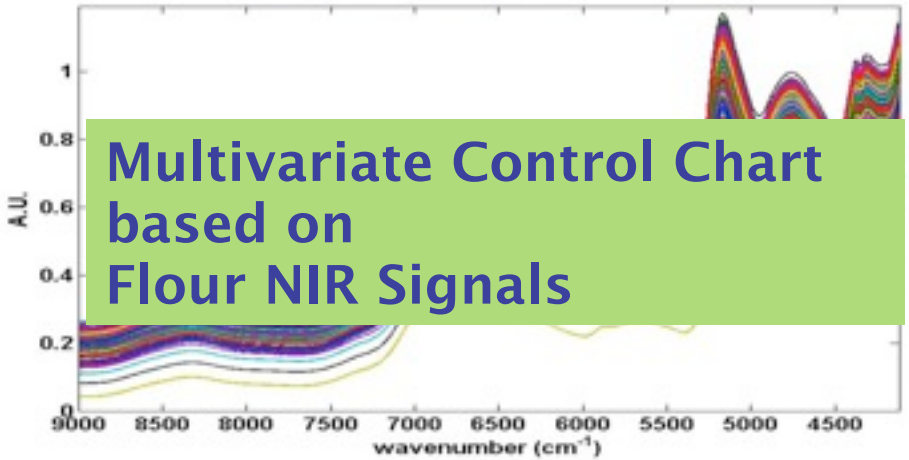
The results

Monitoring Flour Properties



Multivariate Control Chart based on Flour Rheological Properties

Monitoring Flour with NIR



Multivariate Control Chart based on Flour NIR Signals

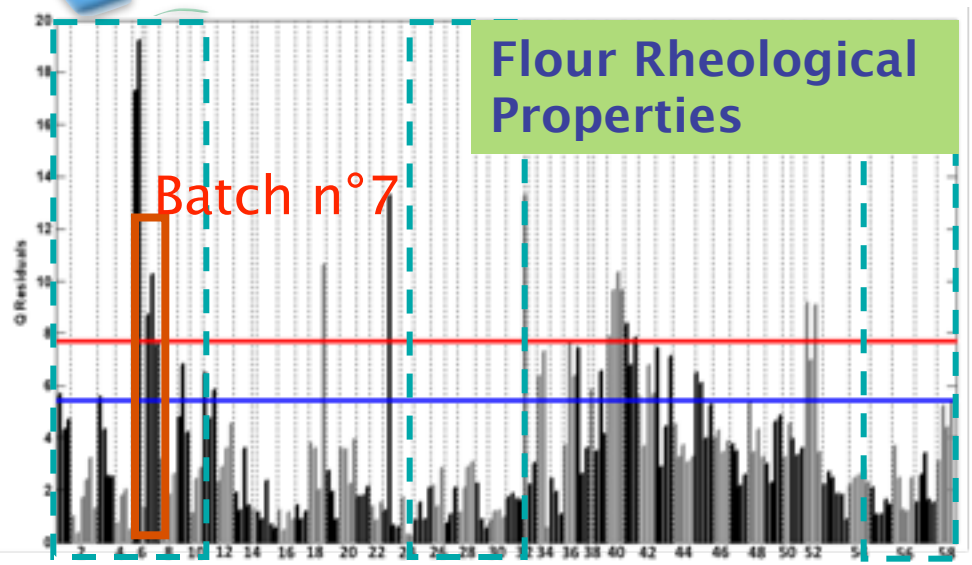
Monitoring Bread Quality

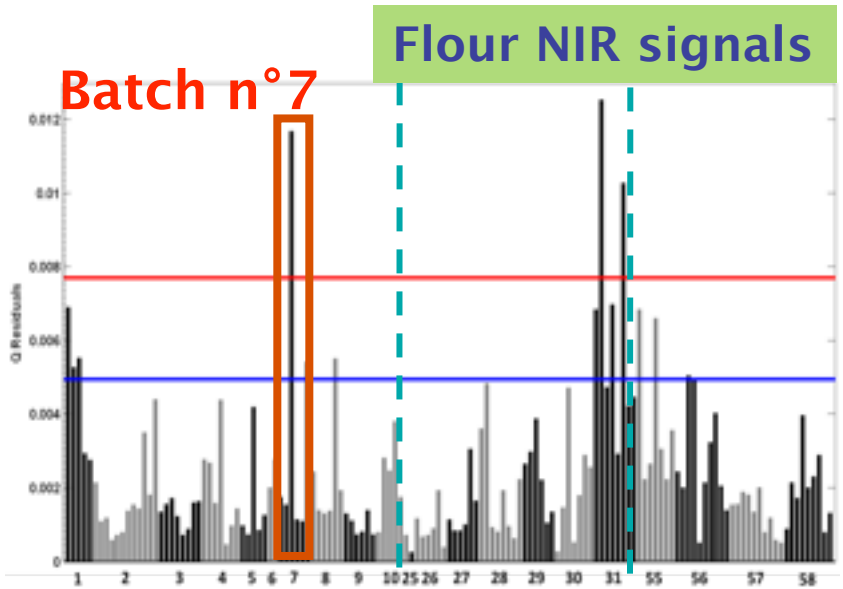
Multivariate Evaluation of Bread Quality

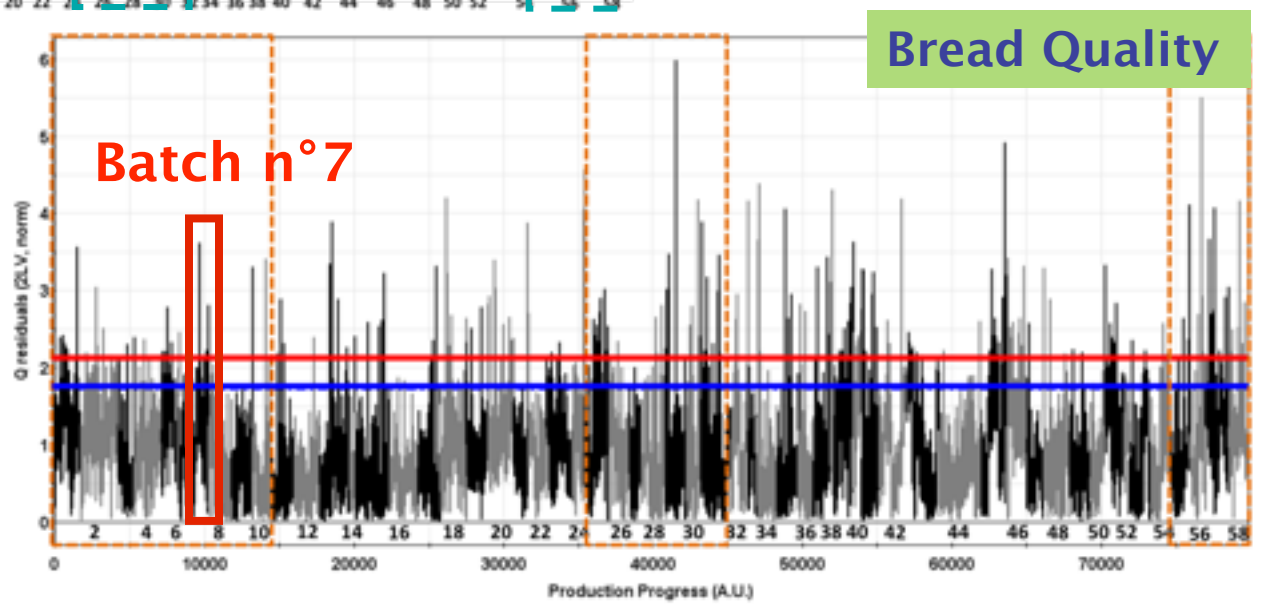
Raw material

Monitoring of raw material quality: Industrial Bread Production

The results



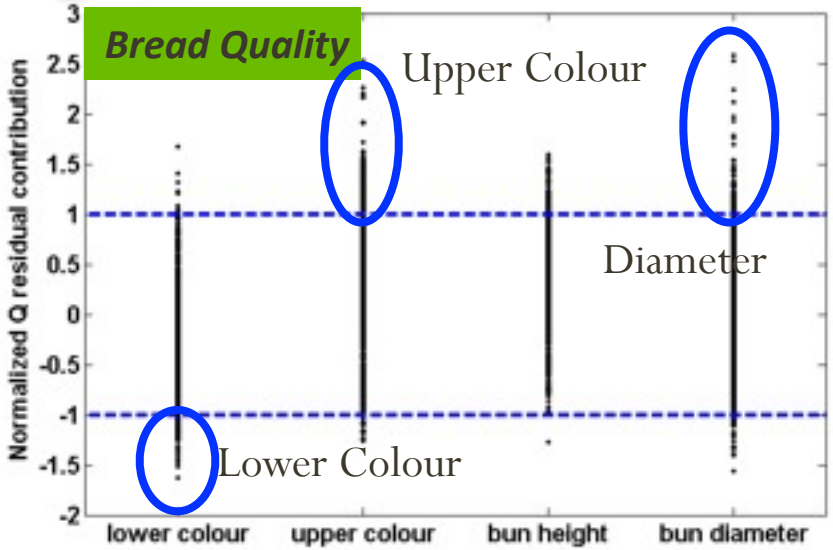


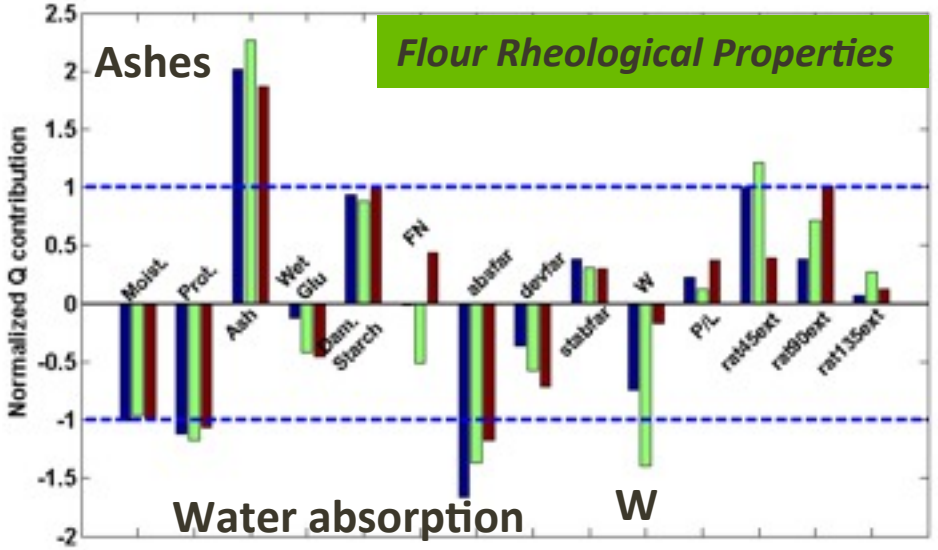


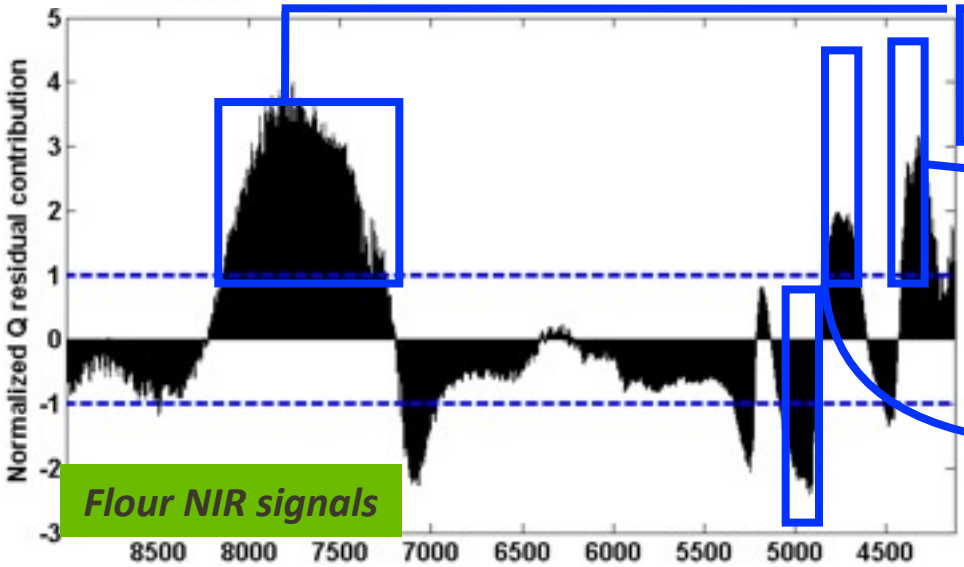
Monitoring of raw material quality: Industrial Bread Production

The results

Contribution Plots for Batch n°7







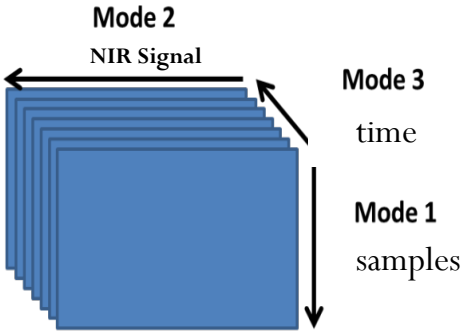
8200–7400 cm^{-1}
C–H 2nd ovt. + comb.

4400–4250 cm^{-1}
starch & protein modes

5285–4990 cm^{-1}
(C=O str. 2nd ovt.)
4760 cm^{-1}
(O-H bnd./C=O str. comb.)

- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

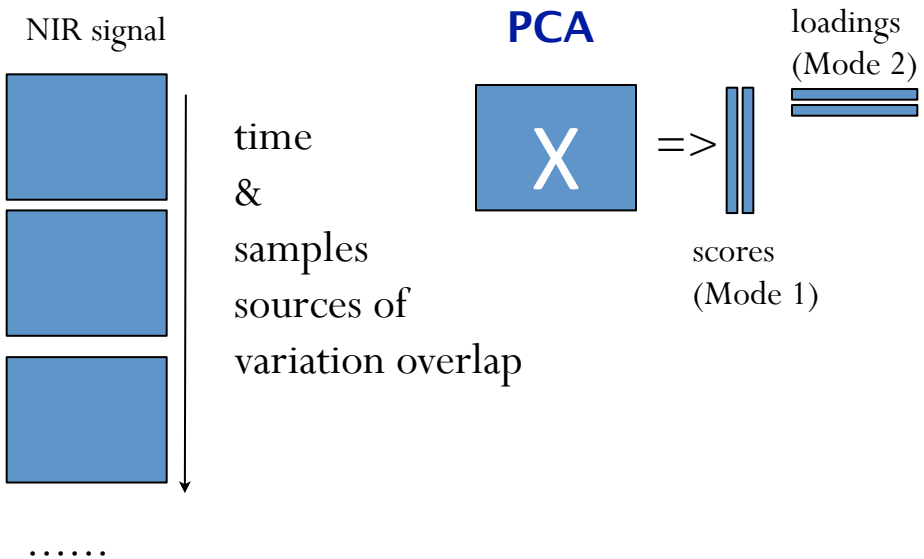
1. Handle variability (preprocessing)
2. Monitoring raw materials
3. Process evolving with time
4. Treatments Effect
5. Multispectral/Hyperspectral imaging



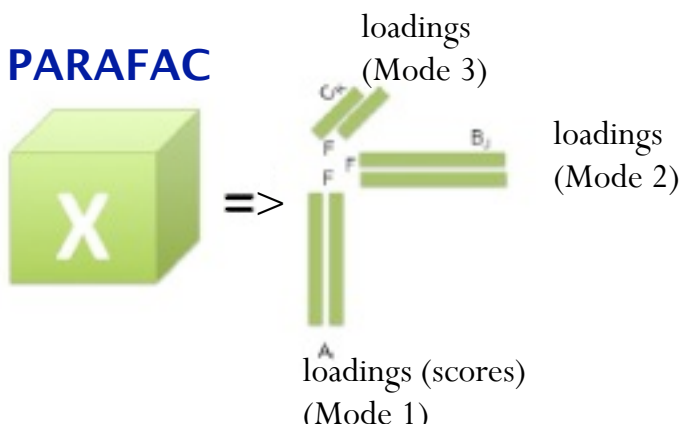
- To highlight trend
  - Explorative Data Analysis
  - Extract features/ time trend /..

How to Model ?

- **unfold (sub-optimal)**



- **keep structure**



- PARAFAC is "recovering" the profile corresponding to each unique phenomenon that is generating a variance source in the analyzed data, for example, a time profile..

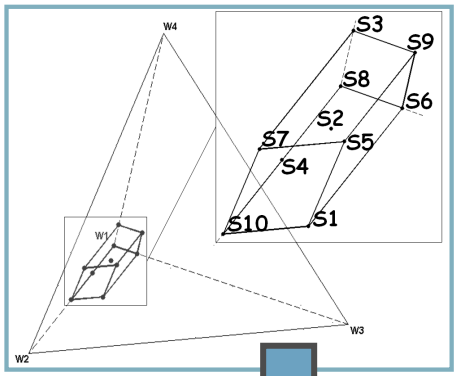
PROCESS EVOLVING WITH TIME: NIR to Monitor dough leavenining

The Aim

Investigation of wheat flour mixture formulation on flour/bread properties

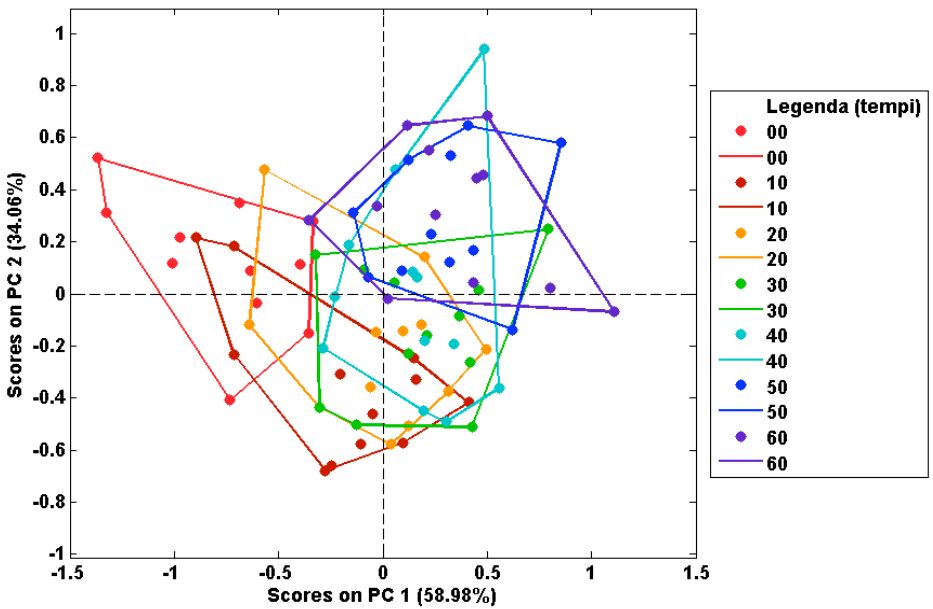
The Data

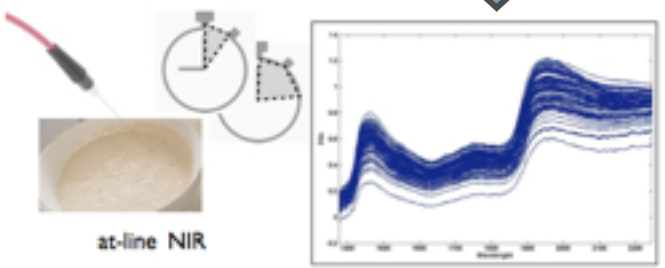
Flours: Mixture of 4 different wheat varieties, obtained by mixture Design of Experiment



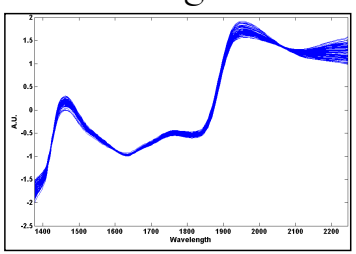
Data Analysis: Compare PCA/PARAFAC

PCA (unfolding sub-optimal)





Smoothing + SNV



Mario Li Vigni, Marina Cocchi, Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties, Anal. Chim. Acta (**2013**), 764, 17-23.

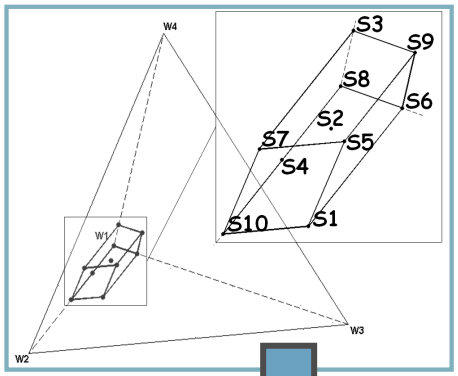
PROCESS EVOLVING WITH TIME: NIR to Monitor dough leavenining

The Aim

Investigation of wheat flour mixture formulation on flour/bread properties

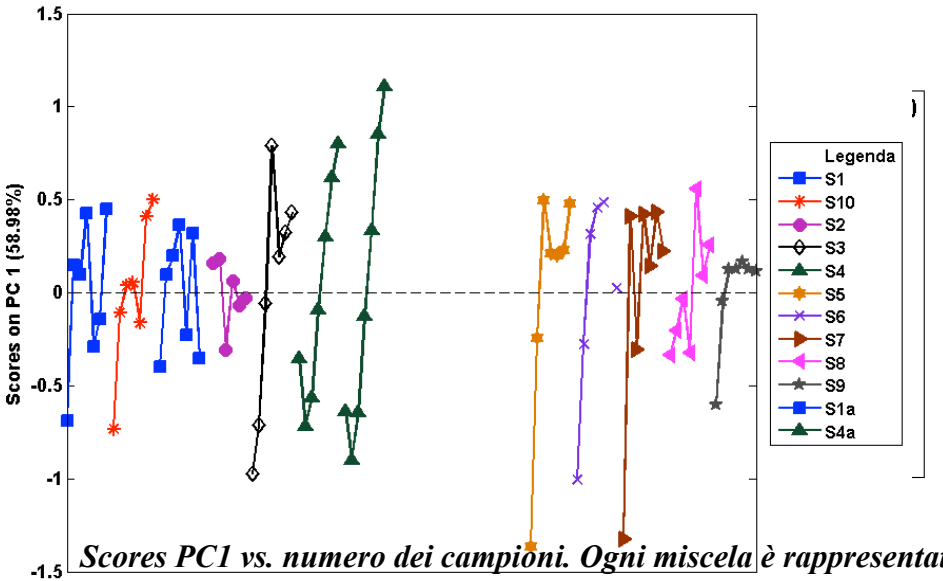
The Data

Flours: Mixture of 4 different wheat varieties, obtained by mixture Design of Experiment

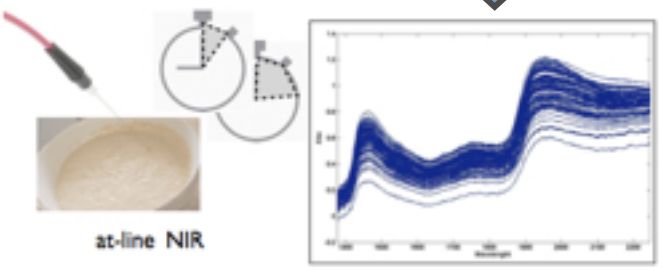


Data Analysis: Compare PCA/PARAFAC

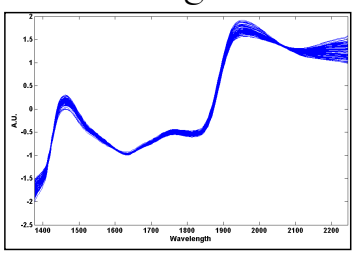
PCA (unfolding sub-optimal)



Scores PC1 vs. numero dei campioni. Ogni miscela è rappresentata da un simbolo/colore diverso, i campioni sono ordinati in base al tempo di lievitazione.



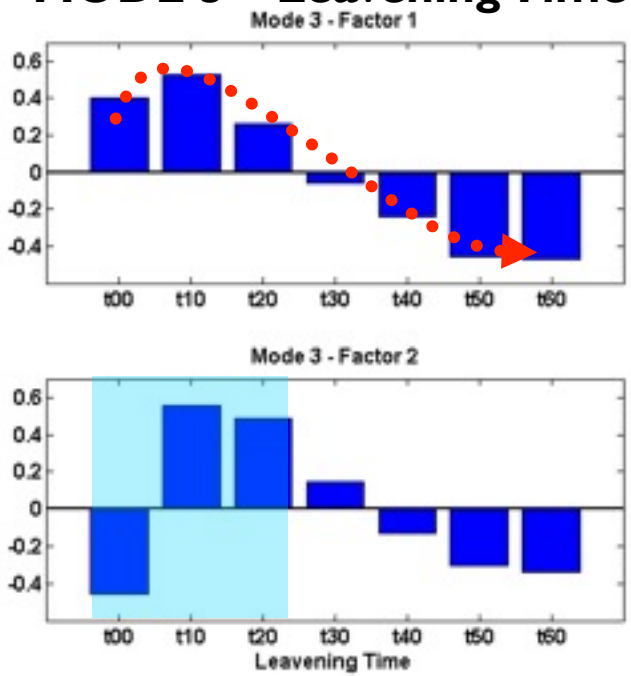
Smoothing + SNV



PARAFAC keep 3D structure!!

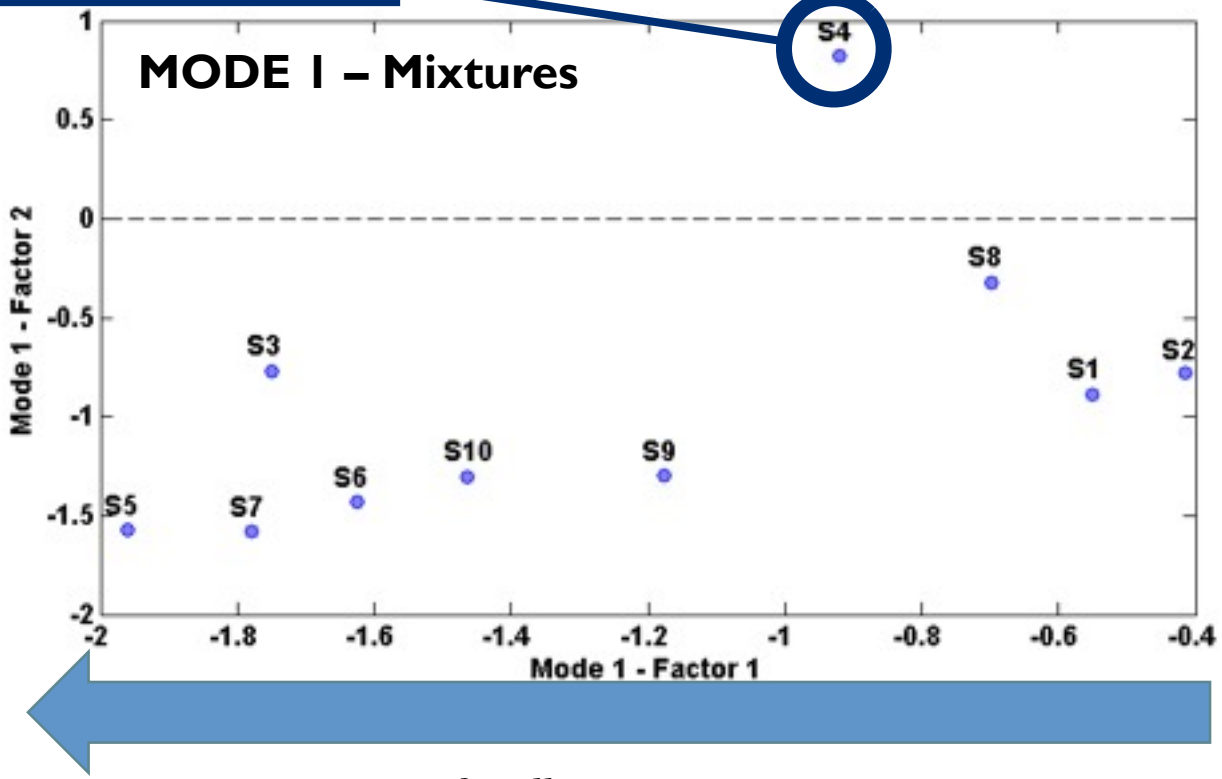
Peculiar behaviour at the early leavening times (Factor 2)

MODE 3 – Leavening Time



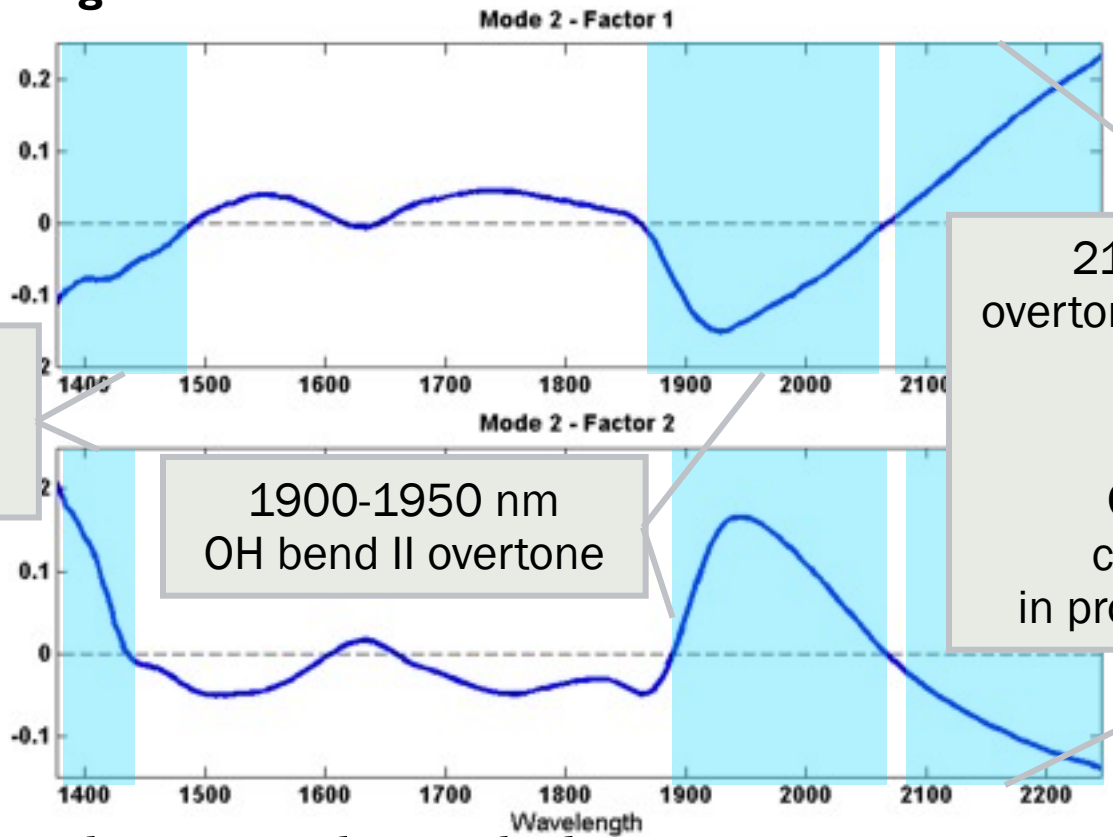
- A common trend for all mixtures
- Times t00 t10 t20 are more relevant

MODE 1 – Mixtures



Same sign for all mixtures on Factor 1: all mixtures present a common trend on Mode 3

MODE 2 – NIR Signal



1380-1480 nm
CH combinations
OH str. I overtone

1900-1950 nm
OH bend II overtone

2100-2250 nm
overtones/combinations
for starch

2180 nm
CH/C=O str.
combinations
in proteins and lipids

NIR contributions can be ascribed to:

- water redistribution in gluten/starch
- formation-consumption of leavening compounds

Wood thermal treatment

Motivation

- Thermally modified wood has become a new commercial product having several advantages over the natural wood.*(treatment process ranging from 24 to 48 hours at temperatures of 180 °C to 230 °C)*
- Chemical-free preservation methods may be preferred, e.g. for the control of pests that may harbored in wood packaging.
  (*ISPM-15 requires a time-temperature schedules that achieves a minimum wood core temperature of 56ºC for a period of at least 30 minutes).*

State of art

- Quite a lot of research has been conducted in order to understand changes to the wood chemistry and wood structure exposed to elevated (150ºC – 250ºC) temperatures.
- Not much attention has been however dedicated for such changes in moderate (50ºC – 150ºC) temperatures

Commercially Relevant questions

1. To have a fast method to estimate if a wood has been treated or not
2. To set the optimal treatment protocol.

***1.* NIR spectroscopy as non-destructive, fast, portable technique**

2. Designed data

M. Cocchi, A. Sandak, J. Sandak, F. Marini

Experimental set up

For each Temperature level

T =1.....n

For each treatment time

time =1.....k

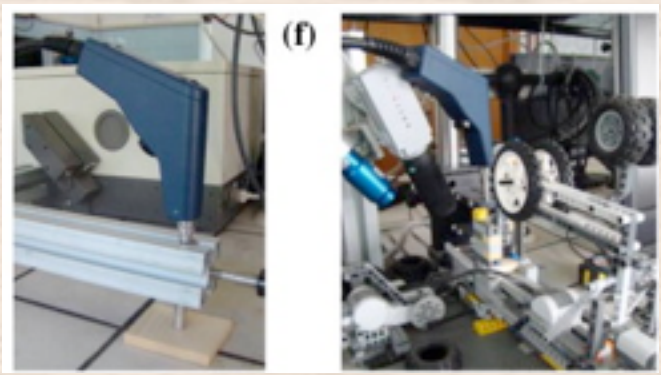
For each storage time

Climate chamber

Measure

40mm x 20mm x 5mm
(length x width x thickness)

Factor	Levels
T °C	40, 50 , 60 , 70 , 80 , 90, 100
t (hours)	1/4 , 1, 3, 6
ts (days)	B (1) C (30) D (90) E (210)



FT-NIR VECTOR 22-N Bruker with fibre-optic probe
Spectral range: 4000 cm-1 - 11000 cm-1 (res: 8 cm-1)

A. Sandak et al. / Construction and Building Materialls (2015), in press

ANALYZING MULTIVARIATE DATA FROM DESIGNED EXPERIMENTS

DoE concept:

– Understand the effect of the controlled factors on the response(s)

– Model the relationship between Y (responses) and X (factors)

performing the minimum number of experiments.

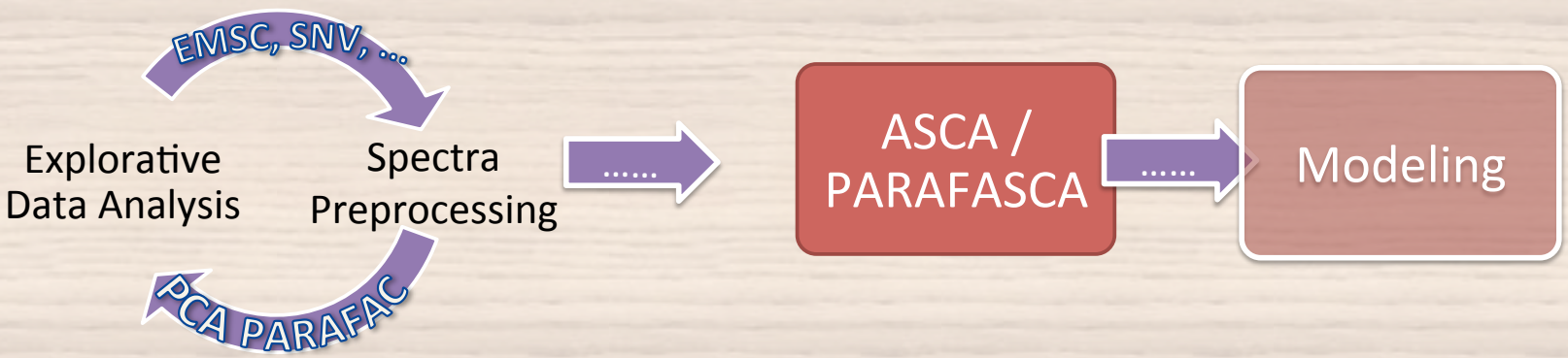
Context here:

– Understand the modification (*effect*), if any, of the controlled factors on the *spectral profile*

  – Best treatment condition

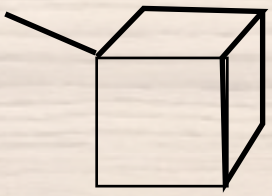
  – Classification Model treated/untreated wood sample from spectra [work in progress]

SUITABLE STRATEGY



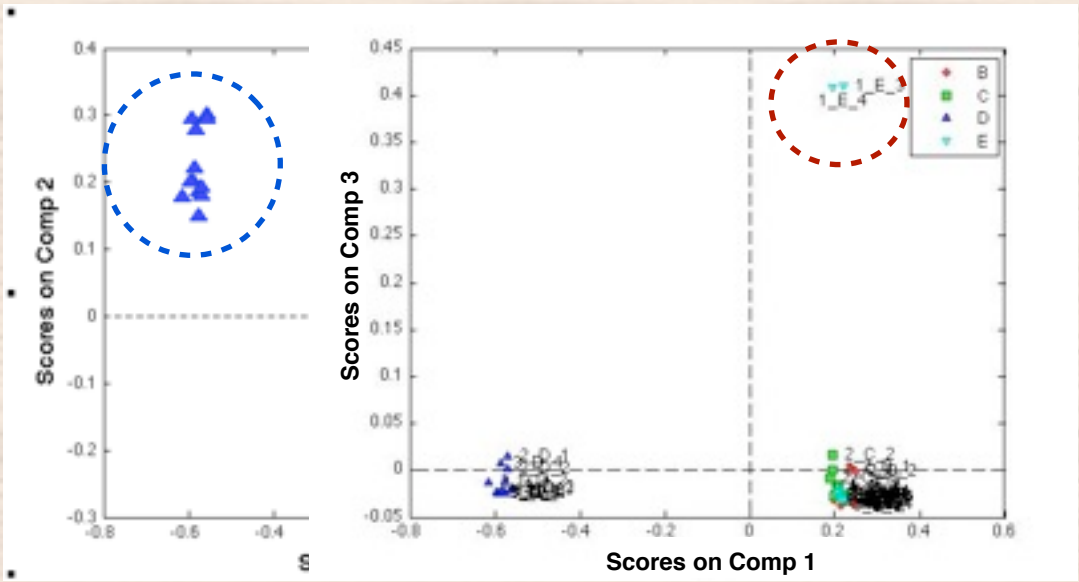
Results - Exploratory Data Analysis

PARAFAC

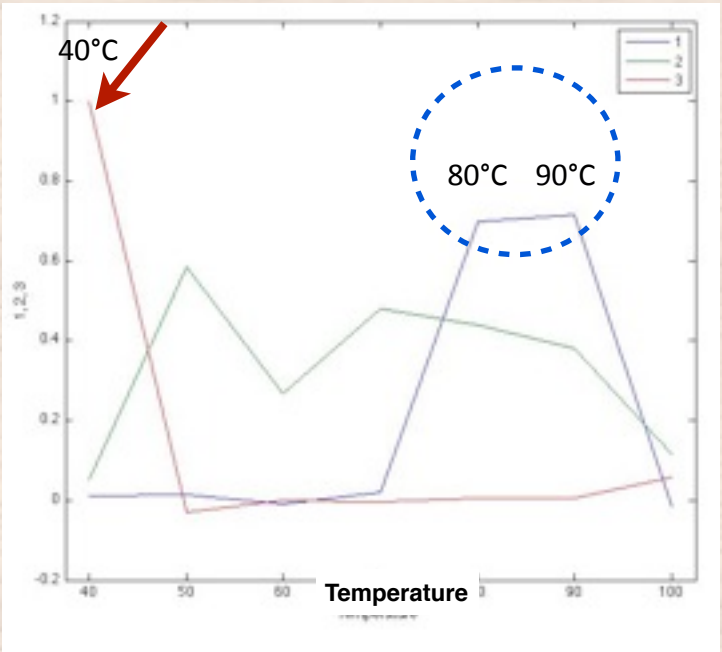


4 way Data array

Mode 1: Replicate; storage time (B,C,D,E)
Mode 2: Spectra (EMSC pret.)
Mode 3: Temperature
Mode 4: treatment time



MODE 3



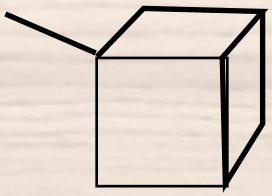
Comp. 1: time storage D from rest because of 80°C, 90°C
Comp. 3: anomalous E samples at T=40°C

EXCLUDE:
all D, 1_E_4, 1_E_3

marina.cocchi@unimore.it

Results - Exploratory Data Analysis

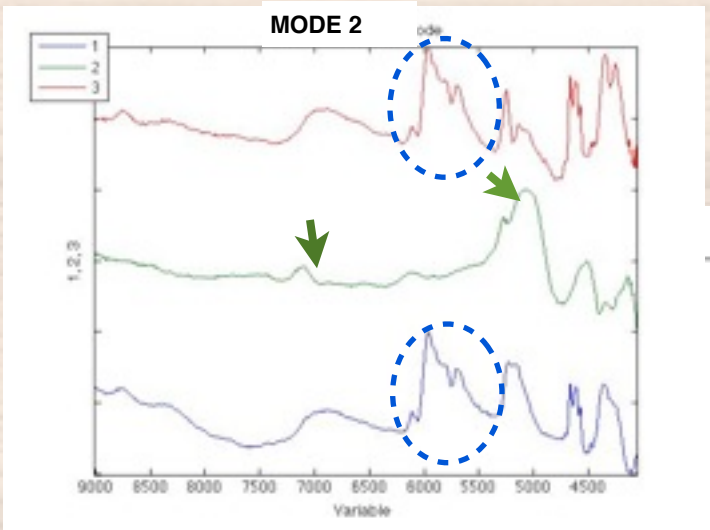
PARAFAC

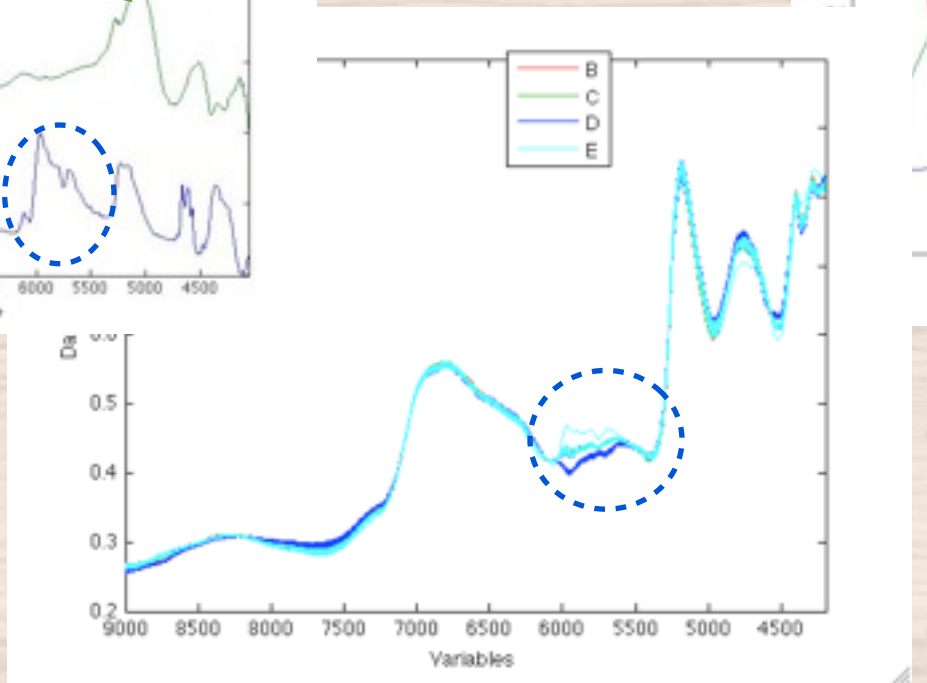


4 way Data array

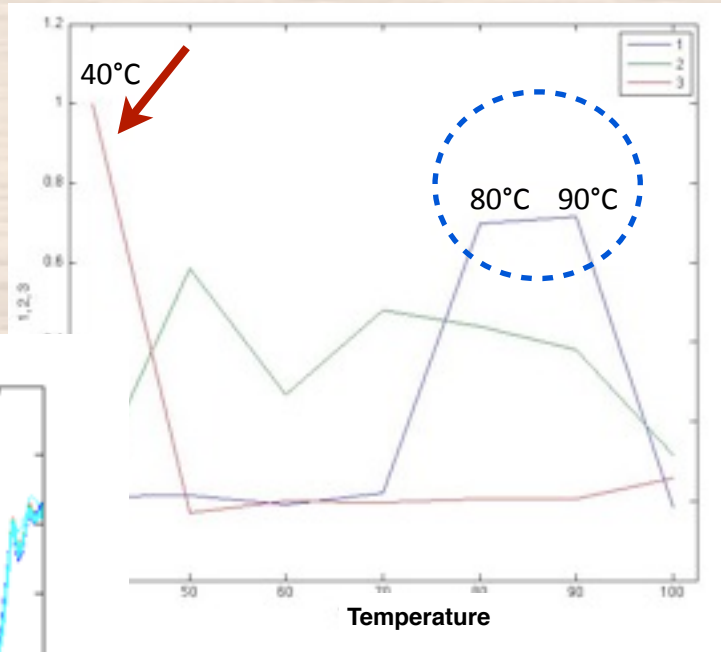
Mode 1: Replicate; storage time (B,C,D,E)
Mode 2: Spectra (EMSC pret.)
Mode 3: Temperature
Mode 4: treatment time

MODE 2





MODE 3



EXCLUDE:
all D, 1_E_4, 1_E_3

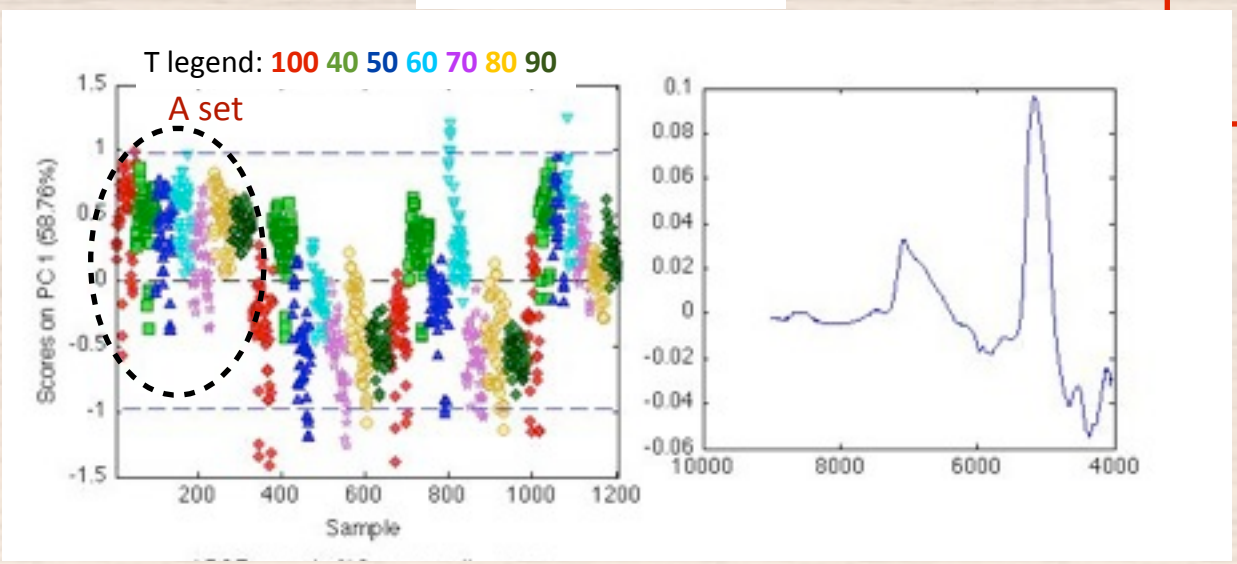
marina.cocchi@unimore.it

PCA : choosing spectra pretreatment

- Minimize variability due to NIR acquisition in different time period
  random selection of T level to measure : **OK;**
  for treatment & storage time: NOT possible
- Minimize variability due to wood block:
  take advantage from replicates and spectra acquired before treatment (A set)

- Standard Normal Variate (SNV)
- MSC, EMSC: average set A as reference spectrum
- MSC, EMSC: the corresponding A (in pair) as reference spectrum
- Subtract the corresponding A spectrum
- + overall normalization

MSC (ref: average A) + SNV



PCA to check the effect

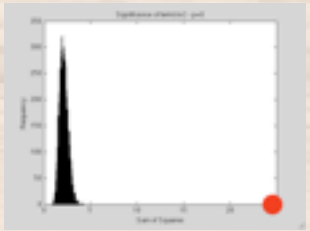
- T variation as main source
- “chemical” loadings
- no trend in A set
- if equivalent the simpler !

marina.cocchi@unimore.it

Results - ASCA

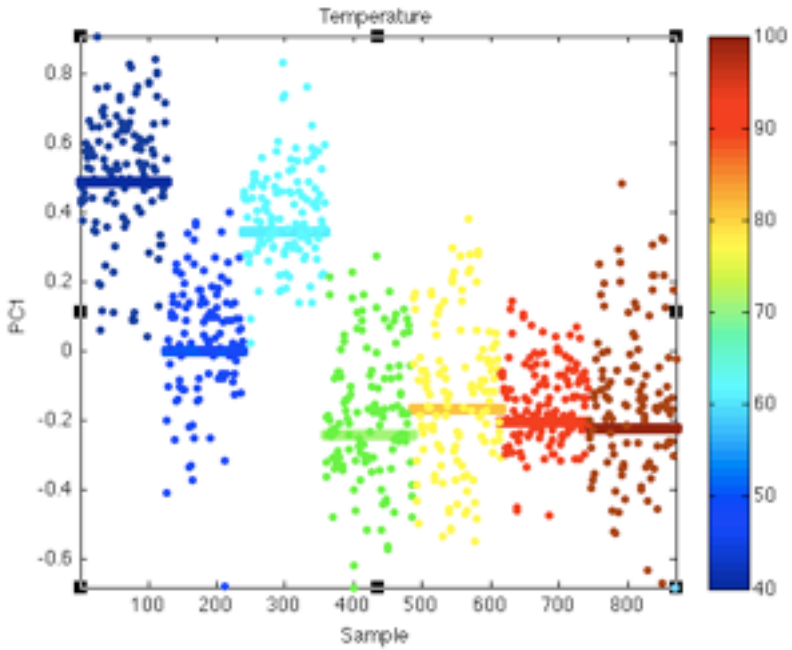
MODEL: $X_c + X_T + X_t + X_{ts} + X_{T*t} + X_{T*ts} + X_{t*ts} + X_{T*t*ts} + X_{res}$

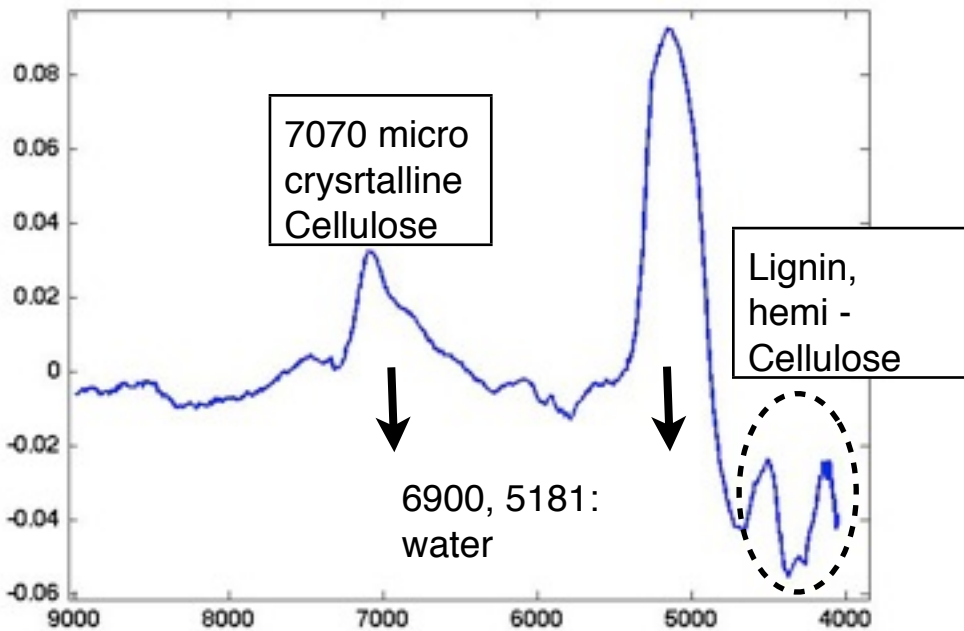
All effects are significants



Effect	Effect Expl Var
T	23.28
storage time (ts)	15.71
T*t	7.07
T*ts	7.41
time treatment (t)	3.66
T*t*ts	3.21

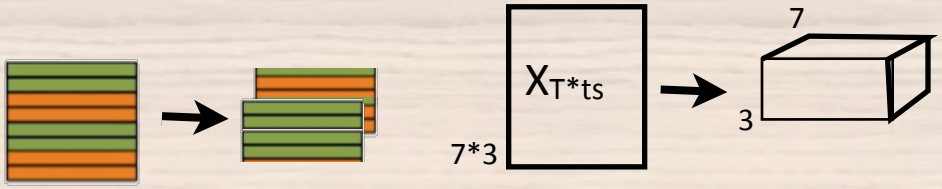
ASCA: T main effect with residuals projected

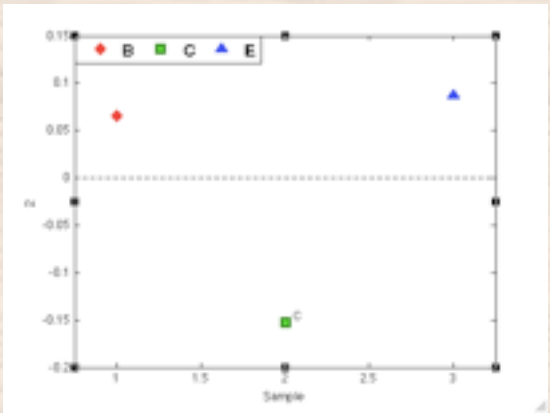


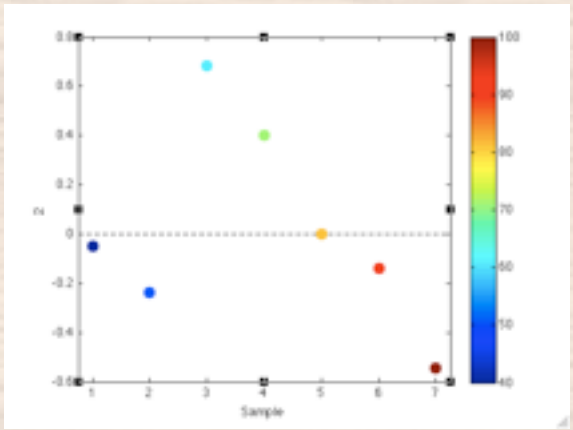


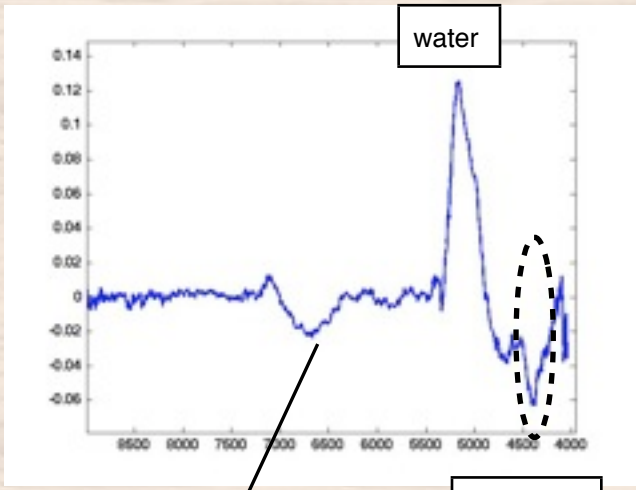
Results - PARAFASCA

3. PARAFASCA refold interaction matrix effect in 3-way and model with PARAFAC.









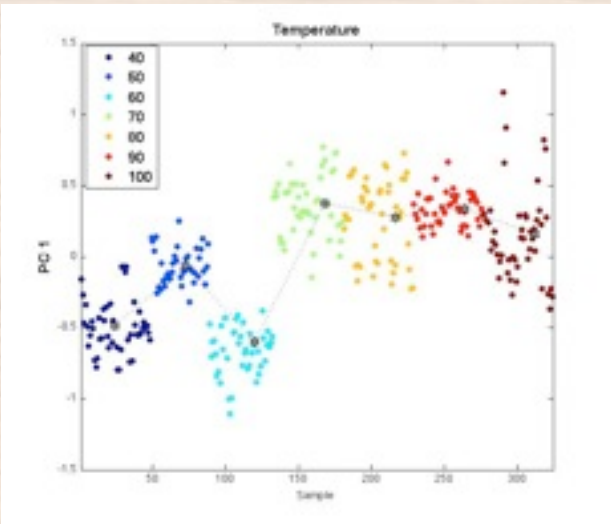
Factor 2 show that at storage time C a trend in T is visible, which is different for 40°C to 50°C (or of little entity) with respect to 60°C to 100°C

Looking at loadings the main contribution is water (decreasing with T) then a other contributions are centered at about 6700 and 4300 and linked to wood structure modifications

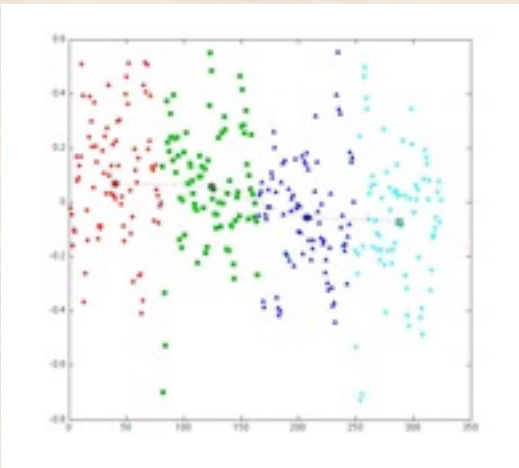
J. J. Jansen et al., J. Chemometrics 2008, 22 114–121

On the basis of previous results only set C analyzed (2 factor)

SCA: T main effect with residuals projected

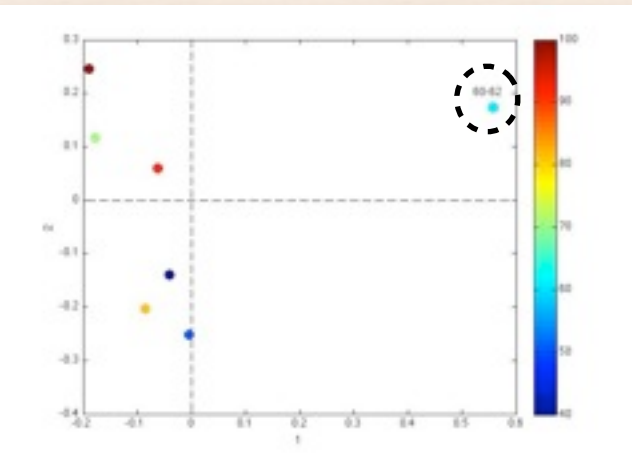


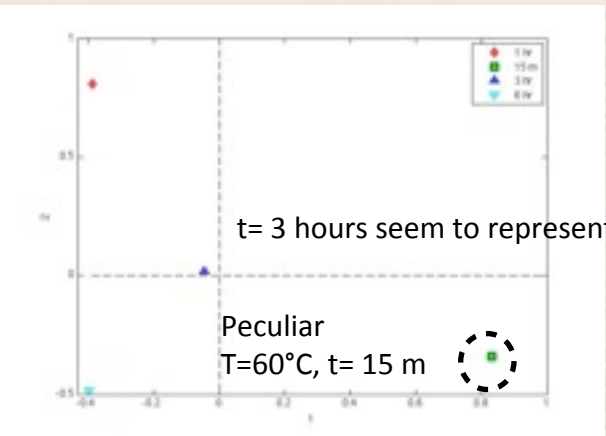
SCA: t main effect with residuals projected



Effect	Expl Var
T	44.13
T*t	10.67
t	2

PARAFASCA: T*t interaction





t= 3 hours seem to represent an average effect

Peculiar
T=60°C, t= 15 m

Remarks

- NIR responds to thermal treatment at low T ("chemical" loadings)
- T induced variation is the main source of variation
- Measure after several months is problematic
- Not a "simple" data set exploratory 2 - 3 ways MVA

and ASCA/PARAFASCA essential

OPTIMAL PROCESS CONDITION (?) .. as starting point time treatment: 1 - 3 hours , from T = 70° a plateau

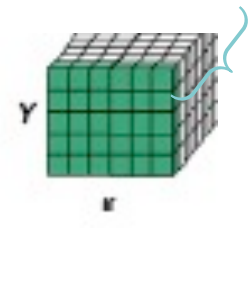
- NIR Peculiarities/Issues
- Chemometrics Tools
- Integration with NIR

1. Handle variability (preprocessing)
2. Monitoring raw materials
3. Process evolving with time
4. Treatments Effect
5. Multispectral/Hyperspectral imaging

Multi - Hyperspectral Images

Multi spectra
tenths of channels

Hyper spectral
whole NIR range



#01 Ultra Blue 430nm
#02 Blue 450nm
#03 Blue 470nm *RGB Blue*
#04 Blue Green 505nm *RGB green*
#05 Green 565nm
#06 Yellow 590nm
#07 Red 630nm *RGB red,*
#08 Red 645nm
#09 Red 660nm
#10 Red 700nm
#11 NIR 850nm *Baseline*
#12 NIR 870nm *Baseline*
#13 NIR 890nm *Fat shoulder (un-saturation)*
#14 NIR 910nm *Protein (C-H)*
#15 NIR 920nm *Carbohydrate (C-H)*
#16 NIR 940nm *Fat*
#17 NIR 950nm *Valley (Carbohydrate, protein)*
#18 NIR 970nm *Water*

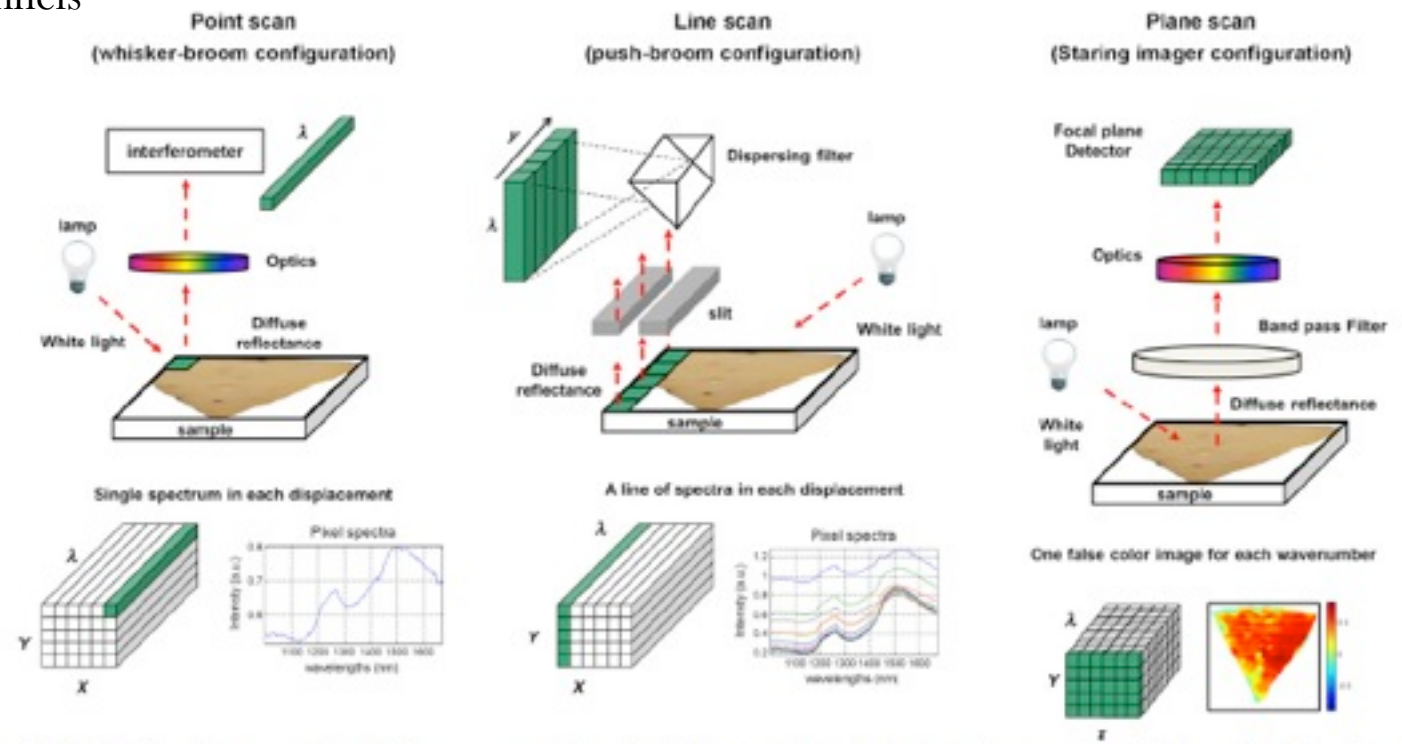
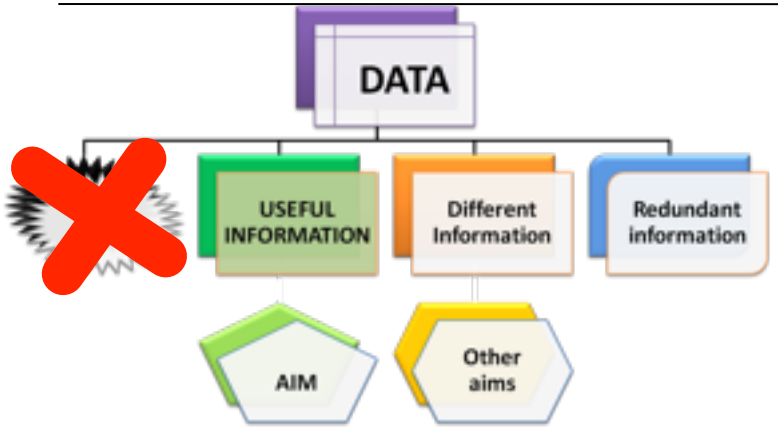


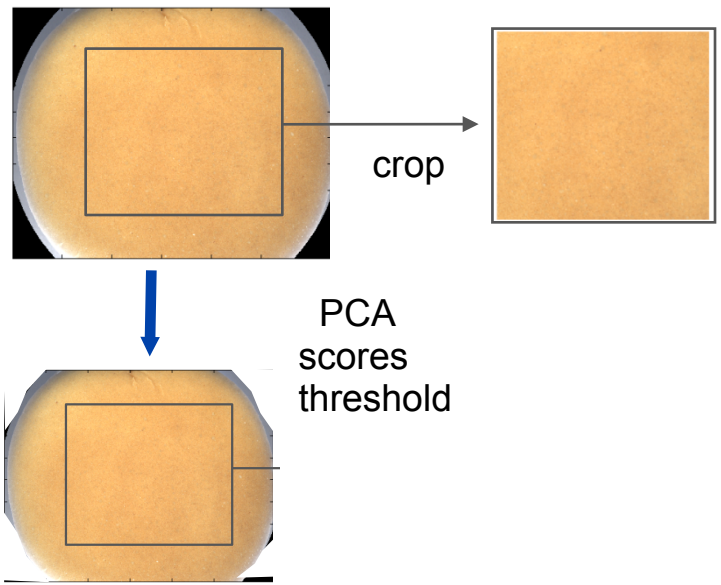
FIGURE 3 Schematic representation of the three common configurations in hyperspectral imaging devices and structure of the final data cube of dimensions ($X \times Y \times \lambda$). *Slightly modified from [14] with permission of Springer.*

from: J. Amigo et al. Food Chemometrics chpt 9.

- quality on line monitoring
- precision agronomy
- .....



- To handle noise
  - Explorative Data Analysis
  - **Preprocessing**

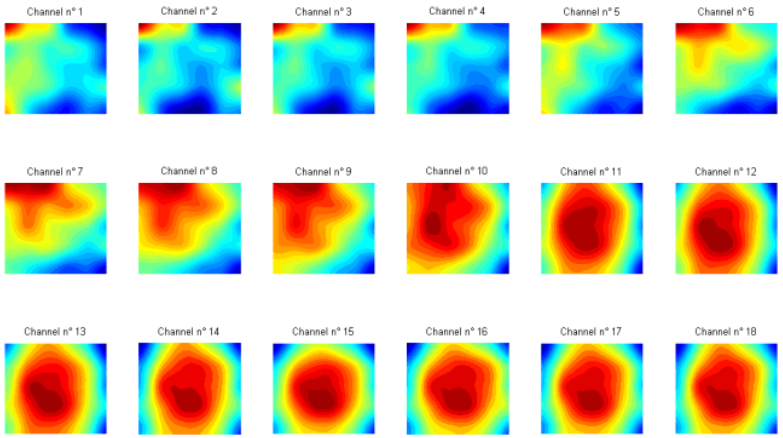


remove “easy” background

Wavelet filter

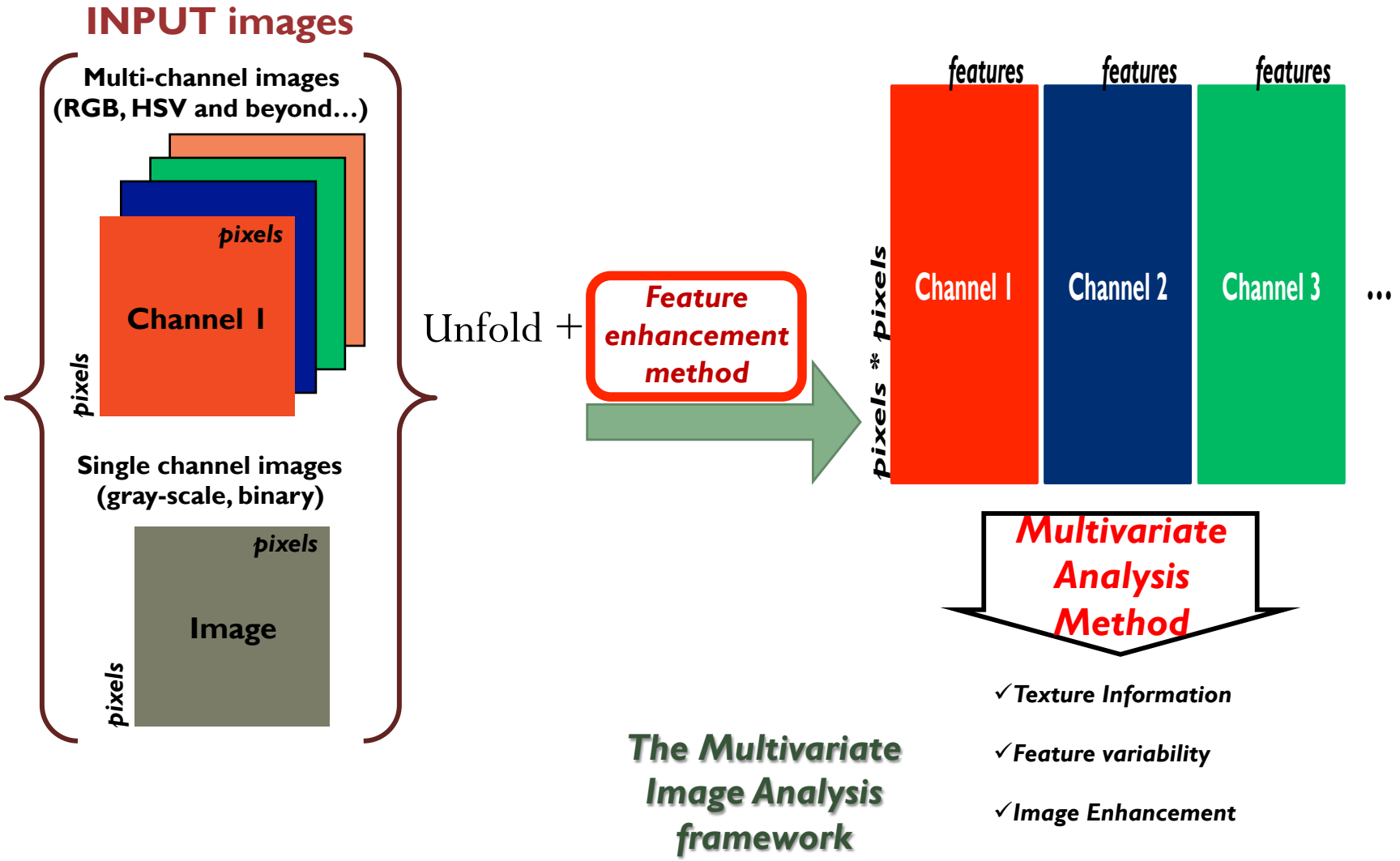
allow to selectively remove

ALLOW to isolate background illumination effects

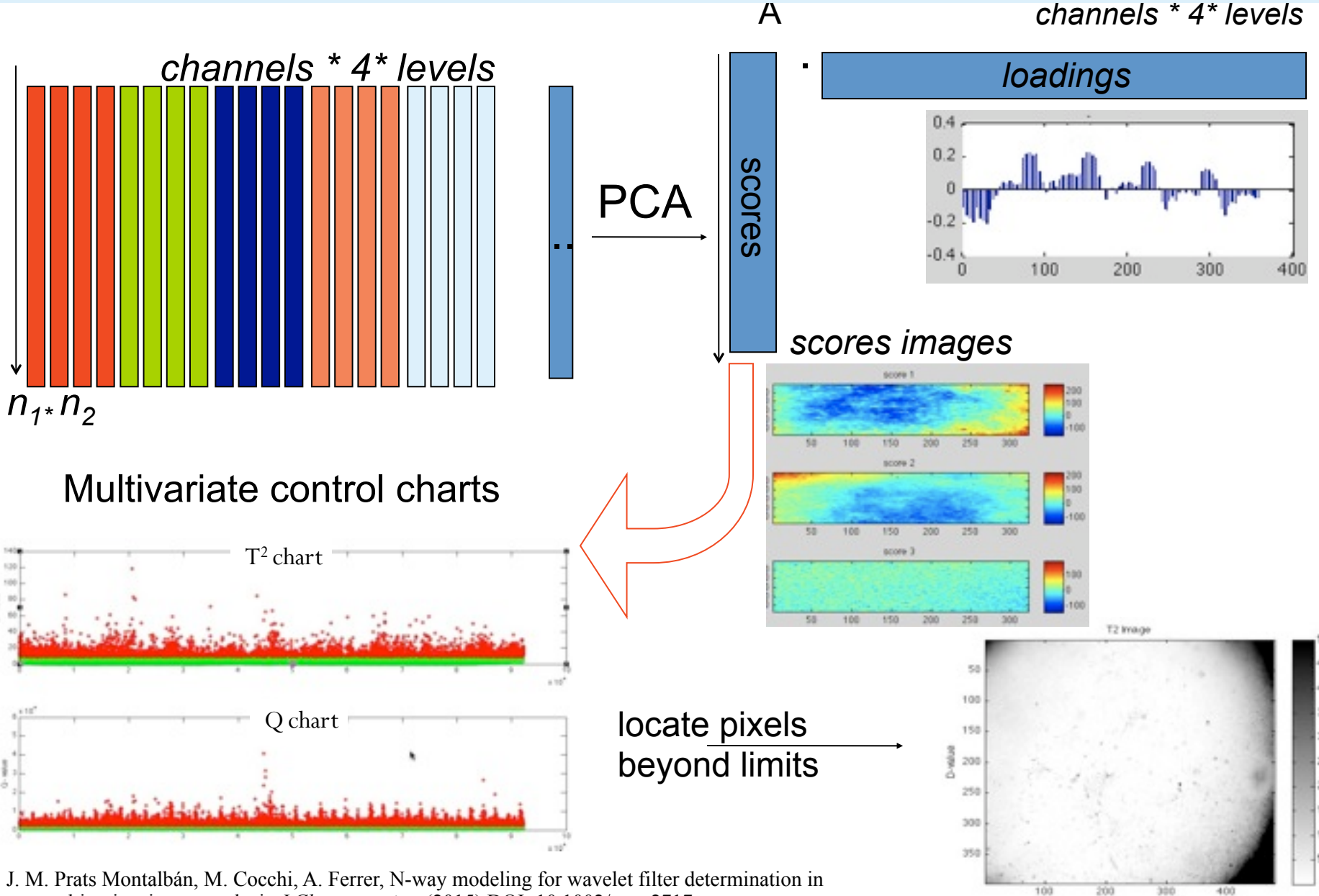


THUS to remove them

- Feature extraction
  - Detect defects / pattern
  - Further Modeling

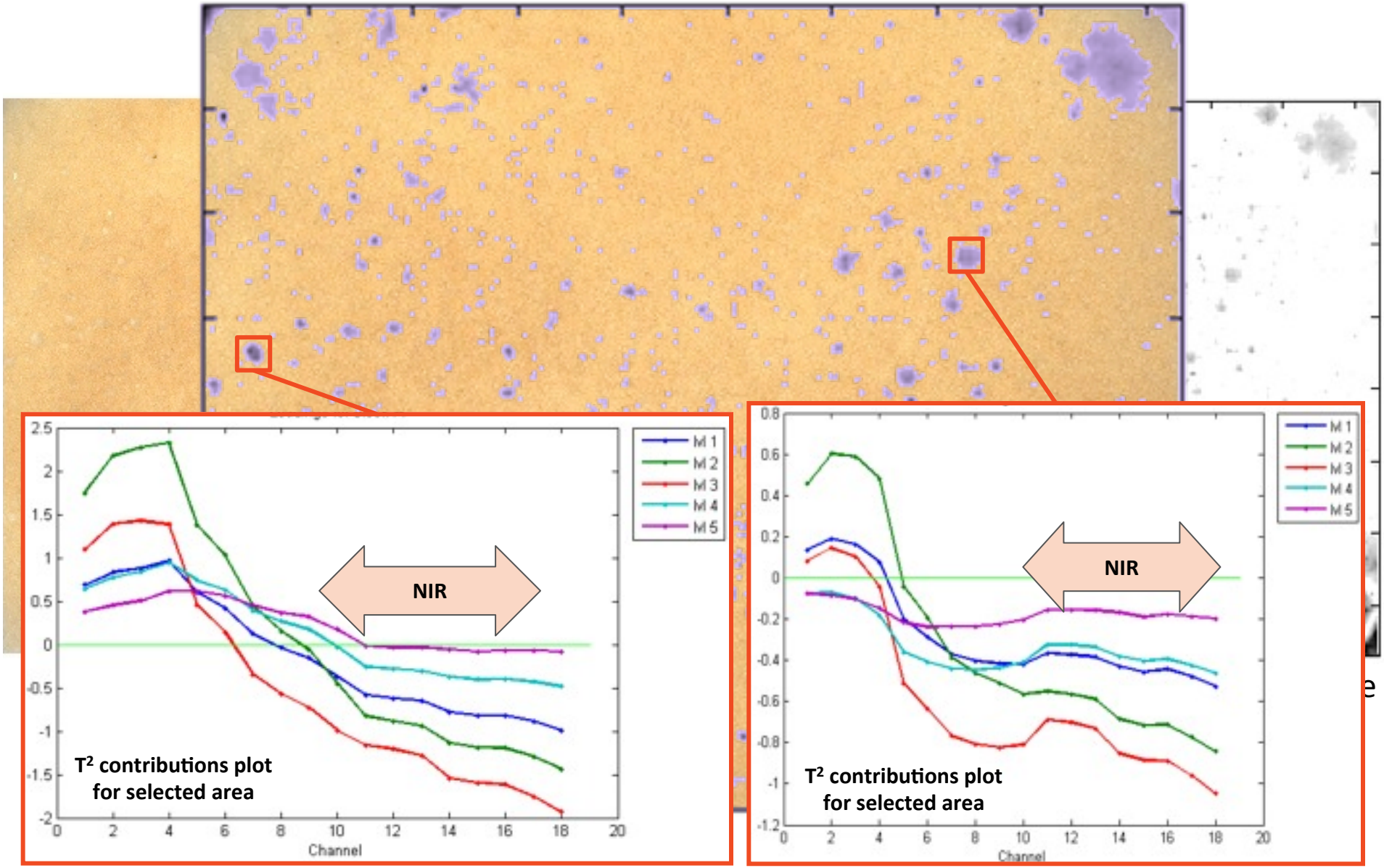


Textural Features extracted by WT-MIA

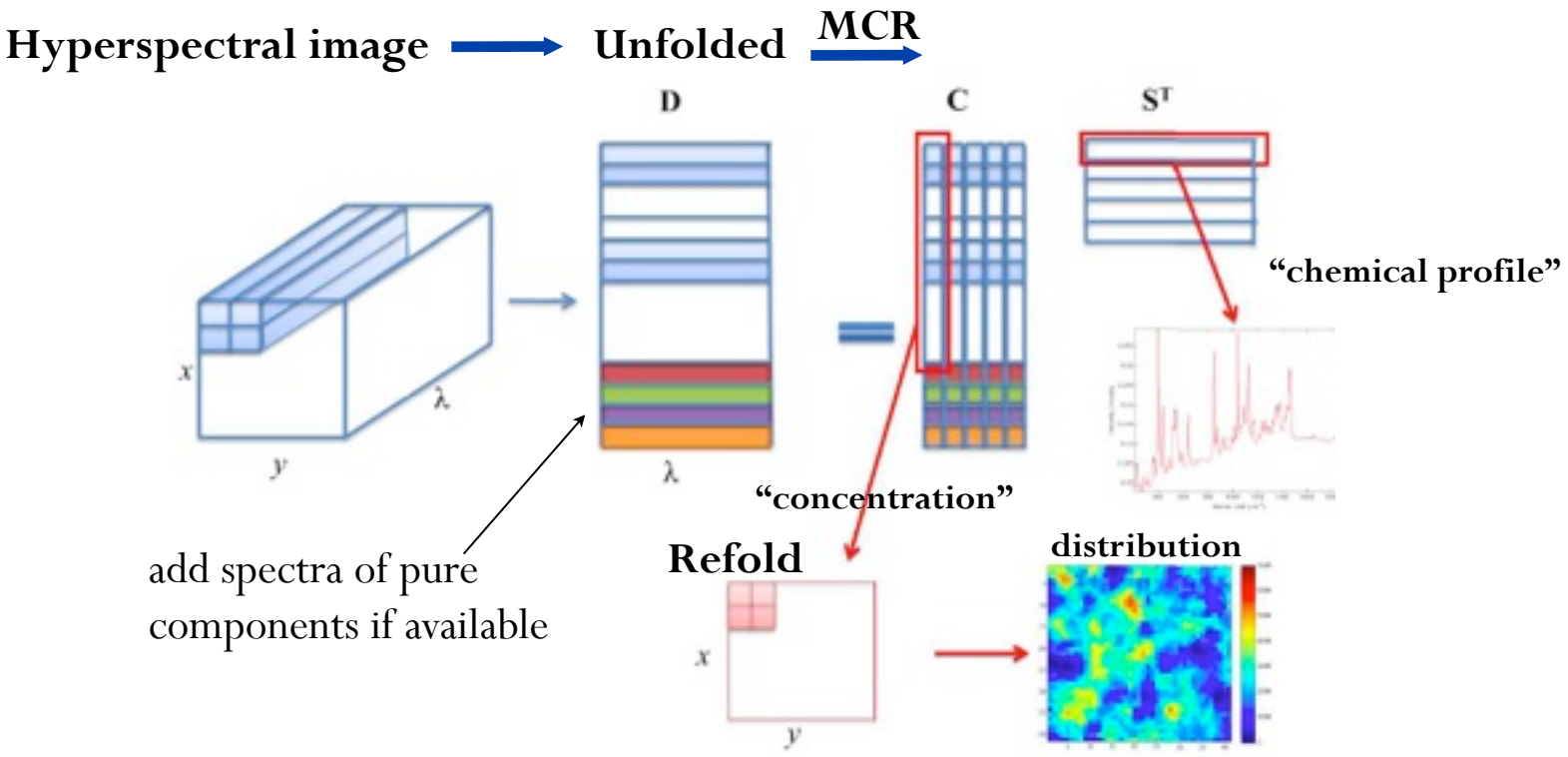


J. M. Prats Montalbán, M. Cocchi, A. Ferrer, N-way modeling for wavelet filter determination in multivariate image analysis *J.Chemometrics* (2015) DOI: 10.1002/cem.2717

M. Cocchi, M. Li Vigni, in C. Ruckebusch Ed. Resolving Spectral Mixtures, Chpt. 13, Data Handling in Science and Technology Vol 30, Elsevier 2016.

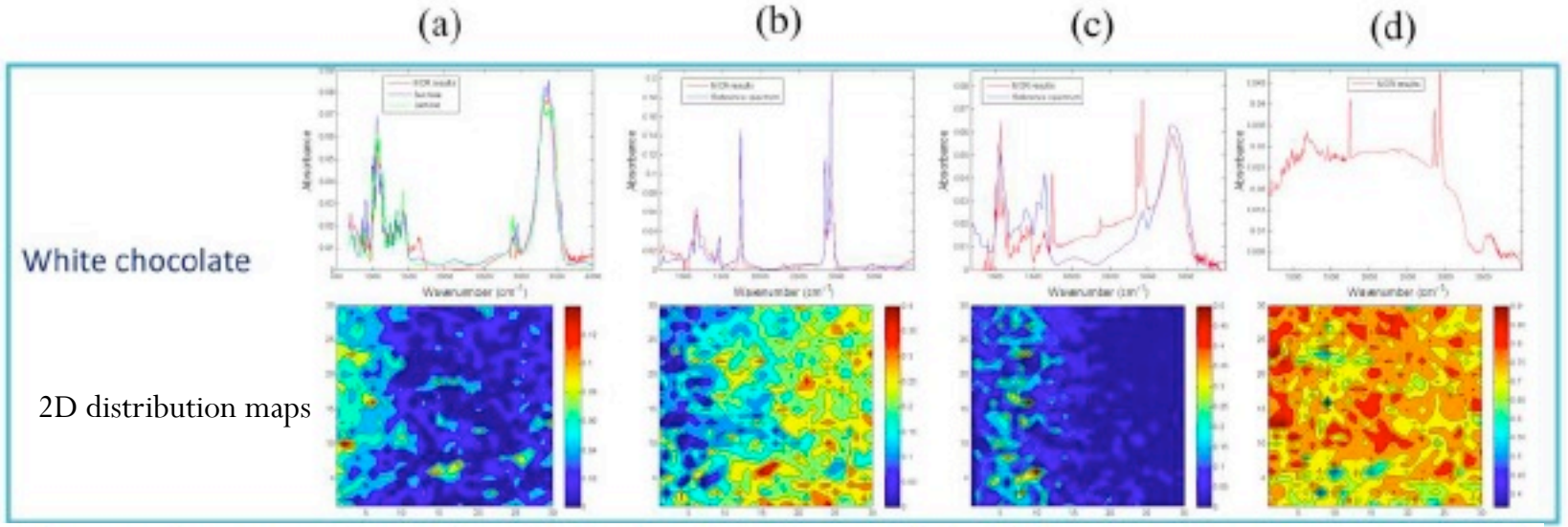

M 1
M 2
M 3
M 4
M 5
NIR
T² contributions plot
for selected area
Channel
M 1
M 2
M 3
M 4
M 5
NIR
T² contributions plot
for selected area
Channel

- Chemical Resolution
- Modeling with MCR
- Extract "chemical profile"

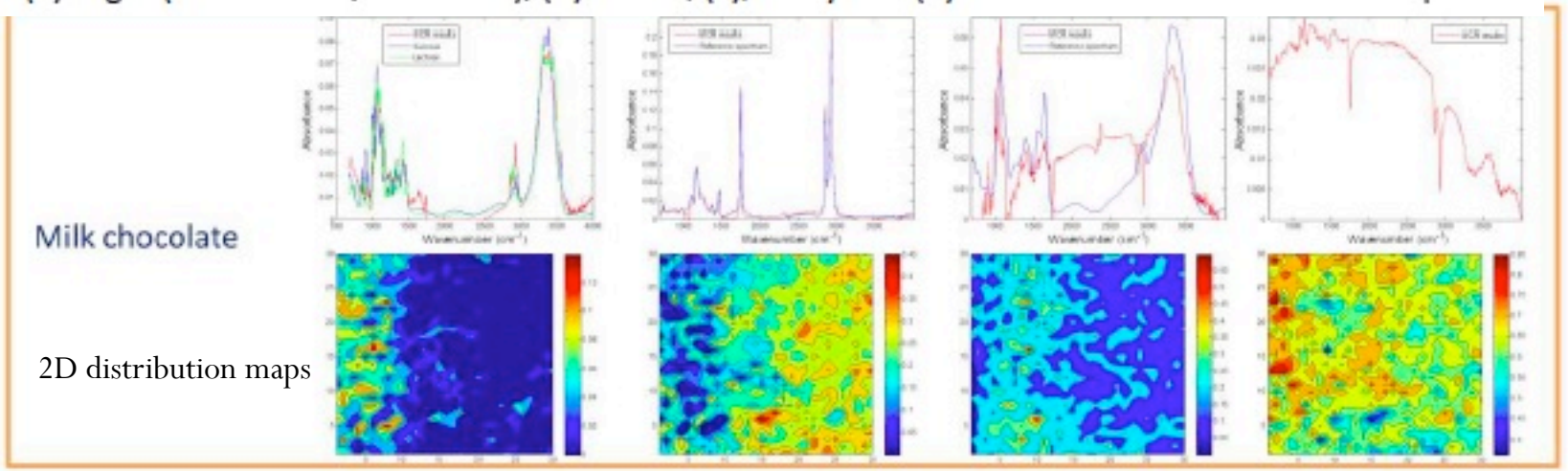


an example with MIR imaging courtesy from A. De Juan

Applied Spectroscopy 69(8) · July 2015



(a) sugar (sucrose and/or lactose), (b) butter, (c), whey and (d) a forth unknown resolved component.



green/blue: reference spectra red: MCR resolved spectra

Take home message

- Data analysis is a step by step strategy
  - when you miss steps you fall (fail)
- Exploratory Data analysis is fundamental
  - do not build your model on “sand”
- Graphical representation essential
  - we believe what we see
- The simpler the better
  - you seldom need fancy stuff.
  - PCA the most used/useful, the less prone to misuse
- We learn by errors
  - experimental design allows you to do errors at the minimum cost

Hints for Discussion

- NIR & Chemometrics:
  - interesting but too difficult ?
  - how much a practitioner need ?
  - where, what....
- Preprocessing
  - really so critical ?
  - have I really to try several ?
- Tell your feelings about

Thanks & questions VERY welcome