INDEPENDENT COMPONENTS ANALYSIS : THEORY, APPLICATIONS AND DIFFICULTIES

Douglas N. Rutledge, Delphine Jouan-Rimbaud Bouveresse

douglas.rutledge@agroparistech.fr delphine.bouveresse@agroparistech.fr

Multivariate Data Analysis

Variables

Samples

(objects)
(individuals)

$$\mathbf{X} = \begin{array}{|c:c:c:c|} x_{11} & x_{1j} & \cdots & x_{1m} \\ \hdashline \cdots & \cdots & \cdots & x_{im} \\ \cdots & \cdots & x_{ij} & \cdots \\ \hdashline x_{n1} & x_{nj} & \cdots & x_{nm} \end{array}$$

Column Vectors

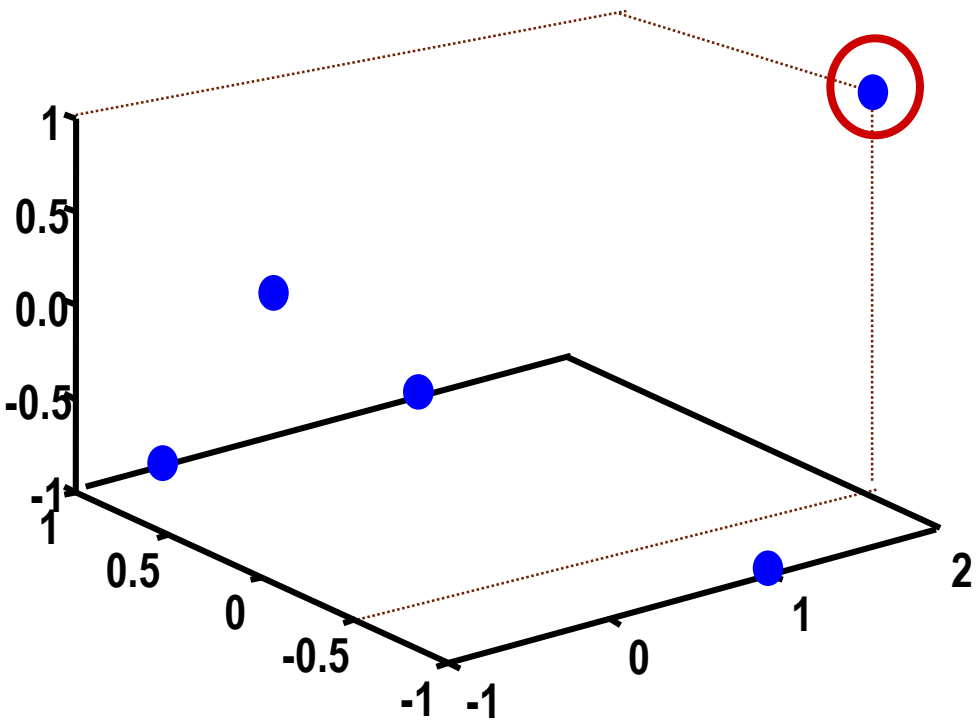
$$\mathbf{x}_1^{\mathrm{T}} = \begin{array}{|c|} x_{11} \\ \cdots \\ x_{i1} \\ x_{n1} \end{array}$$

Line Vectors

$$\mathbf{x}_1 = \begin{array}{|cccc|} x_{11} & \cdots & x_{1j} & x_{1m} \end{array}$$

A matrix as points in space

$$\mathbf{X} = \begin{matrix} -1.00 & 0.00 & 0.50 \\ 1.00 & -1.00 & -1.00 \\ -0.50 & 1.00 & -1.00 \\ 1.00 & 1.00 & -1.00 \\ 2.00 & -0.50 & 1.00 \end{matrix}$$



Principal Components Analysis

Pearson, 1901

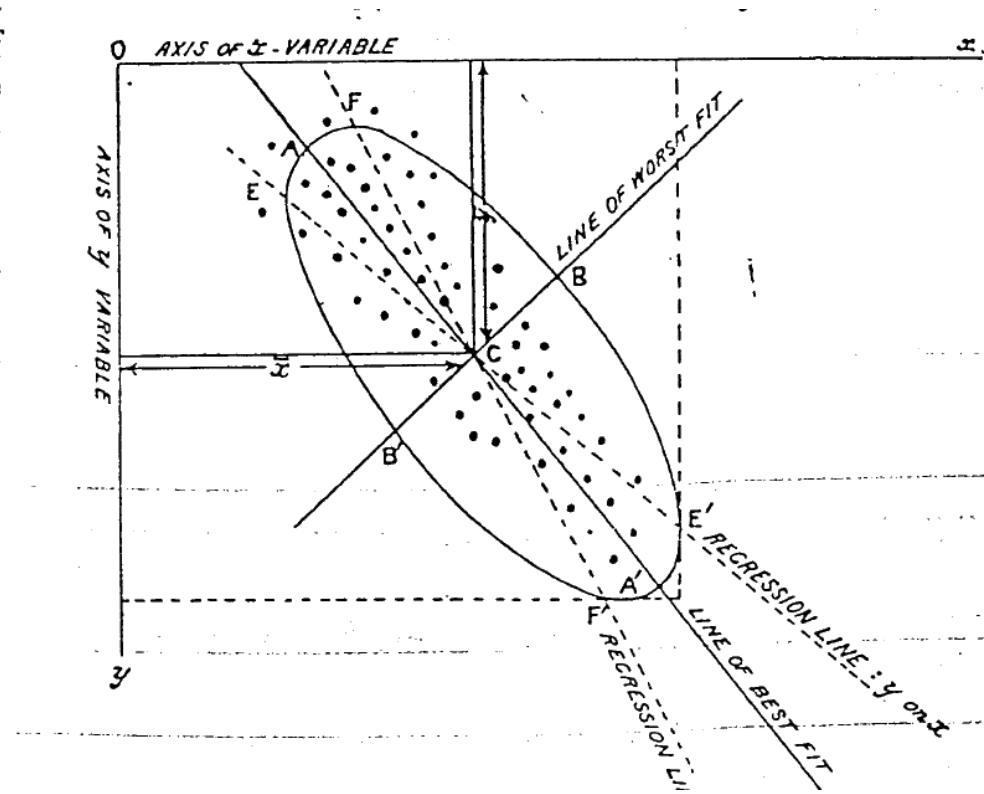
LIII. *On Lines and Planes of Closest Fit to Systems of Points in Space.* By KARL PEARSON, *F.R.S., University College, London* *.

(1) IN many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane. Analytically this consists in taking

$$y = a_0 + a_1 x, \quad \text{or} \quad z = a_0 + a_1 x + b_1 y,$$

$$\text{or} \quad z = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n,$$

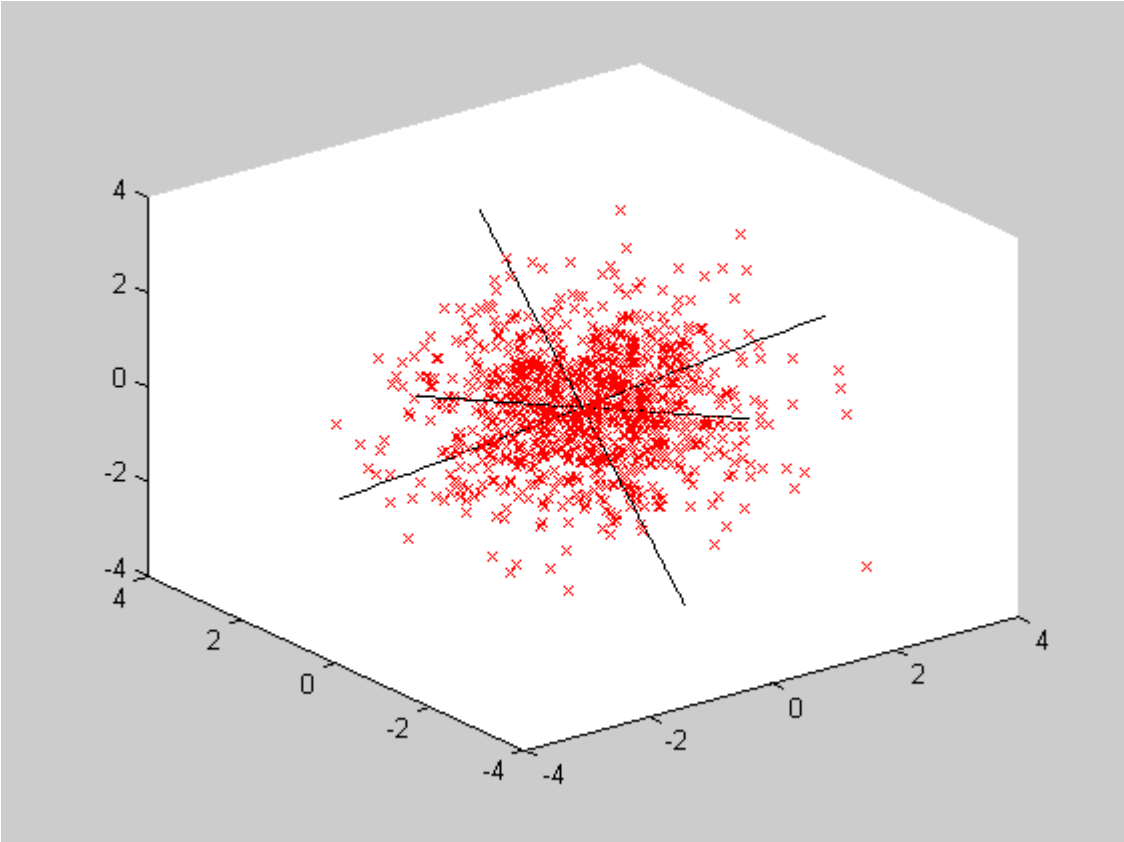
where $y, x, z, x_1, x_2, \ldots x_n$ are variables, and determining the "best" values for the constants $a_0, a_1, b_1, a_0, a_1, a_2, a_3, \ldots a_n$ in relation to the observed corresponding values of the variables. In nearly all the cases dealt with in the text-books of least squares, the variables on the right of our equations are treated as the independent, those on the left as the de-



Principal Components Analysis

Samples projected onto space defined by variables

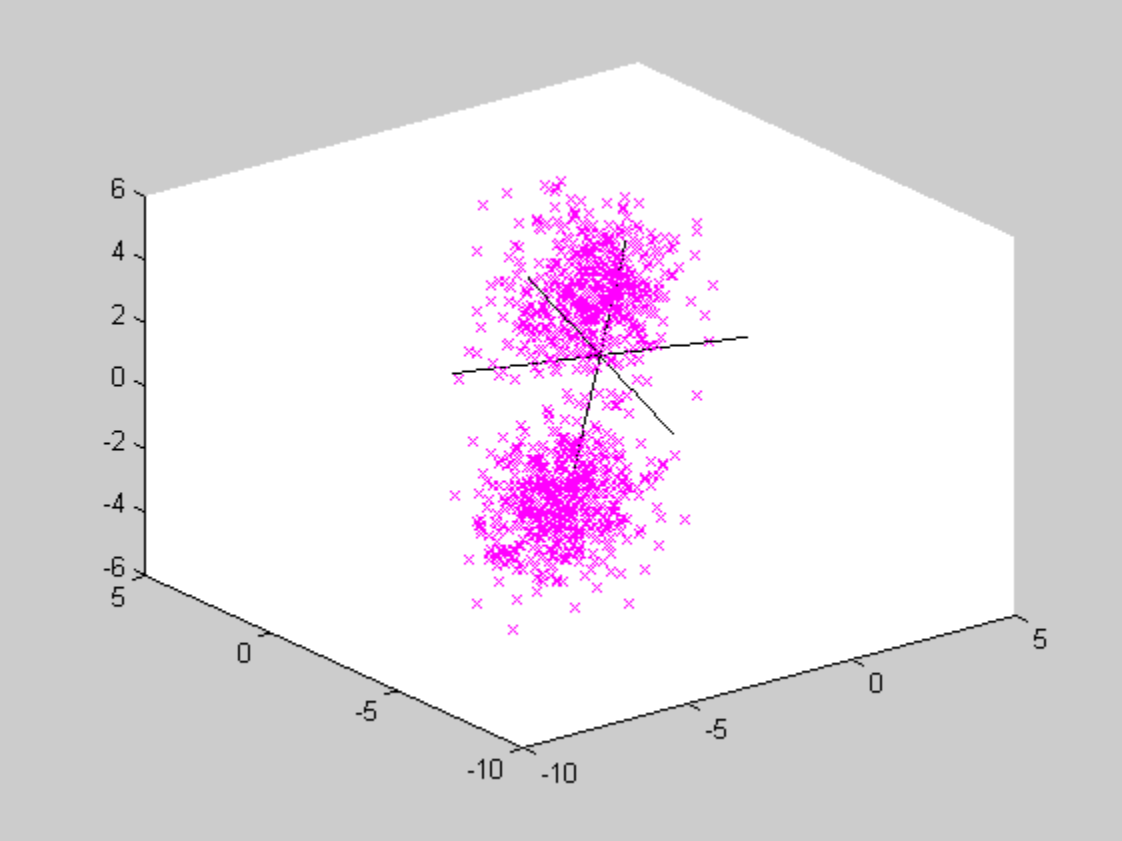
- if ***NO information*** → spherical distribution
  → ***NO preferred*** axes



Principal Components Analysis

Samples projected onto space defined by variables

- if ***information*** → non-spherical distribution
  → ***preferred*** axes



“Blind Source Separation”

Observed
Sensor Signals

Variables

$$\mathbf{X} = \begin{array}{c|c|c|c} x_{11} & x_{1j} & \cdots & x_{1m} \\ \hline \cdots & \cdots & \cdots & x_{im} \\ \cdots & \cdots & x_{ij} & \cdots \\ \hline x_{n1} & x_{nj} & \cdots & x_{nm} \end{array}$$

Row Vectors

$$\mathbf{X} = \begin{array}{c} \mathbf{x}_1^{\mathbf{T}} \\ \cdots \\ \mathbf{x}_i^{\mathbf{T}} \\ \mathbf{x}_n^{\mathbf{T}} \end{array}$$

Independent Components Analysis

Principal Components Analysis (PCA)

data matrix $\longrightarrow$ **points in the multivariate space**

Independent Components Analysis (ICA)

data matrix $\longrightarrow$ **a set of observed signals**, where

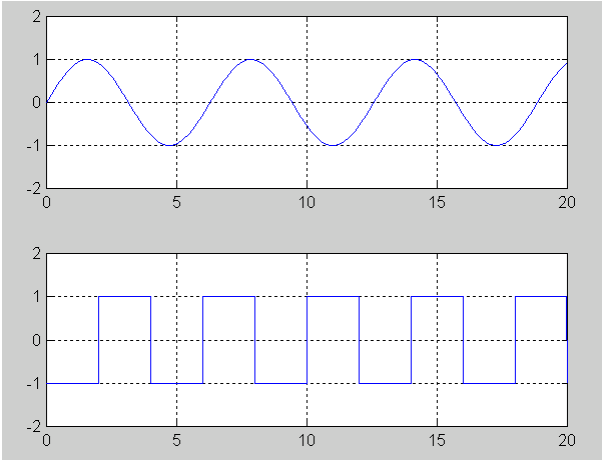
- each observed sensor signal is the weighted sum of pure source signals
- the weighting coefficients are proportions of the source signals

$$x_1 = a_{11} * s_1 + a_{12} * s_2$$
$$x_2 = a_{21} * s_1 + a_{22} * s_2$$
$$\ldots$$

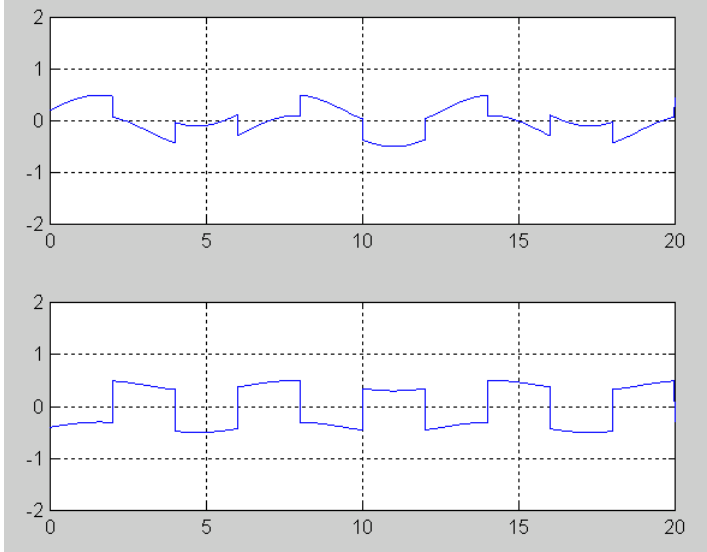
$\Longrightarrow$ **In matrix notation :**

$$X = A * S$$

Example



Two Independent Sources



Two mixtures

$$x_1 = a_{11}s_1 + a_{12}s_2$$

$$x_2 = a_{21}s_1 + a_{22}s_2$$

a_{ij} ... Proportion of source signal s_j in observed signal x_i

The objective of ICA is to find "*physically significant*" vectors.

Hypotheses :

1) No reason for the variations in one source signal to depend *in any way* on the variations in another source signal

   Source signals should therefore be « independent »

2) The measured signals being combinations of several independants sources, they should be more gaussian than the sources
   (Central Limit Theorem)

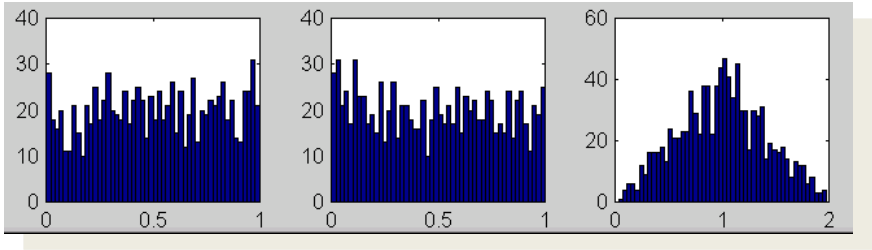
“Nongaussian is independent”: Central Limit Theorem

The measured signals being combinations of several independants sources, they should be more gaussian than the sources

Idea of ICA is to search for sources *the least gaussian possible*

ICA Principal (Non-Gaussian is Independent)

- Key to finding **S** is *non-Gaussianity*
- The distribution of a sum of independent random variables tends toward a Gaussian distribution (by CLT)



$f(s_1)$ $f(s_2)$ $f(x_1) = f(s_1 + s_2)$

Measures of Non-Gaussianity

Quantitative measure of non-gaussianity for ICA estimation

- Kurtosis : gauss=0 (sensitive to outliers)

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2$$

- Entropy : gauss=largest

$$H(y) = -\int f(y) \log f(y) dy$$

- Neg-entropy : gauss = 0 (difficult to estimate)

$$J(y) = H(y_{gauss}) - H(y)$$

- Approximations

$$J(y) = \tfrac{1}{12} E\{y^2\}^2 + \tfrac{1}{48} kurt(y)^2$$

$$J(y) \approx \left[E\{G(y)\} - E\{G(v)\}\right]^2$$

where v is a standard gaussian random variable and :

$$G(y) = \tfrac{1}{a} \log \cosh(a.y)$$

$$G(y) = -\exp(-a.u^2 / 2)$$

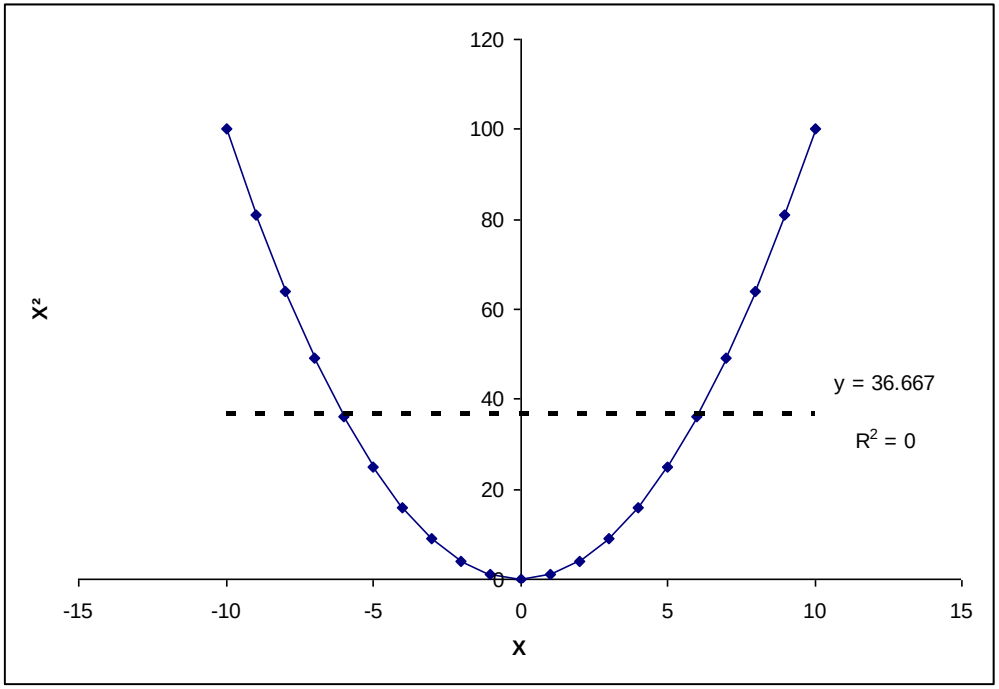
ICA :

- decomposition of a set of vectors into linear components
  that are “as independent as possible”

“Independence” :

- goes beyond (second-order) decorrelation
  involves the non-Gaussianity of the data

“Independent” versus “non-correlated”



Deux variables **x** et **y** sont indépendantes si,
quelles que soient les valeurs non nulles *a* et *b*,

r(**x**^*a*, **y**^*b*) = 0

Dans le cas présenté,

si *a*=2 et *b*=1, r=1

donc **x** et **y** ne sont pas indépendantes.

ICA attempts to recover the original signals by estimating a linear transformation, using a criterion that measures statistical independence among the sources

This may be achieved by the use of higher-order information that can be extracted from the densities of the data

PCA does not really look for (*and usually does not find*) components with physical reality

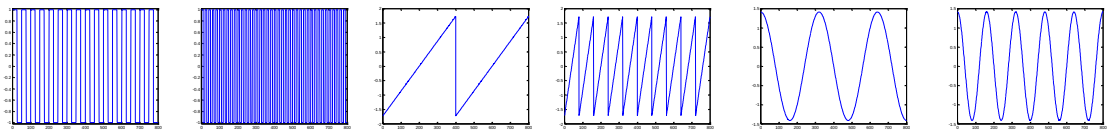
Why not ?

- **PCA finds "direction of greatest dispersion of the samples"**
- **this is the wrong direction for "signal separation" !**

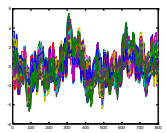
A simulated example

Pure source signals

Observed sensor signals

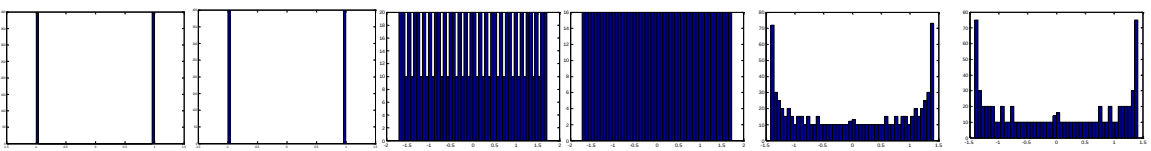


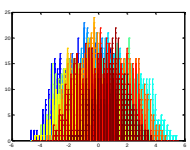
+ Gaussian Noise



Histograms of source signals

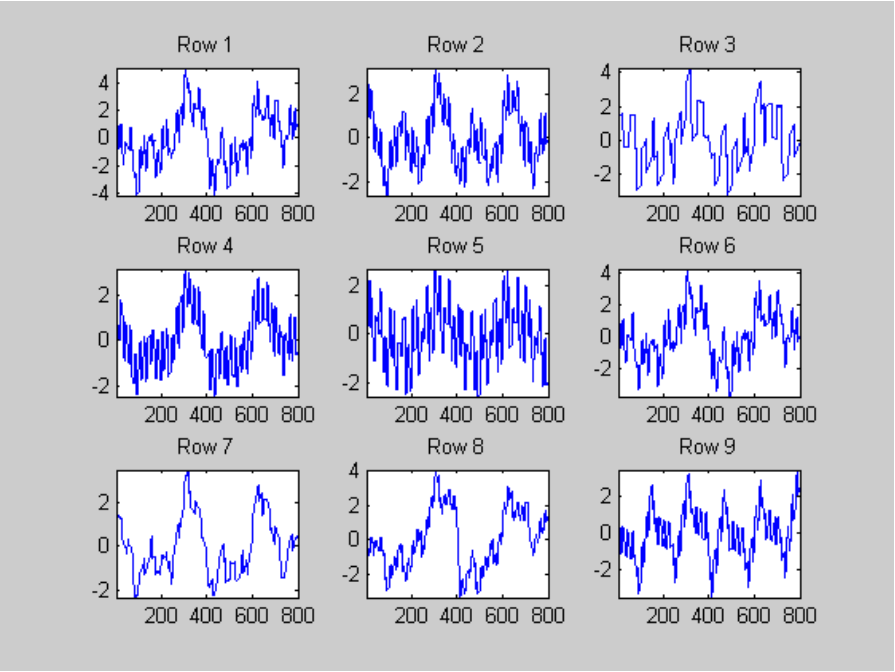
Histograms



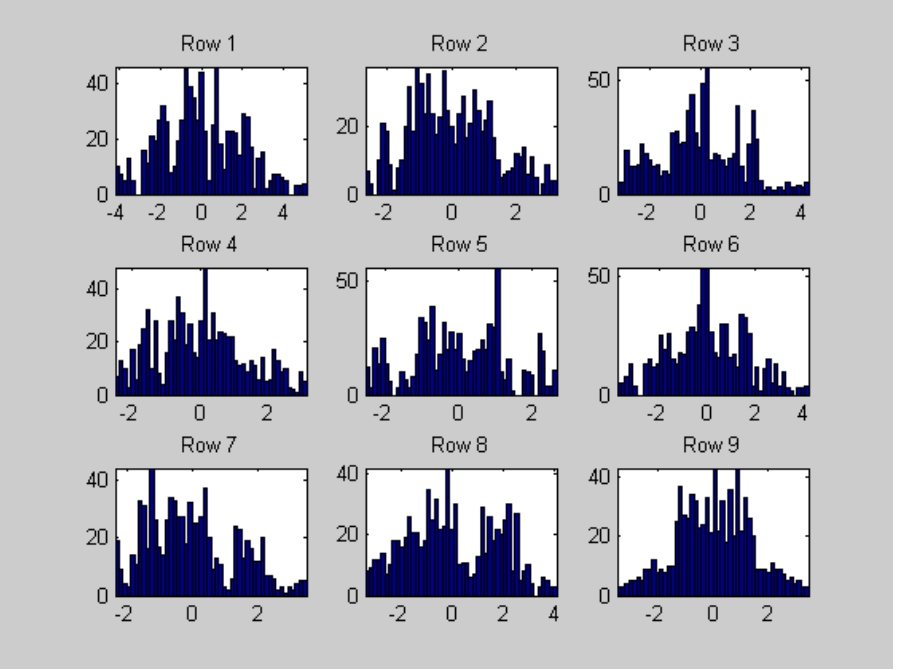


Initial data matrix

Matrix rows

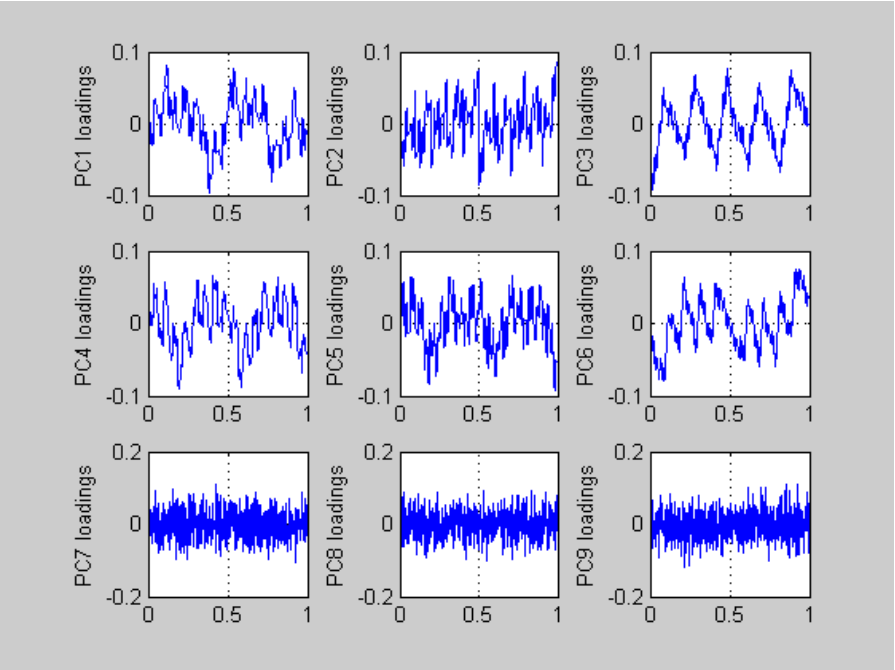


Histograms

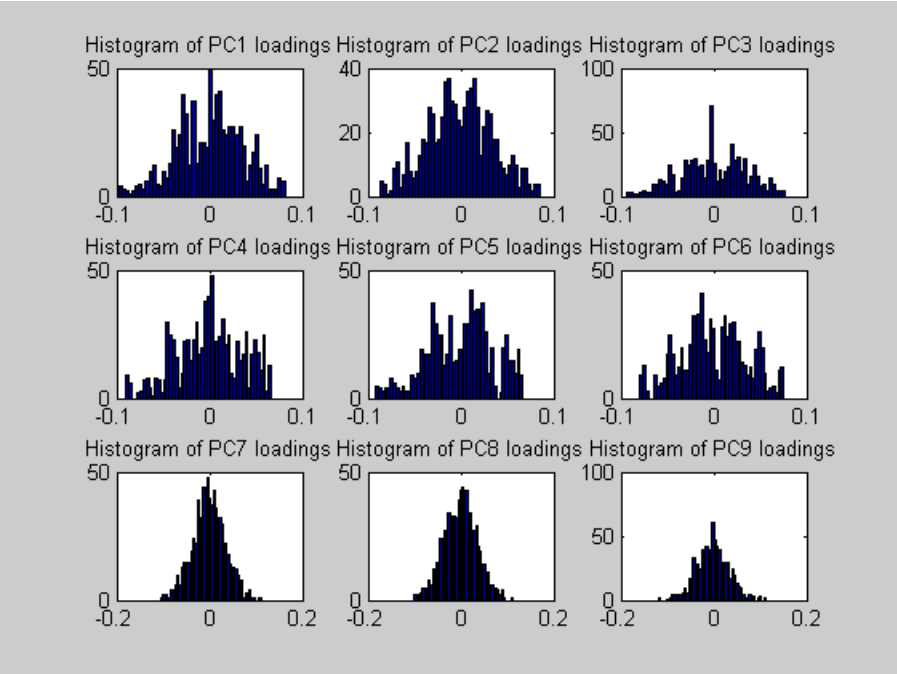


PCA applied to the example

Loadings

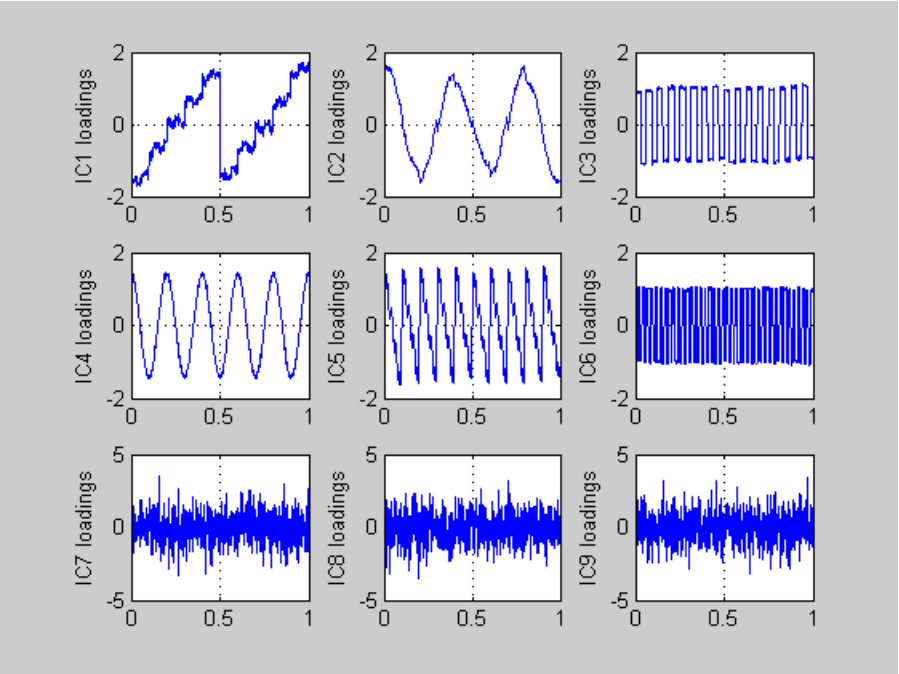


Histograms

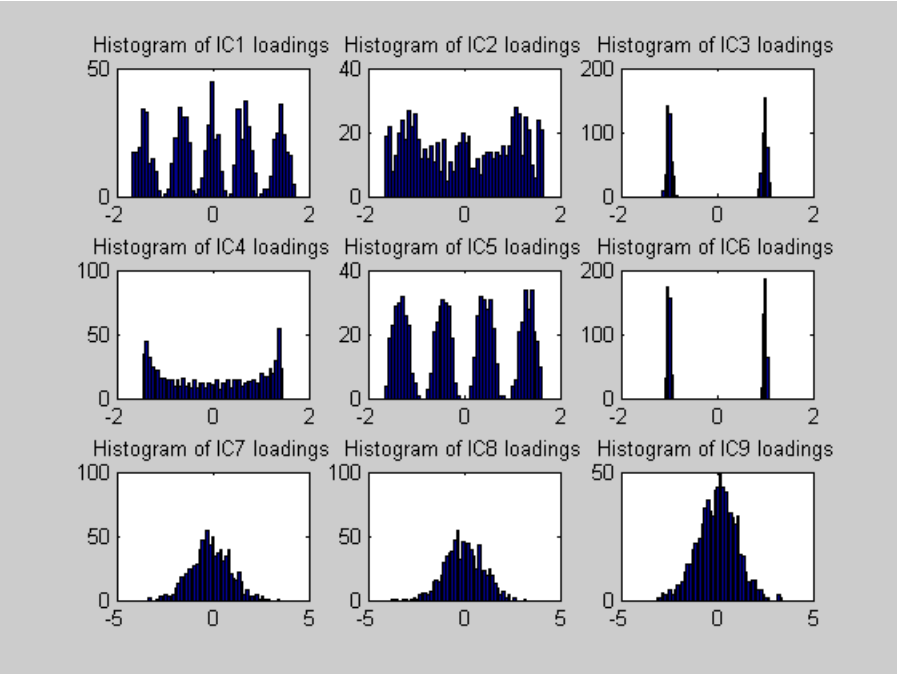


ICA applied to the example

Source Signals



Histograms



Procedure

ICA calculates a *demixing* matrix, W

W approximates A^{-1} , the inverse *mixing* matrix

The pure component signals are recovered from the measured mixed signals by :

$$S = W*X$$

The trick is to calculate W when we only have X !

Calculating the demixing matrix

Many algorithms available :

FastICA :	**numerical gradient search**
Projection Pursuit :	**interestingly structured vectors**
Complexity Pursuit :	**minimise common information**
JADE :	**based on "*cumulants*"**

….

JADE
(Joint Approximate Diagonalization of Eigenmatrices)

- JADE was developed by Cardoso *et al.* in 1993 [1].
- A *blind source separation* method to extract independent non-Gaussian sources from signal mixtures with Gaussian noise
- Based on the construction of a fourth-order *cumulant* array from the data
- It is freely downloadable from

  http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m

Cardoso, J-F. and Souloumiac, A.
Blind beamforming for non-Gaussian signals.
IEE proceedings-F, (1993). 140 (6) 362-370

Cumulants

For a set of values : x

2° order auto-cumulant :

$$\sigma^2(x) = \mathrm{Cum}_2\{x, x\} = E\{x^2\}$$

4° order auto-cumulant

$$\kappa_x(x) = \mathrm{Cum}_4\{x, x, x, x\} = E\{x^4\} - 3.E^2\{x^2\}$$

Cumulants

Cumulants can be expressed in tensor form.
A tensor can be seen as a generalization of the notion of matrices.

Cumulant tensors are a generalization of covariance matrices.

The main diagonal of cumulants tensors are *auto-cumulants*
and non-diagonal elements are *cross-cumulants*.

Statistically mutually independent vectors (components) give :

- *null* cross-cumulants
- *maximum* auto-cumulants

Cumulants

In PCA :

The Eigenvalue decomposition of a covariance matrix results in its diagonalization, i.e. the cancellation of its non-diagonal terms.

In JADE :

A generalization of this methodology was proposed by Cardoso and Souloumiac to diagonalise fourth-order cumulants tensors.

4th-order Cumulants

For the 3 observed signal vectors : $\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$

Create a four-way array with dimensions [3, 3, 3, 3]

Each of the 81 elements is the mean of products of the vectors.

For $v_i, v_j, v_k, v_l = \mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}$

$$\begin{aligned} Cum_4\{v_i, v_j, v_k, v_l\} = E\{v_i v_j v_k v_l\} \\ - E\{v_i v_j\} \,.\, E\{v_k v_l\} \\ - E\{v_i v_k\} \,.\, E\{v_j v_l\} \\ - E\{v_i v_l\} \,.\, E\{v_j v_k\} \end{aligned}$$

3 Auto-Cumulants

3 different values

For v_i, v_j, v_k, v_l = x_p (p=1:3)

$$\mathrm{Cum}_4\{p, p, p, p\} = E\{x_p^4\} - E\{x_p^2\} \,.\, E\{x_p^2\} - E\{x_p^2\} \,.\, E\{x_p^2\} - E\{x_p^2\} \,.\, E\{x_p^2\}$$

$$\mathrm{Cum}_4\{p, p, p, p\} = \mathrm{mean}\,(x_p^4) - \mathrm{mean}\,(x_p^2) \,.\, \mathrm{mean}\,(x_p^2) - \mathrm{mean}\,(x_p^2) \,.\, \mathrm{mean}\,(x_p^2) - \mathrm{mean}\,(x_p^2) \,.\, \mathrm{mean}\,(x_p^2)$$

$$\mathrm{Cum}_4\{p, p, p, p\} = \mathrm{mean}\,(x_p^4) - 3$$

18 Even Cross-Cumulants

3 different values

For $v_i, v_j = x_p$ et $v_k, v_l = x_q$ (p=1:3, q=1:3, p≠q)

$$\mathrm{Cum}_4\{p, p, q, q\} = E\{x_p^2 x_q^2\} - E\{x_p^2\} \,.\, E\{x_q^2\} - E\{x_p x_q\} \,.\, E\{x_p x_q\} - E\{x_p x_q\} \,.\, E\{x_p x_q\}$$

$$\mathrm{Cum}_4\{p, p, q, q\} = \mathrm{mean}\,(x_p^2 x_q^2) - \mathrm{mean}\,(x_p^2) \,.\, \mathrm{mean}\,(x_p^2) - \mathrm{mean}\,(x_p x_q) \,.\, \mathrm{mean}\,(x_p x_q) - \mathrm{mean}\,(x_p x_q) \,.\, \mathrm{mean}\,(x_p x_q)$$

$$\mathrm{Cum}_4\{p, p, q, q\} = \mathrm{mean}\,(x_p^2 x_q^2) - 1$$

60 Odd Cross-Cumulants

9 different values

For $V_i = x_p$ et $V_j, V_k, V_l = x_q$ (p=1:3, q=1:3, p≠q)

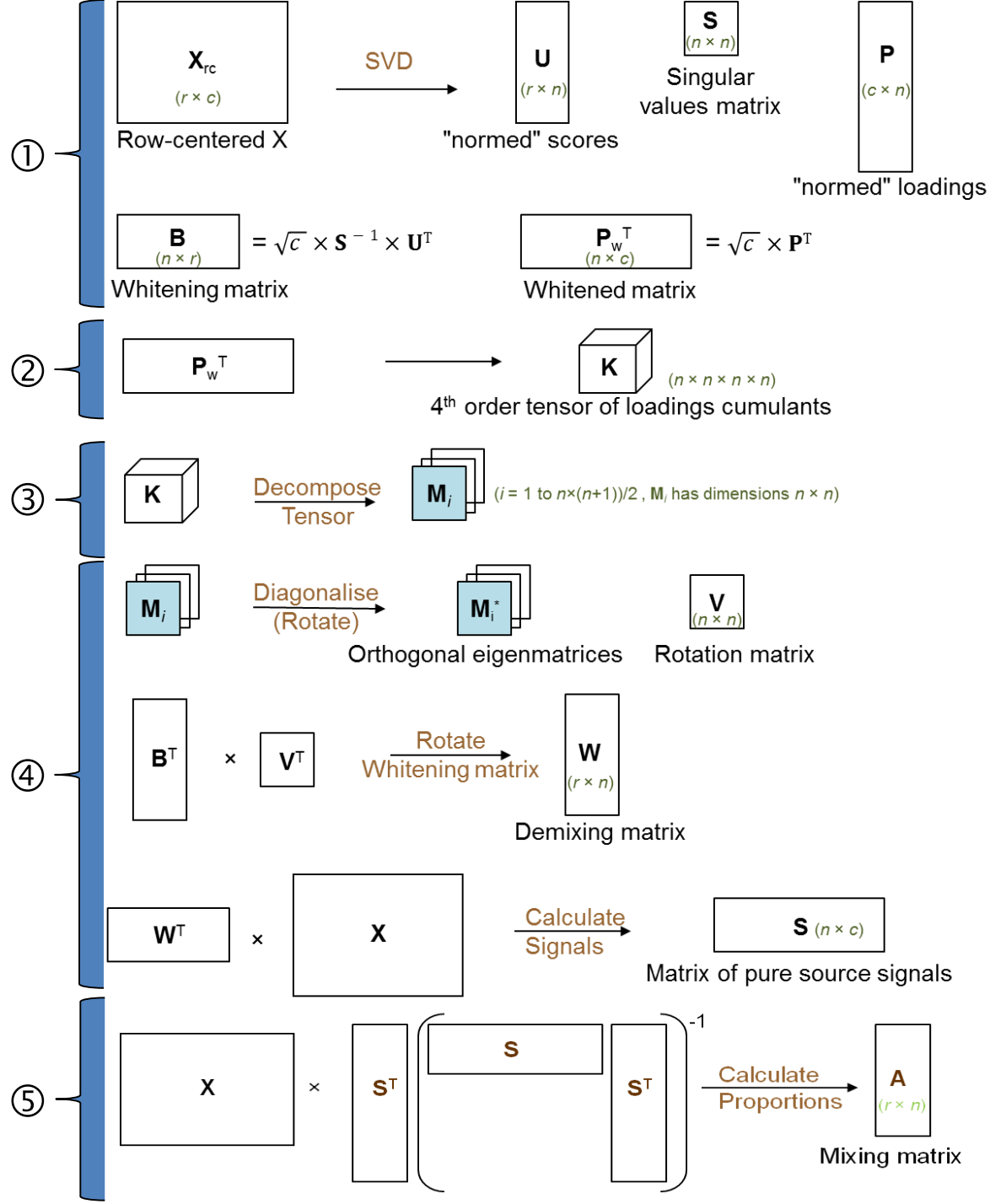
$$\mathrm{Cum}_4\{p, q, q, q\} = E\{x_p x_q^3\} - E\{x_p x_q\} \,.\, E\{x_q^2\} - E\{x_p x_q\} \,.\, E\{x_q^2\} - E\{x_p x_q\} \,.\, E\{x_q^2\}$$

$$\mathrm{Cum}_4\{p, q, q, q\} = \mathrm{mean}\,(x_p x_q^3) - \mathrm{mean}\,(x_p x_q) \,.\, \mathrm{mean}\,(x_p^2) - \mathrm{mean}\,(x_p x_q) \,.\, \mathrm{mean}\,(x_q^2) - \mathrm{mean}\,(x_p x_q) \,.\, \mathrm{mean}\,(x_q^2)$$

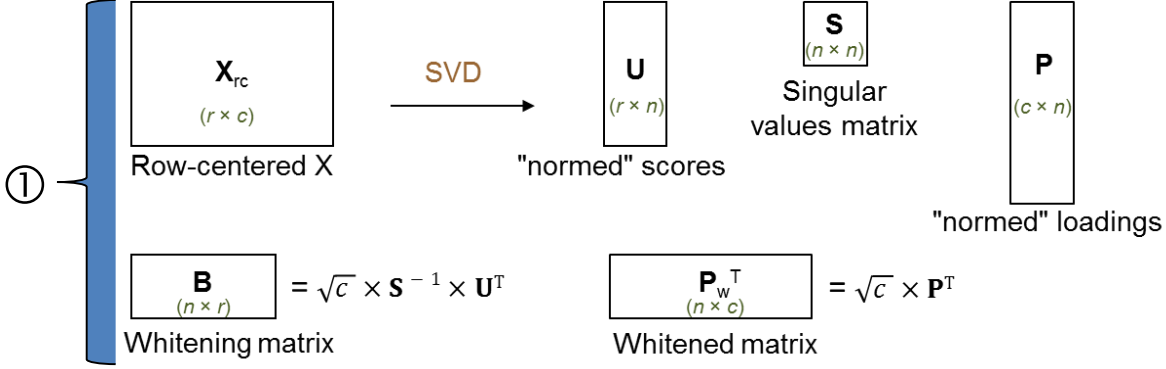
$$\mathrm{Cum}_4\{p, q, q, q\} = \mathrm{mean}\,(x_p x_q^3) - 0$$

The JADE algorithm:

a multi-step procedure

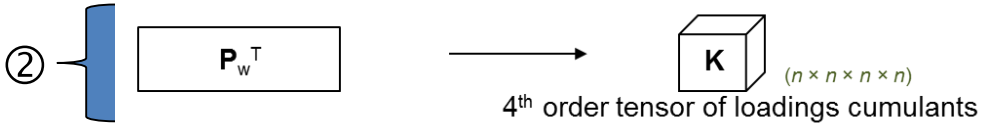


If **X** is very wide or tall, it can be replaced by the loadings obtained from Kernel-PCA, or by using (segmented-) PCT



① Whitening (sphering) of the original data matrix **X**

- Reduce the data dimensionality
- Orthogonalize the data
- Normalise the data on all dimensions



② Cumulants computation

Cumulants :

Generalization of *moments* such as the **mean** (1st-order auto-cumulant) and the **variance** (2nd-order auto-cumulant) to statistics of order > 2:

Independent signal vectors have :

null 4th-order cross-cumulants and **maximal auto-cumulants**

➔Find rotations of $\mathbf{P}_w$ so that loadings vectors become independent

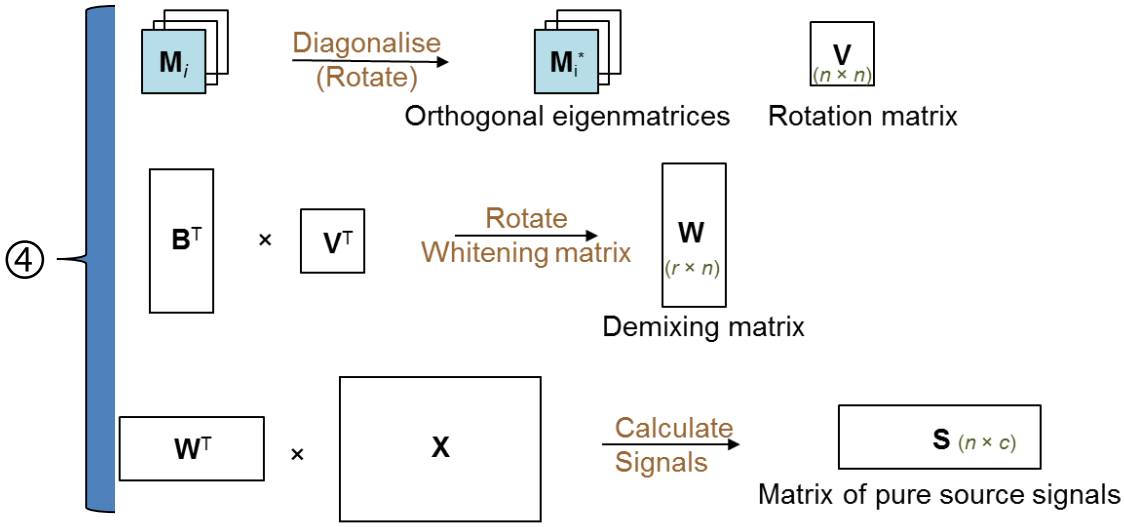
③ Decomposition of the cumulant tensor

Generalization to higher-order statistics

of the eigenvalue decomposition of the variance-covariance matrix:

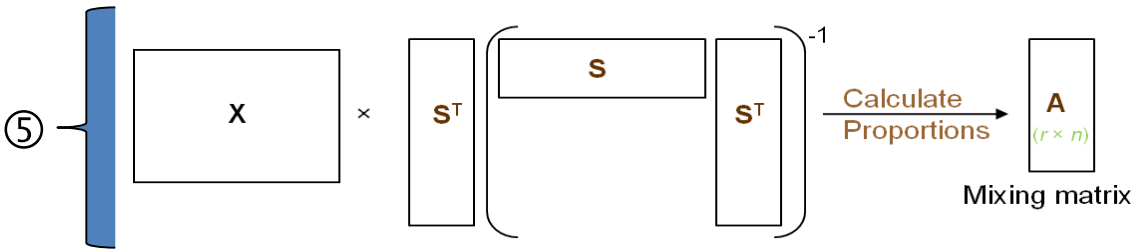
- Decompose the 4th-order tensor into a set of $n(n+1)/2$ orthogonal eigenmatrices $\mathbf{M}_i$
- Project the cumulant tensor onto these planes

④ Joint Diagonalization of eigenmatrices



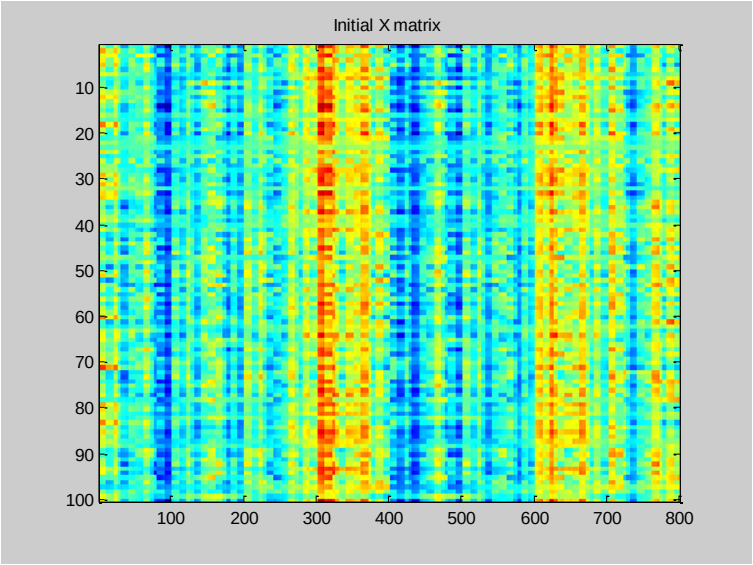
(based on the Jacobi algorithm)

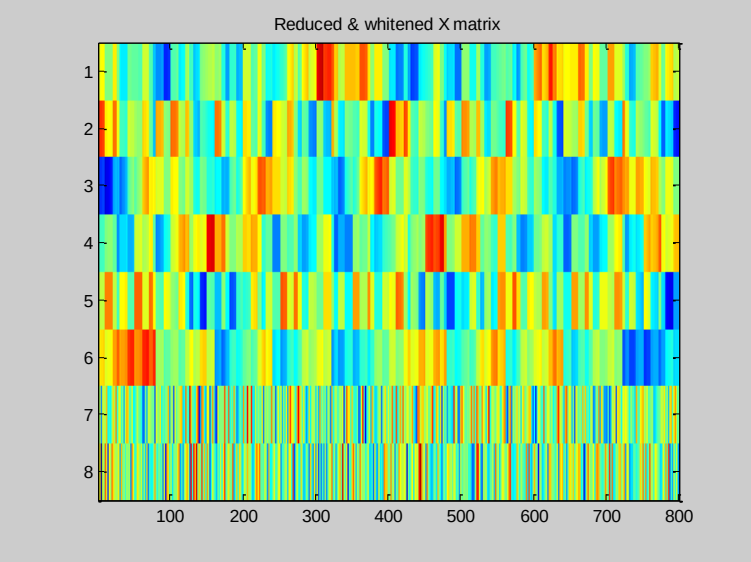
⑤ Calculate the vector of proportions



A simulated example

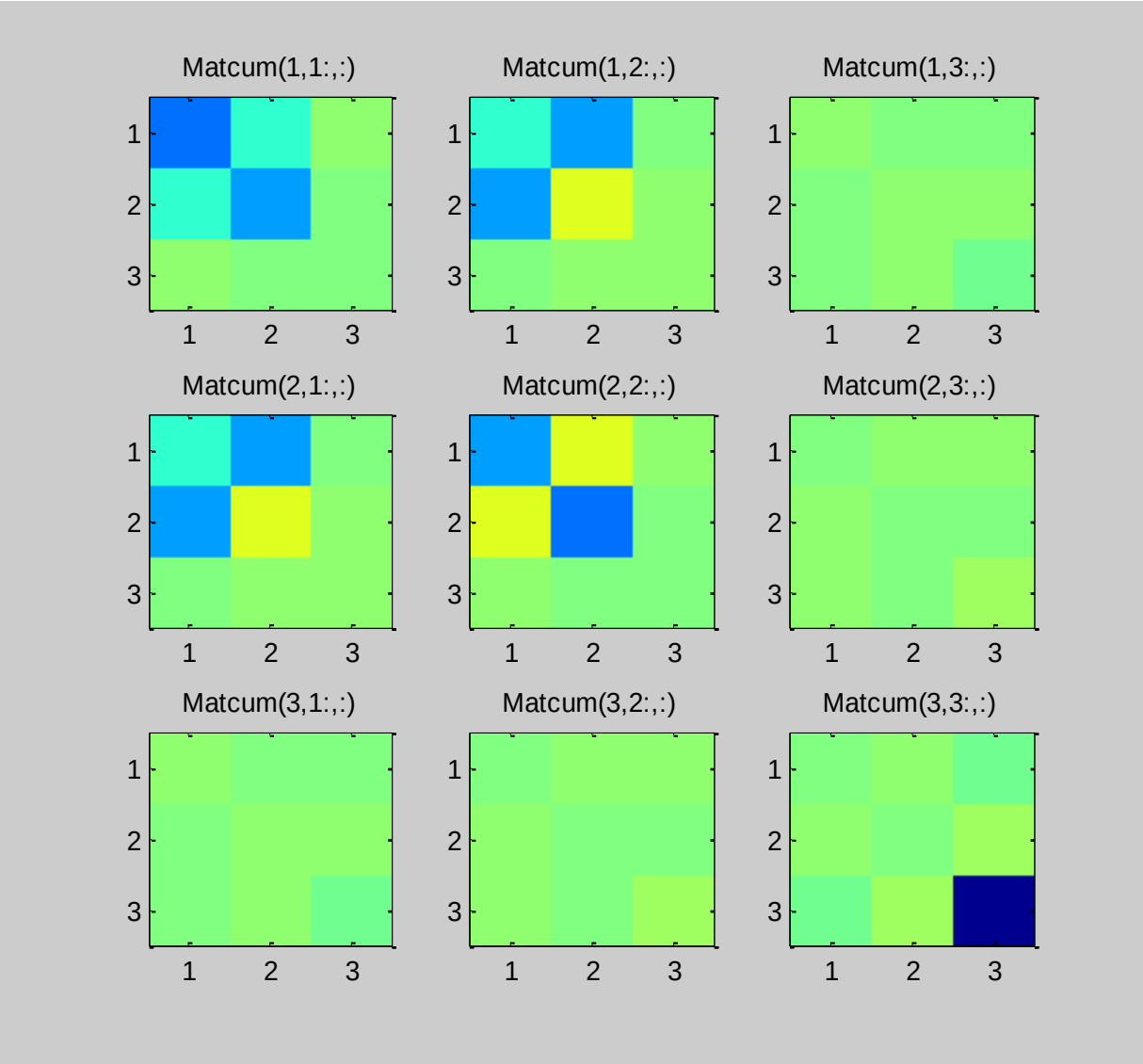
Initial X matrix



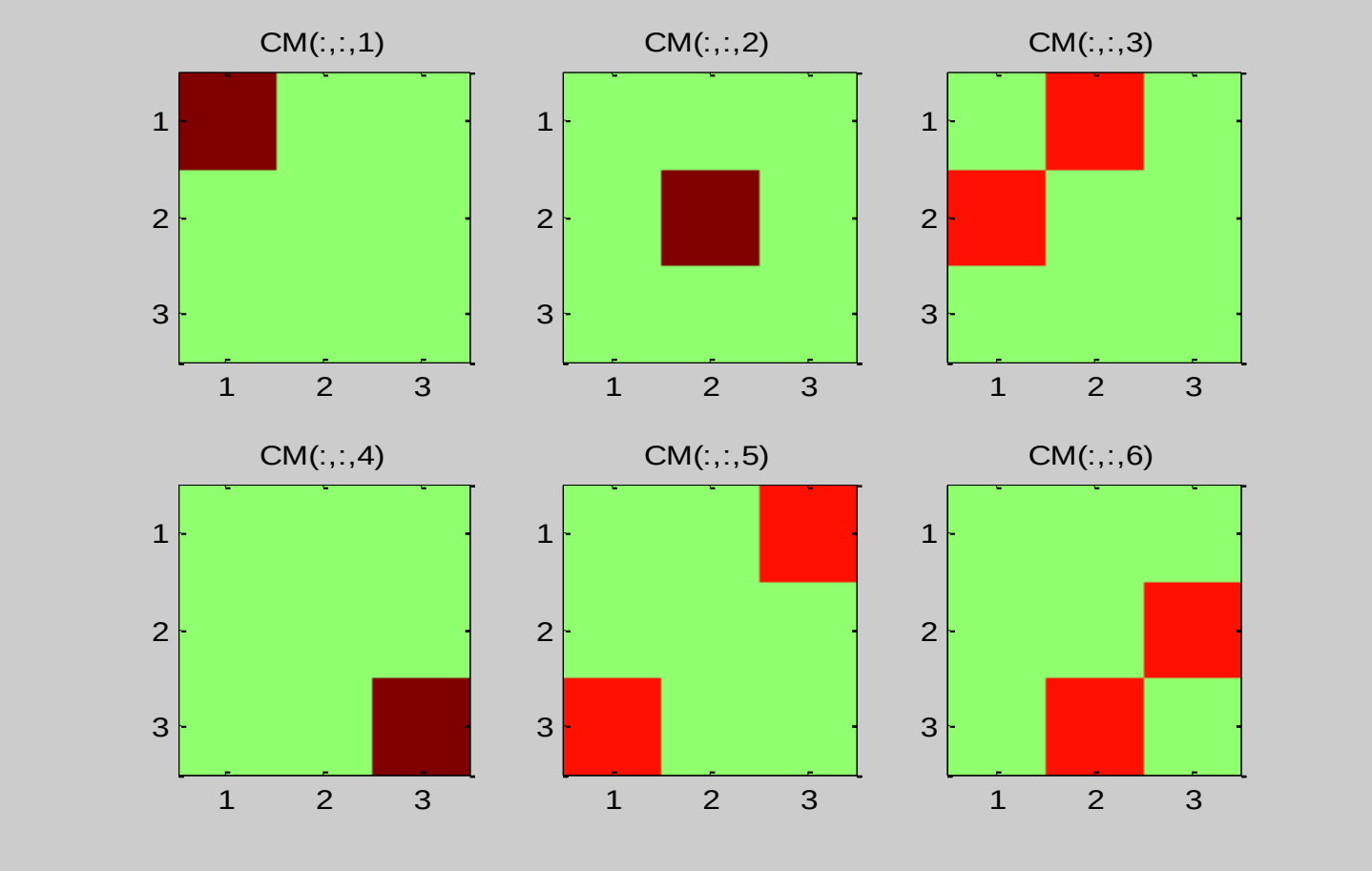


Smaller, whitened X matrix

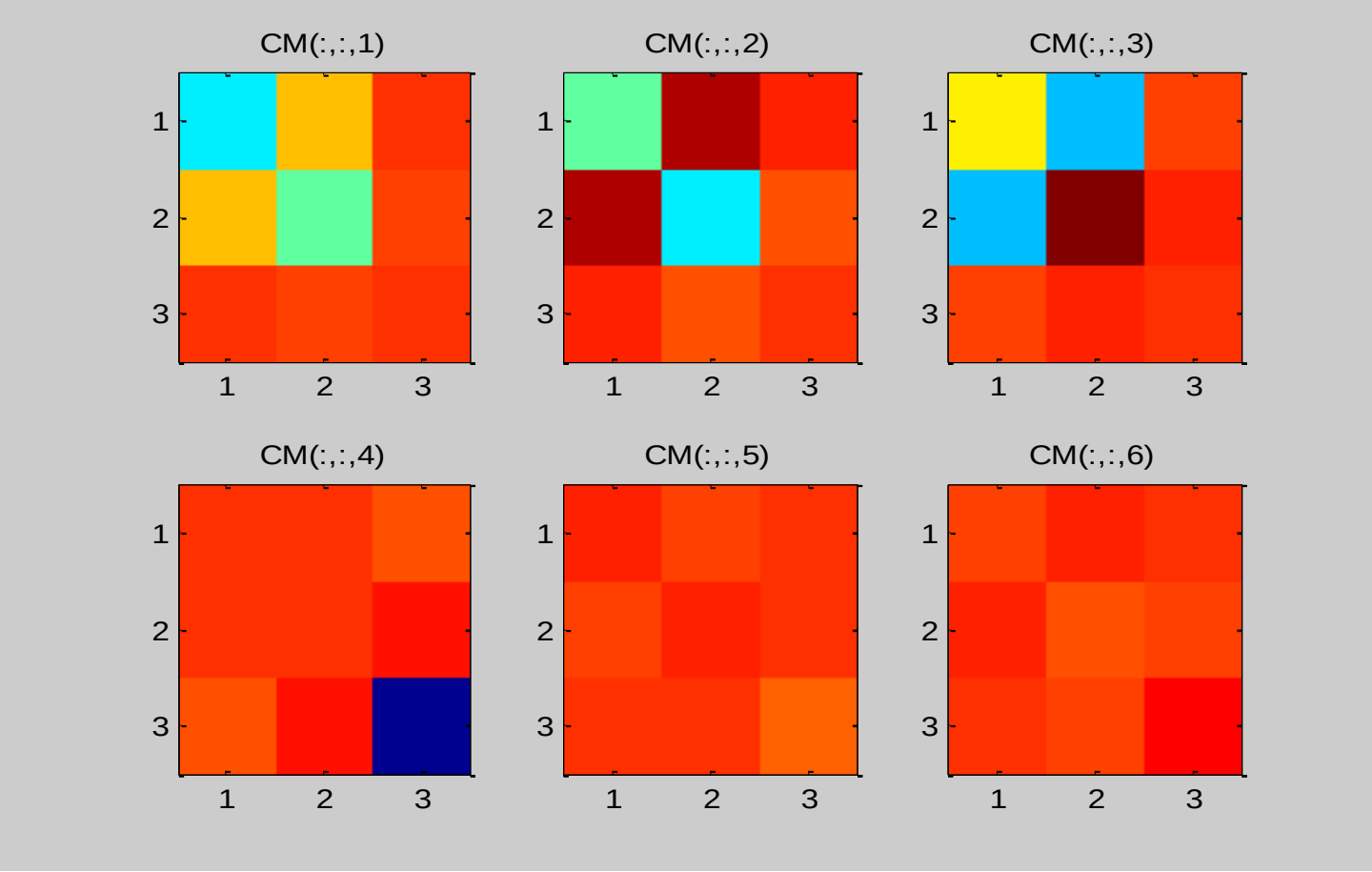
4-way tensor of cumulants



Initial orthogonal matrices to project cumulants



Eigenmatrices of cumulants

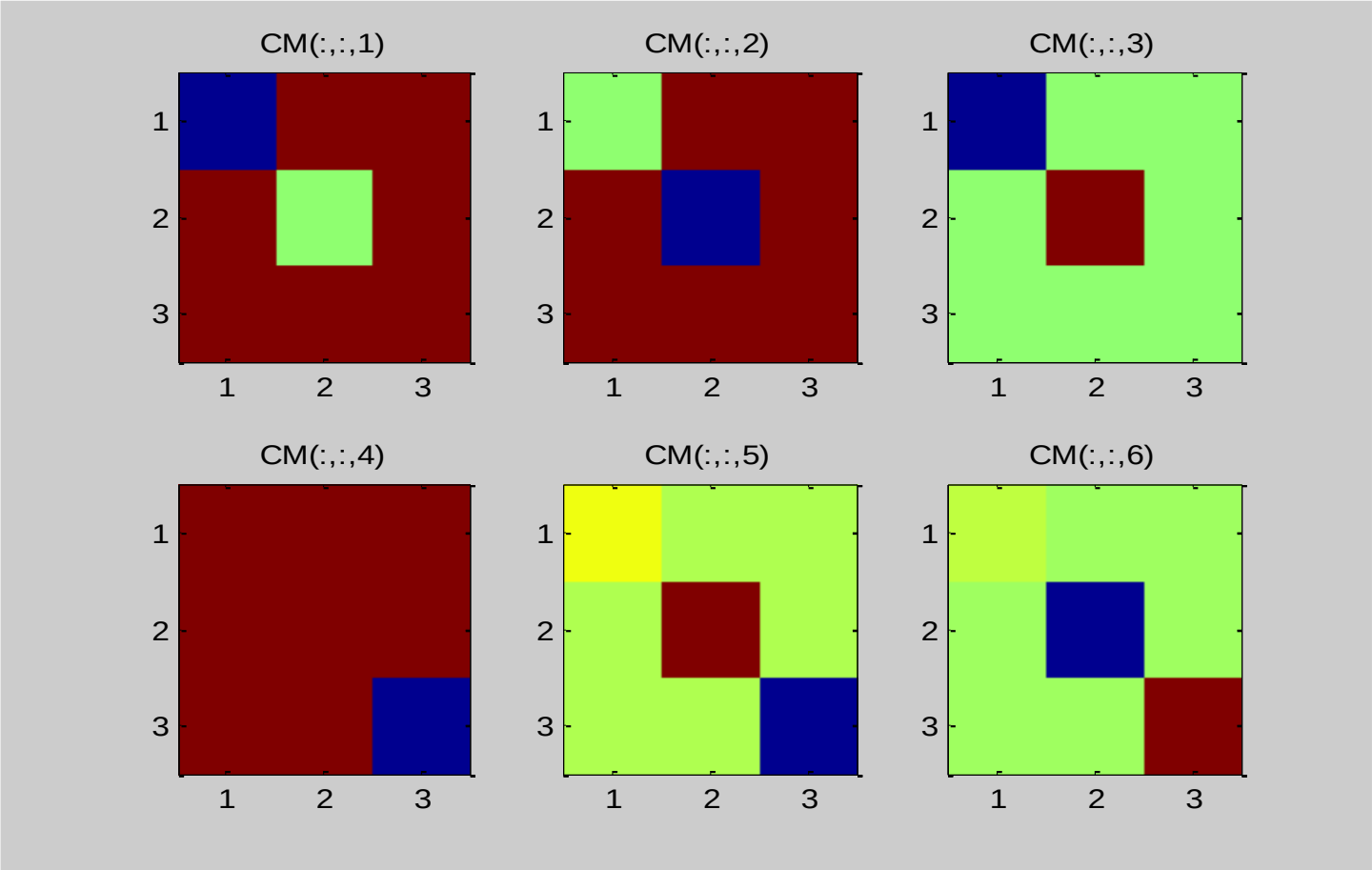


JADE

Joint diagonalization using the Jacobi algorithm,
based on Givens rotation matrices :

Minimize the sum of squares of the “off-diagonal" 4th-order cumulants
(i.e. the cumulants between different signals)

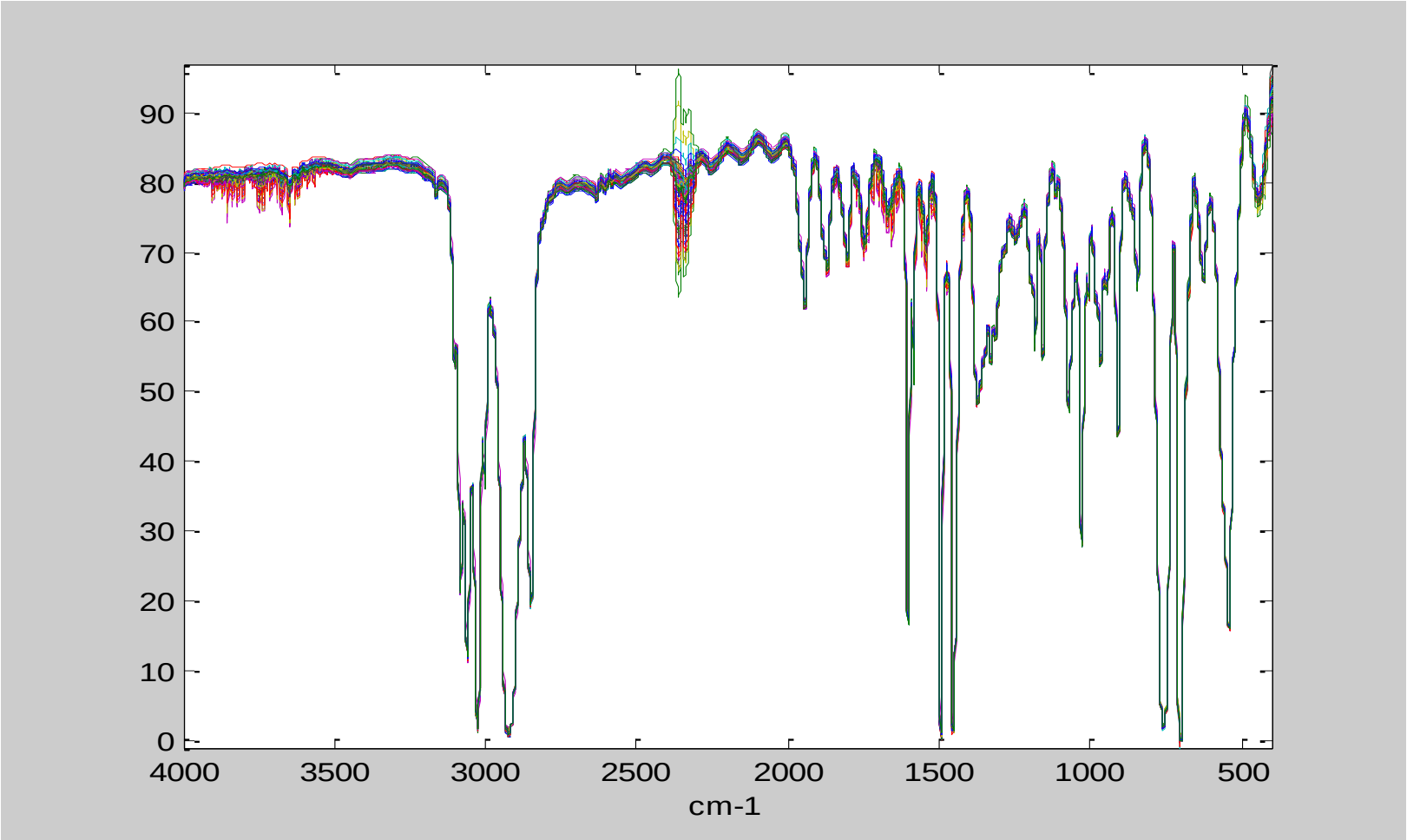
Diagonalised matrices



Applications of ICA

- Signal vectors (spectra, chromatograms, ...)
- Multi-way signals (3D-fluorescence, LC-MS, hyperspectral images, ...)
- Non-structured data vectors

Information Hidden in MIR Spectra



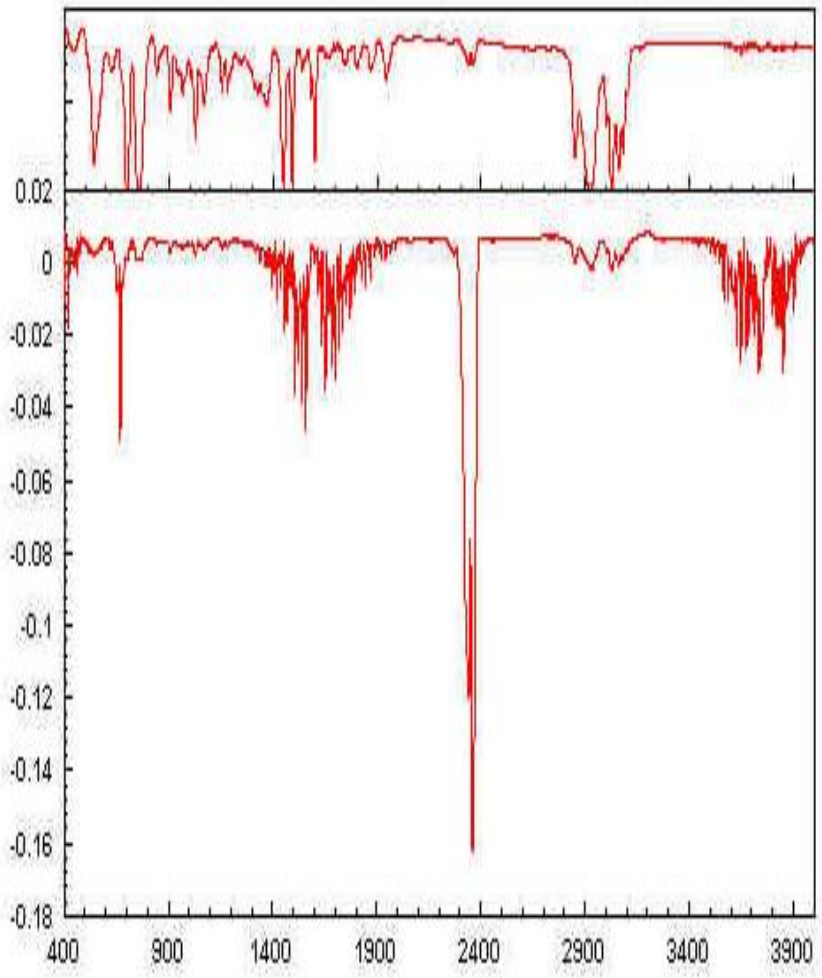
R.Aries, D. Lidiard, R. Spragg,

Spectrosc. Internat., 1990, 5(3) 41-44

ICA on 100 Mid-Infrared spectra

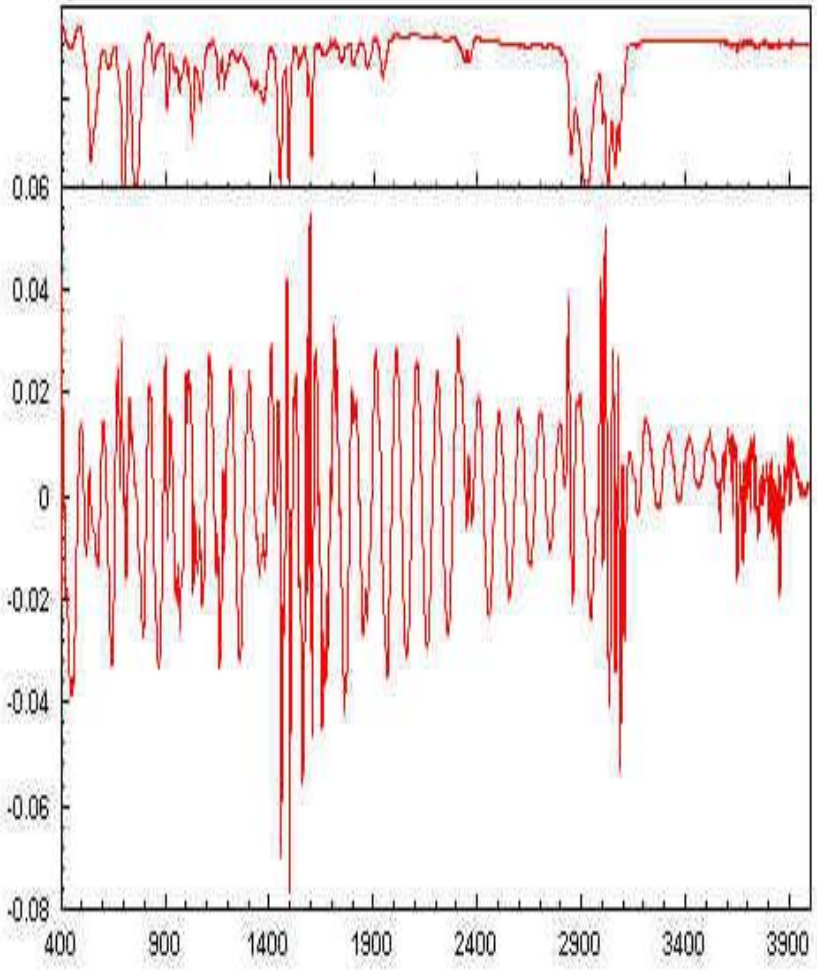
- Samples : a single polystyrene film
- MIR spectra taken on 100 spectrometers at end of production line
- Acquired from 4000cm^{-1} to 400cm^{-1}, at 1 cm^{-1}
- Pre-treatment of spectra :
  - normalised between 0 and 1
  - neither centred nor standardised

Results of a PCA on the polystyrene MIR data



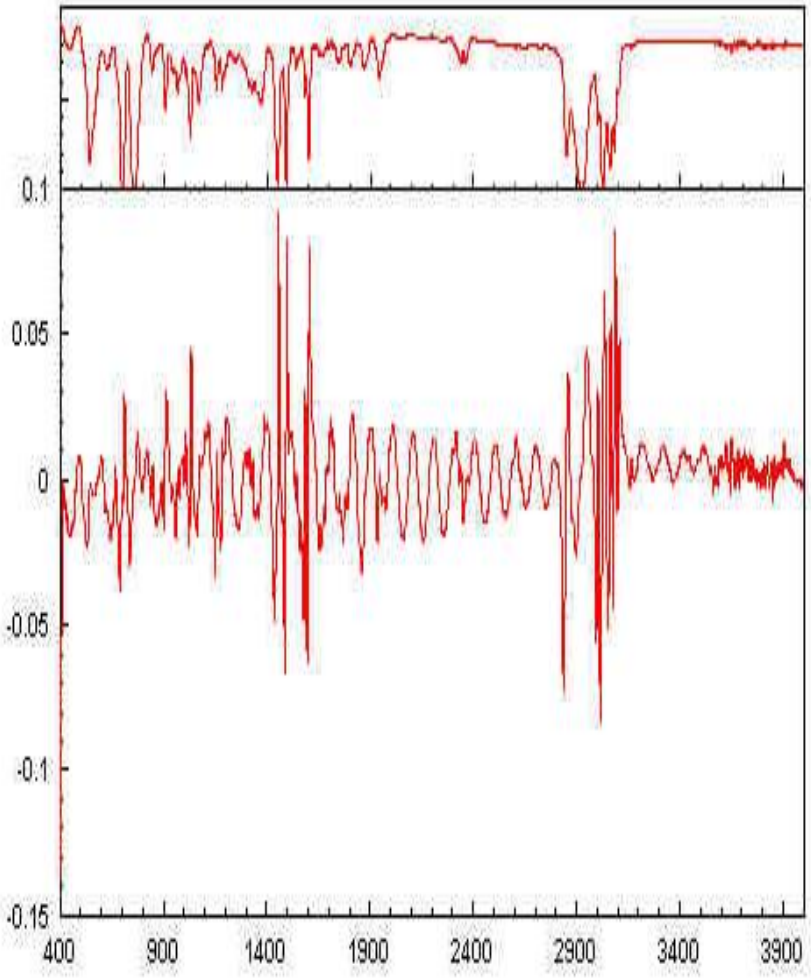
PC2 ≡ CO_2 & H_2O

Several phenomena in 1 PC



PC5

Interferogram & 1° derivative

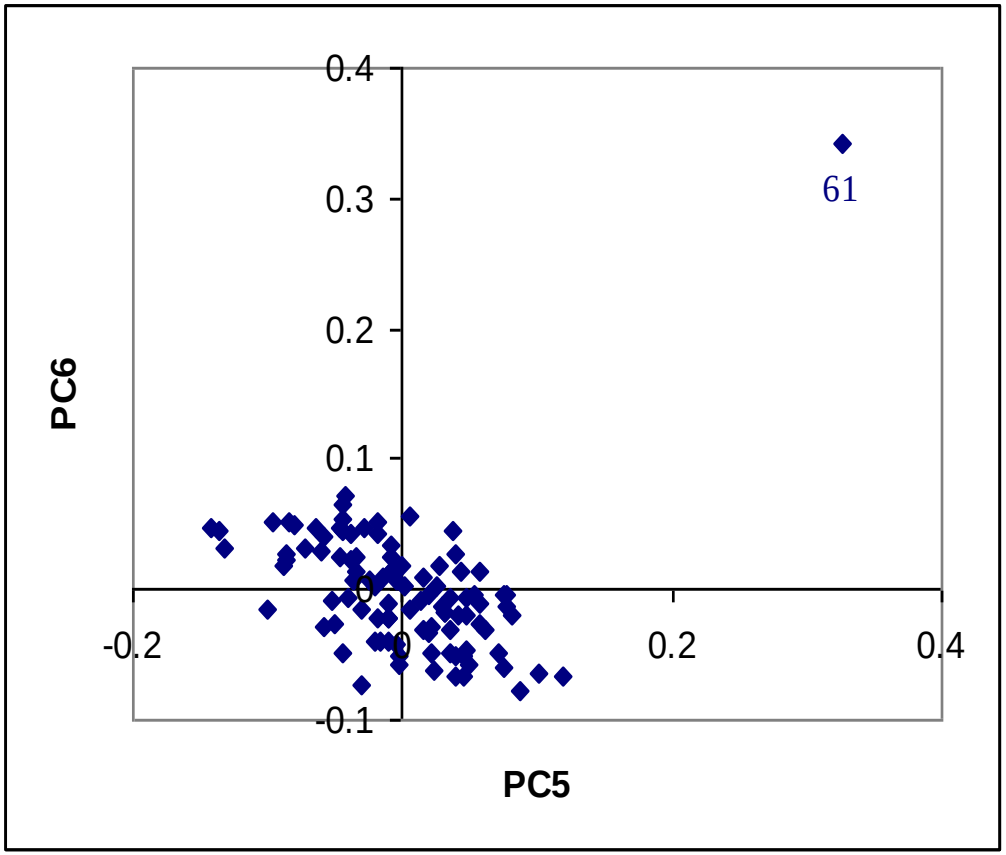


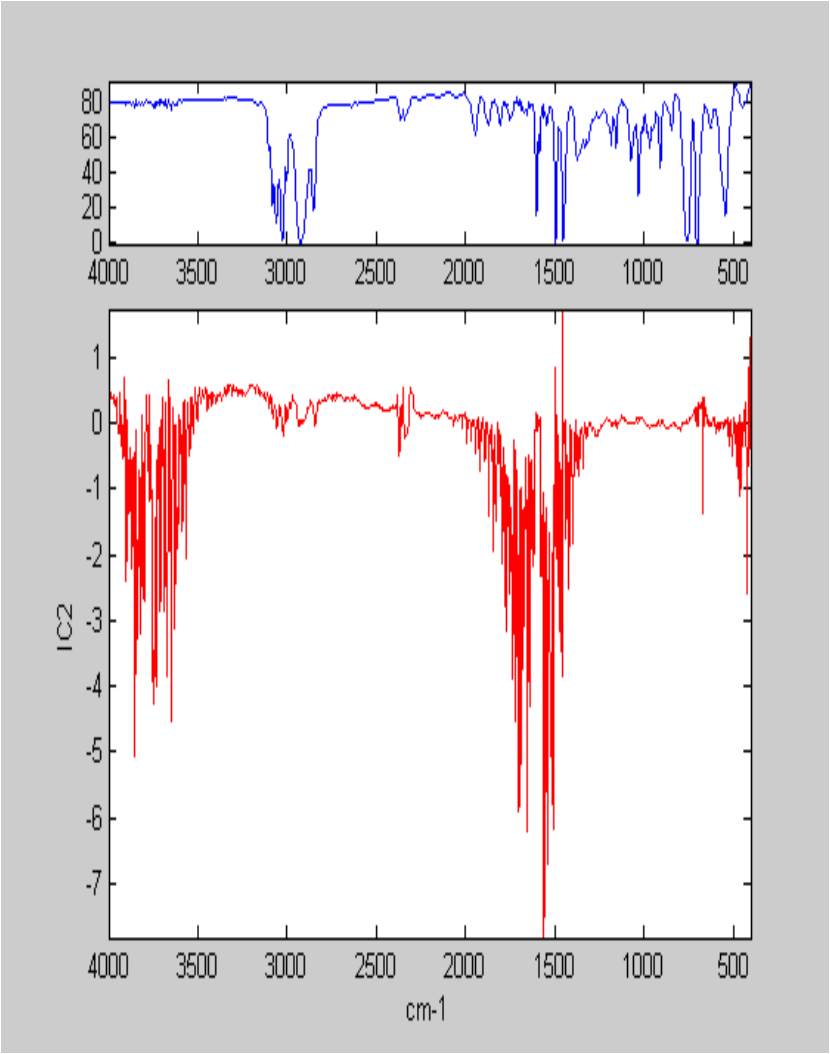
PC6

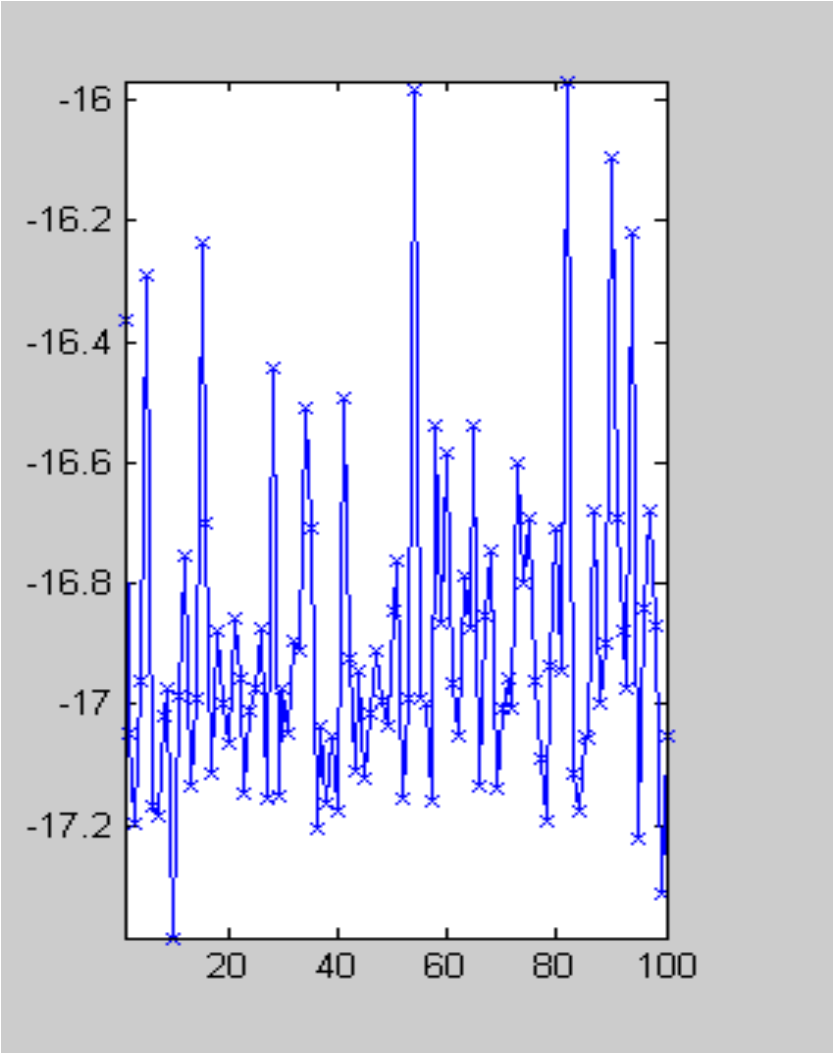
Interferogram & 1° derivative

Same phenomena in several PCs

PC5-PC6 Scores Plot

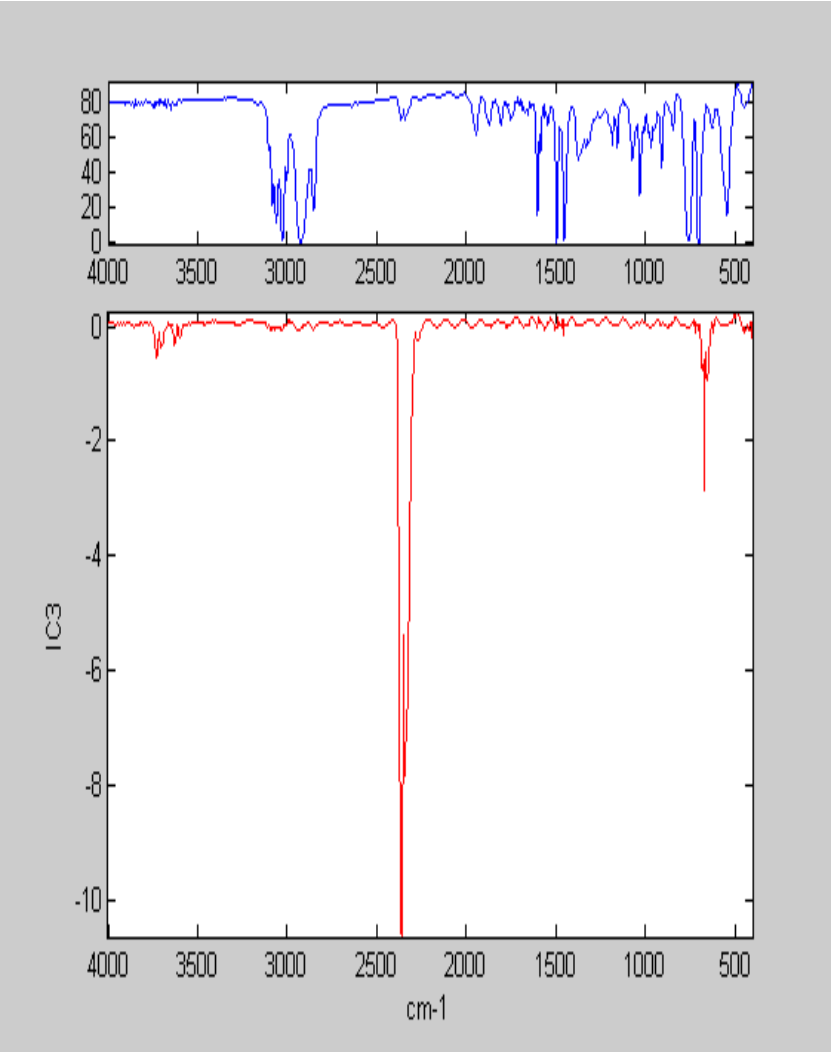


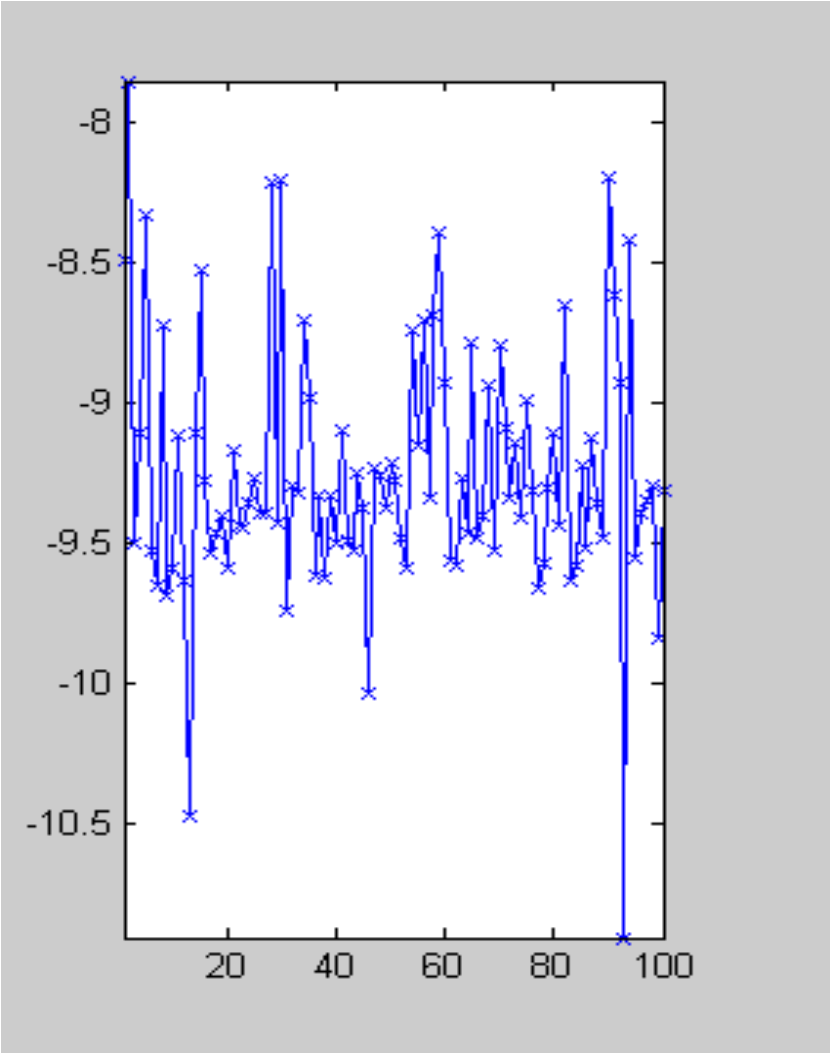




IC2/7 signal looks like water vapour spectrum

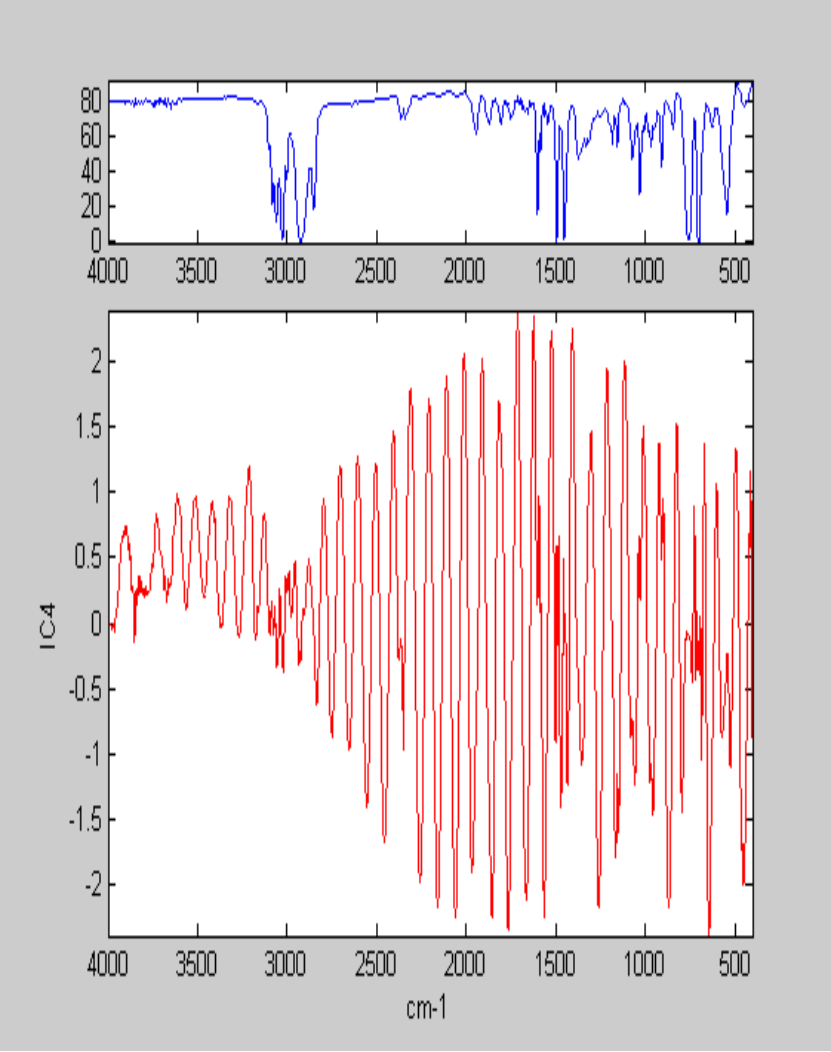
Due to variations in moisture content of air

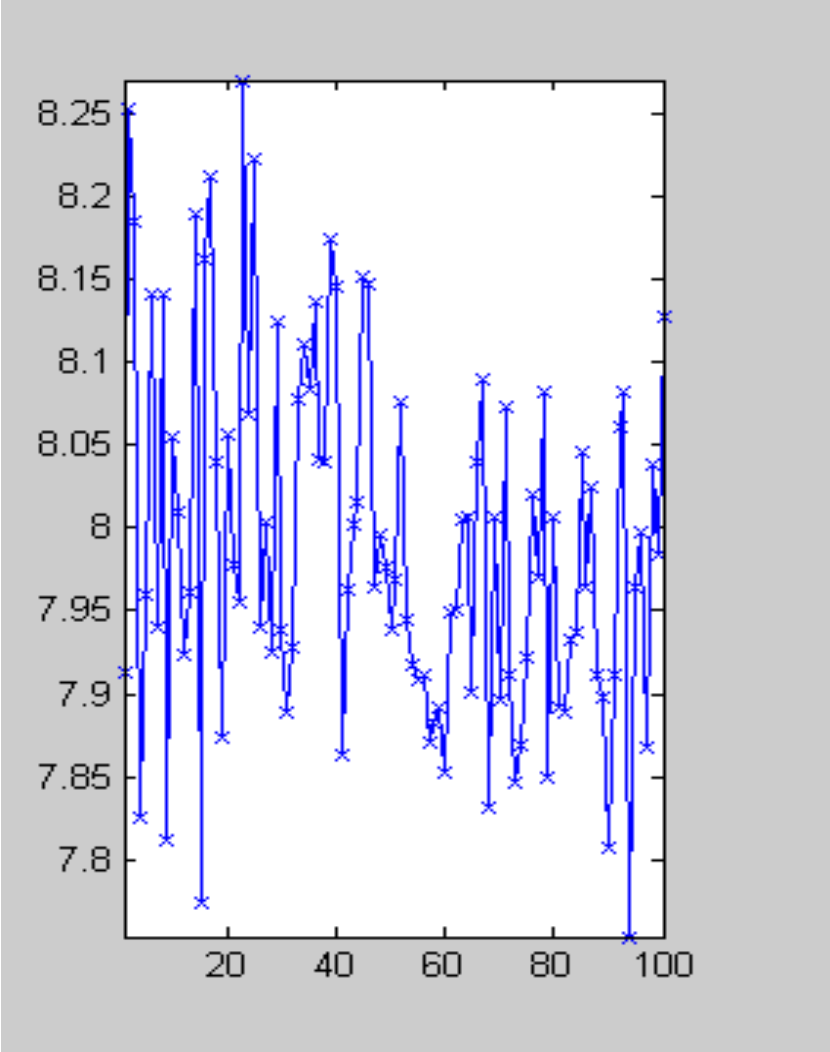




IC3/7 signal looks like spectrum of CO_2

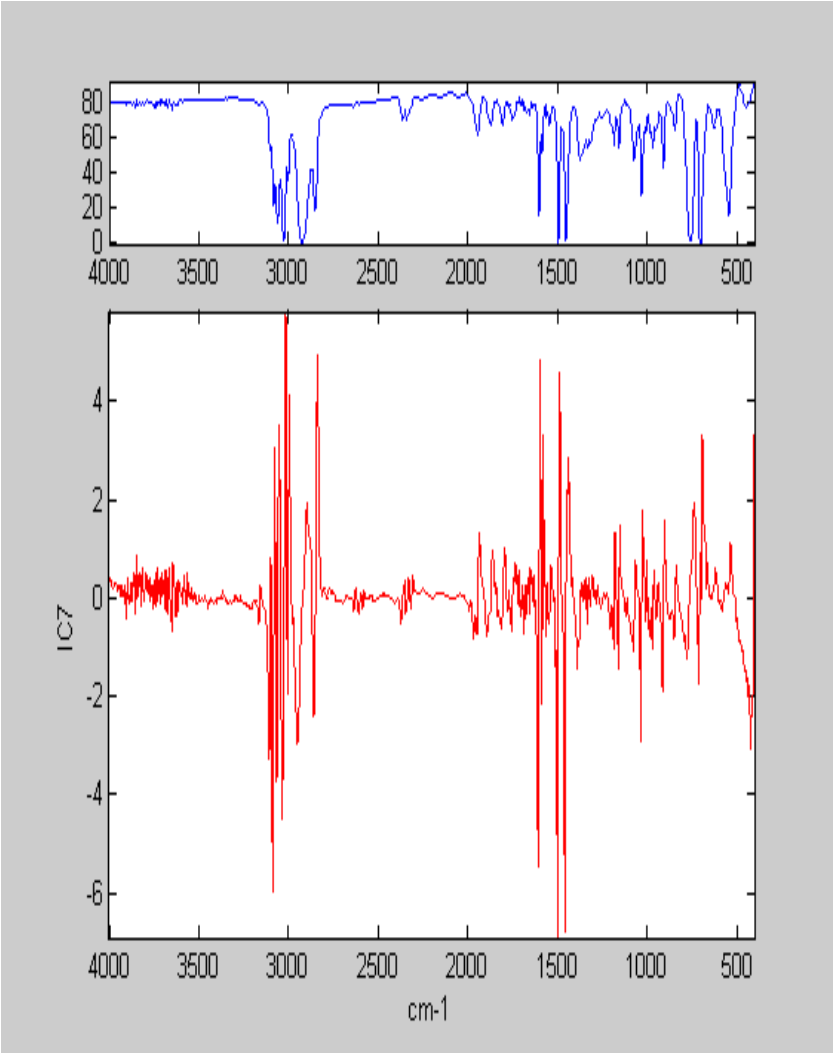
Due to variations in CO_2 content of air

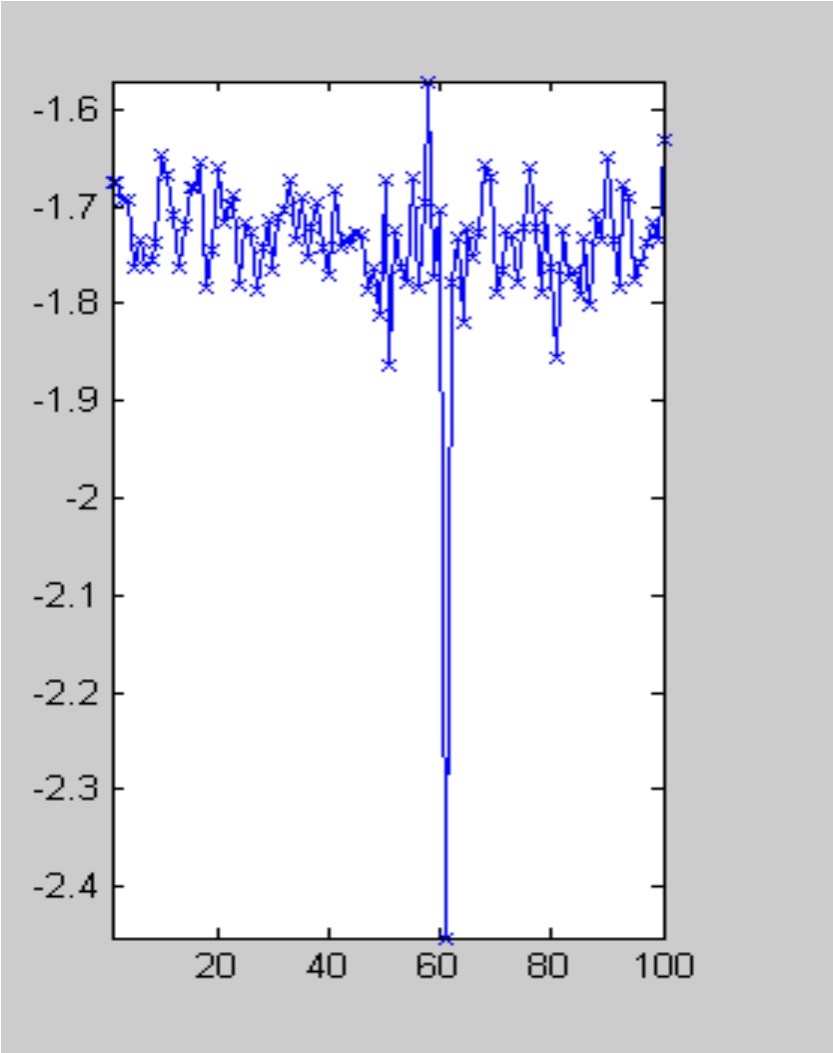




IC4/7 signal looks like beats of a simple interferogram

Due to variations in optical path of polystyrene sample ?





IC7/7 signal looks like first derivative of average MIR spectrum

Due to one spectrometer (N°61) being badly adjusted (frequency shift)

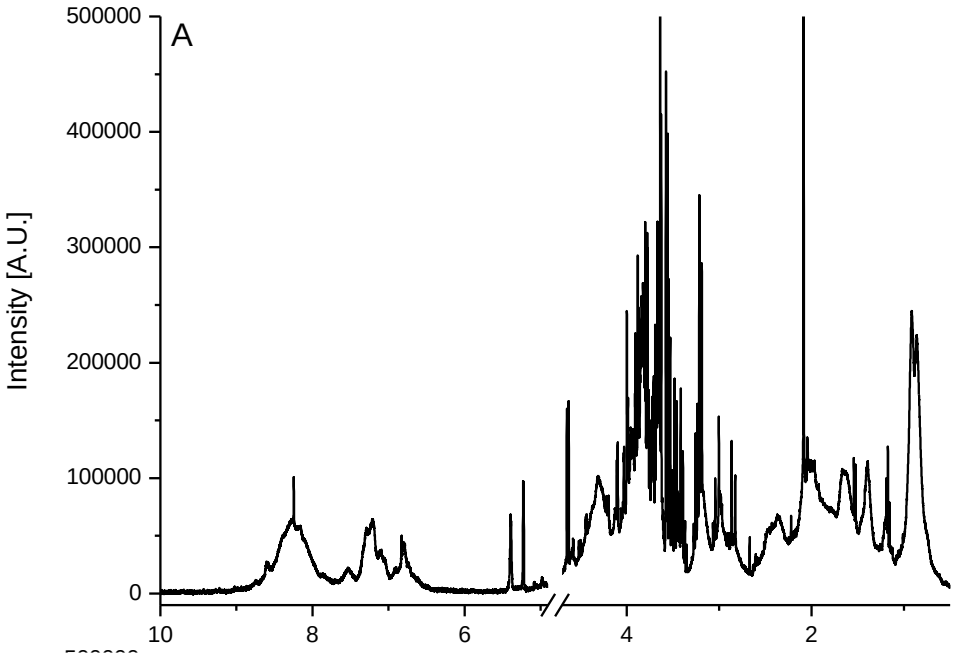
Determination of rice type by ^{1}H NMR

Calibration set :

18 Basmati,
18 non-Basmati long grain rice,
6 round grain rice

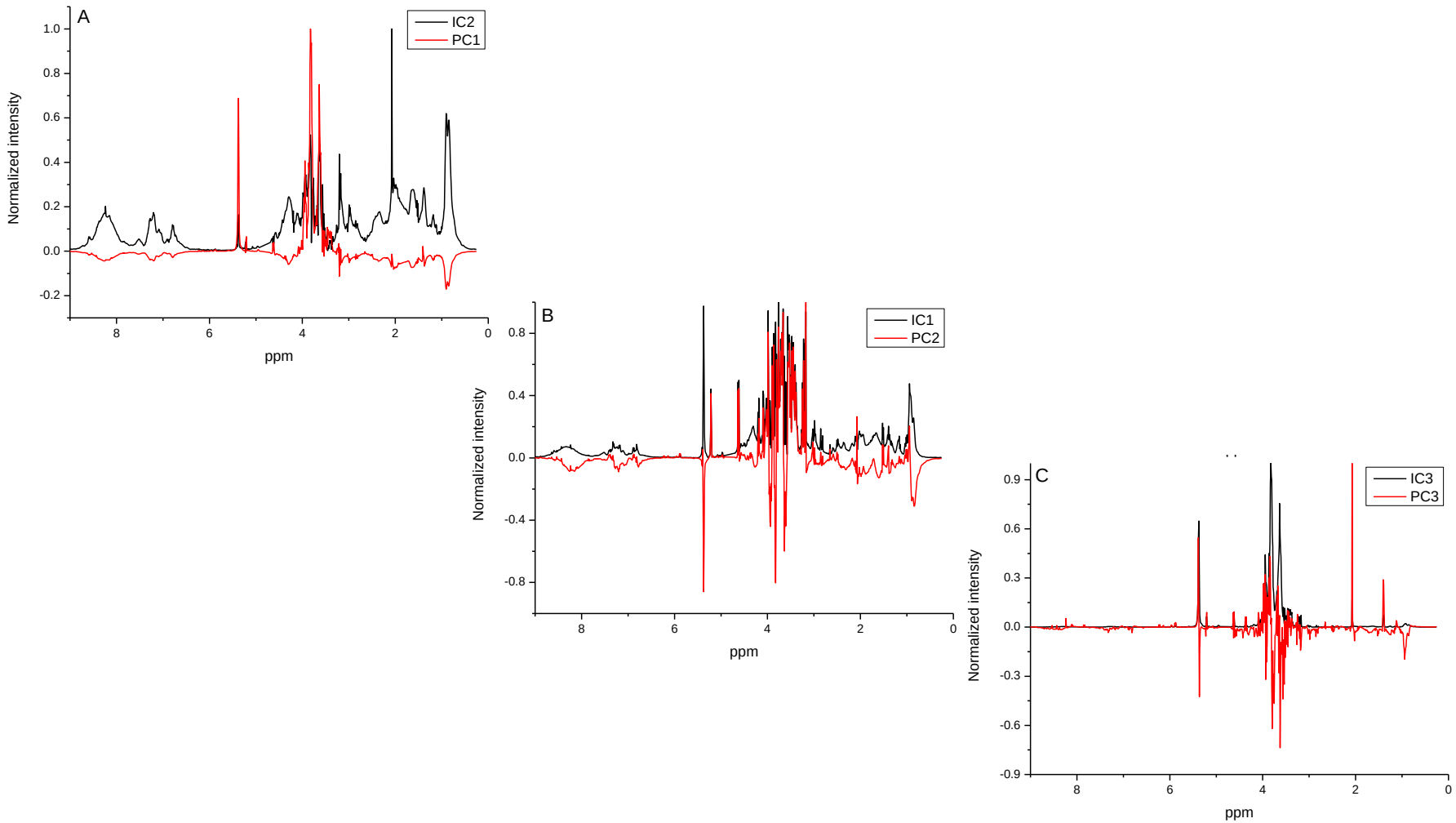
Test set :

4 Basmati,
4 non-Basmati long grain rice
2 round grain rice



Yulia B. Monakhova, Douglas N. Rutledge, Andreas Roßmann, Hans-Ulrich Waiblinger, Manuela Mahler, Maren Ilse, Thomas Kuballa and Dirk W. Lachenmeier
Determination of rice type by 1 H NMR spectroscopy in combination with different chemometric tools
J. Chemometrics 2014, 28: 83–92

ICs & PCs extracted from NMR spectra



Classification results based on LDA, FDA, PLS-DA & ICA

Method		LDA	FDA	PLS-DA	ICA
False-positive samples [a,b]		3	1	1	**0**
False-negative samples [a,b]		4	2	1	**1**
Sensitivity [%] [a,b]		83	94	94	**100**
Specificity [%] [a,b]		88	89	96	**96**
Overall correct classification rate [%][d]	[b]	83	81	92	**90**
	[c]	50	70	80	**90**

Summary of the classification results by different methods :

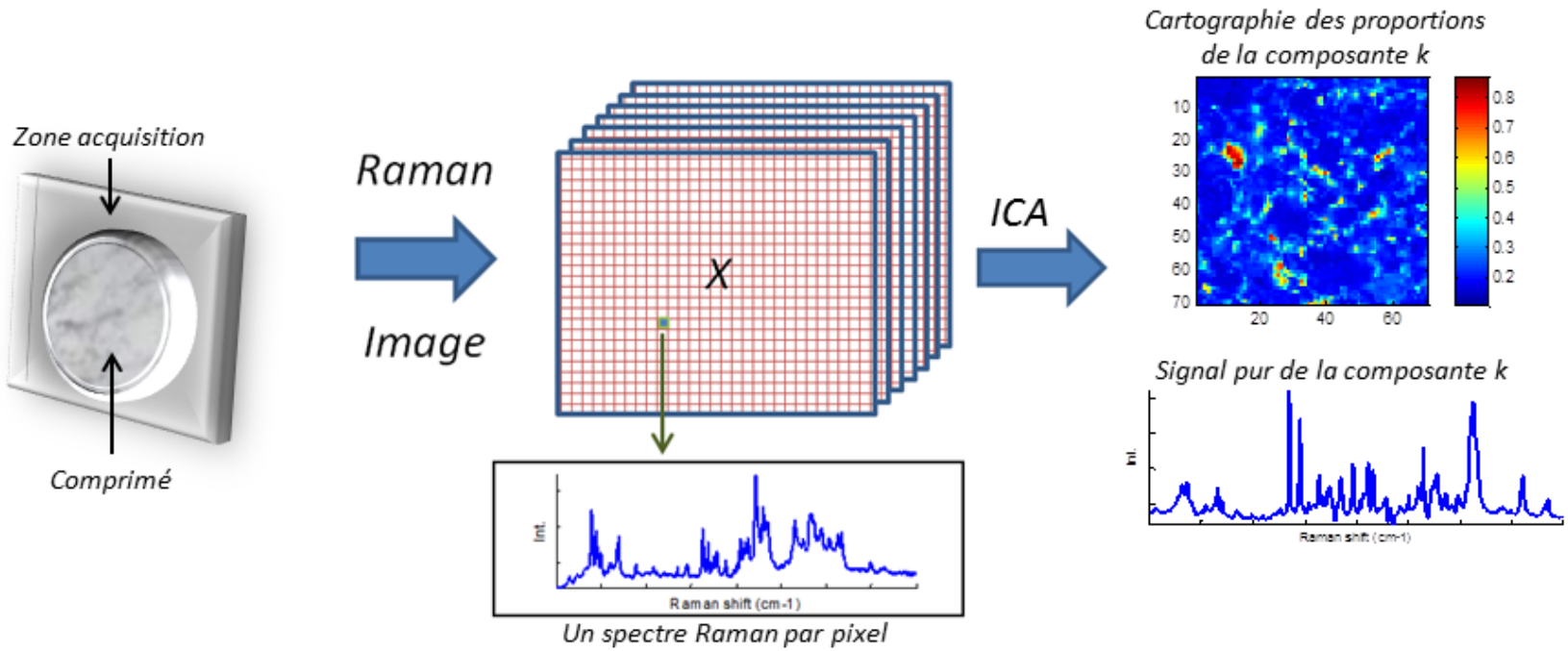
[a] was calculated for the Basmati group (n=18)

[b] leave-out-one cross validation (n=42)

[c] independent test set (n=10)

[d] considering all three groups (Basmati, non-Basmati long grain rice, round grain rice)

ICA applied to Multi-way data
Raman hyperspectral images



Hyperspectral images
acquired for an authentic and a suspect pharmaceutical pill

M Boiret, D N Rutledge, N Gorretta, Y-M Ginot, J-M Roger
Utilisation de la microscopie Raman et des methodes chimiometriques pour la detection de comprimes contrefaits,
Spectra Analyse, 2014, 298, pp 74-80

Images collected using a PerkinElmer RS400 system

Microscope coupled to spectrophotometer with 785nm 400mW laser

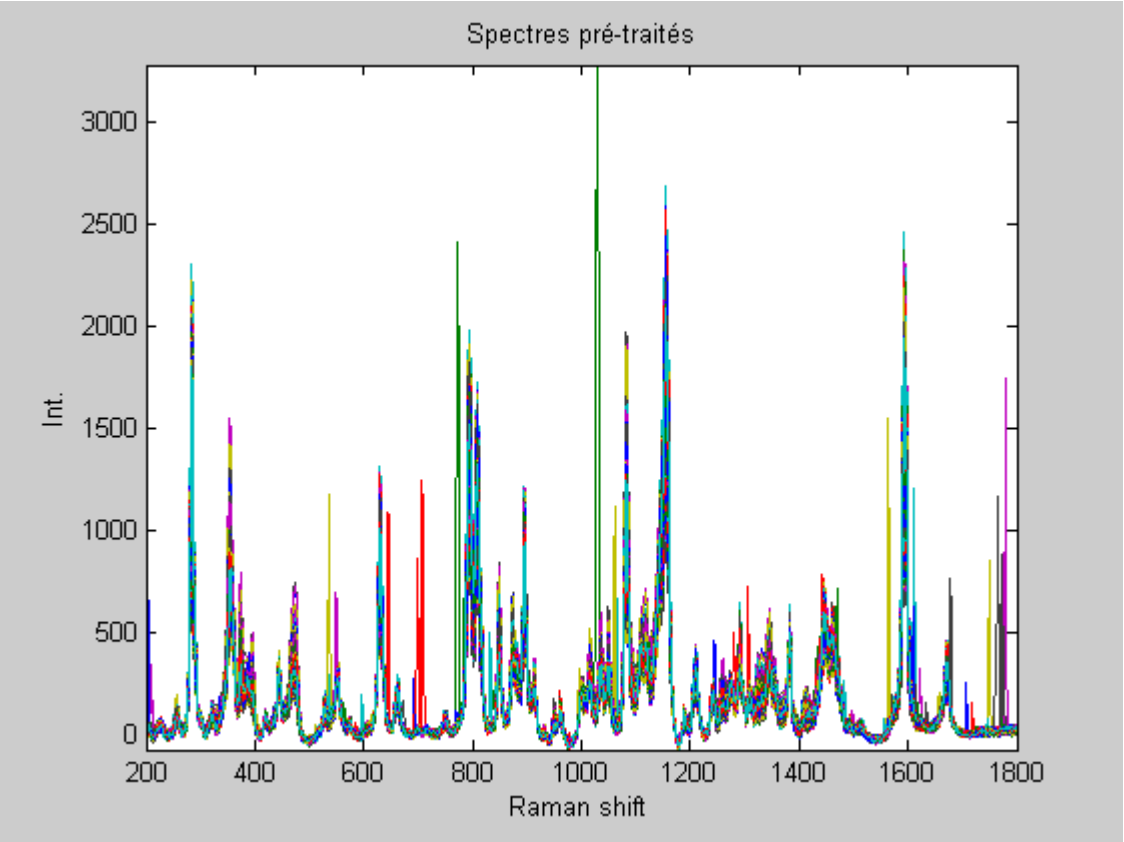
CCD sensor (Charge-Coupled Device)

Sample on a motorized stage with a pitch of 50 microns

Raman spectra from 3200cm $^{-1}$ to 200cm $^{-1}$
Spectral resolution 2 cm $^{-1}$

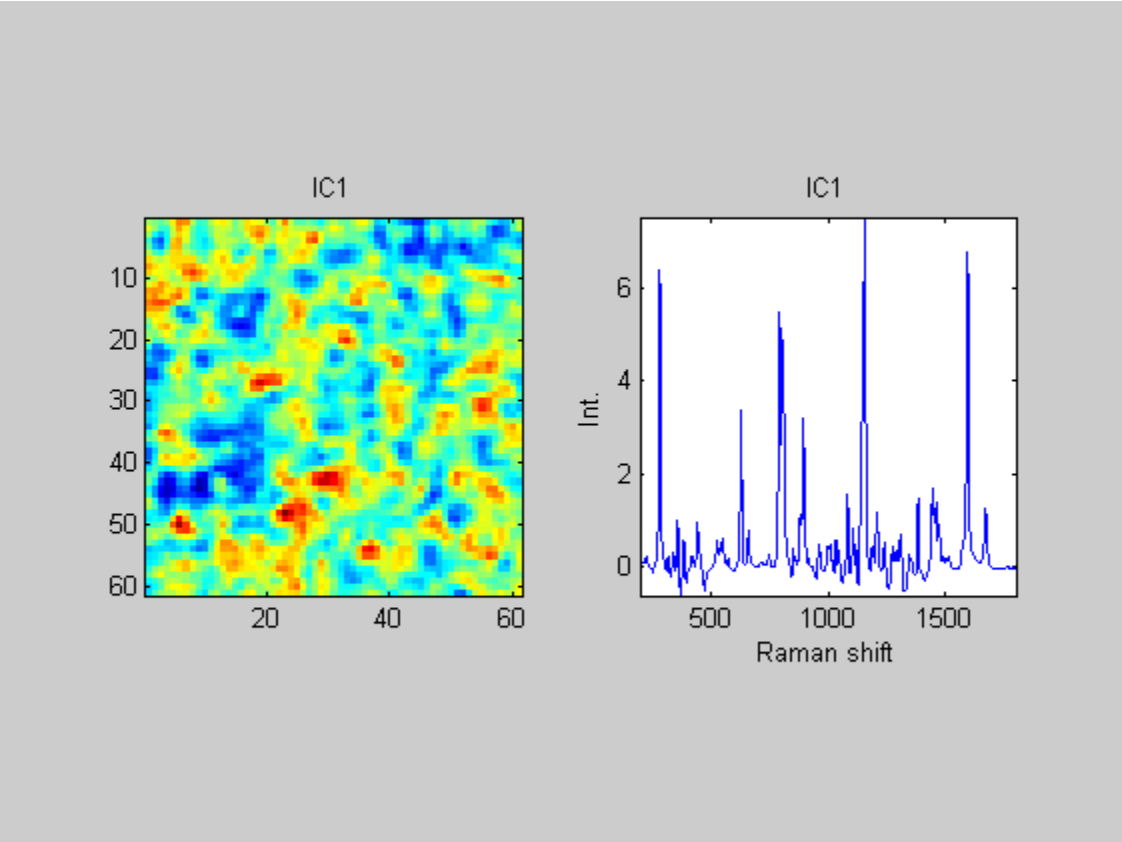
26 000 spectra on a surface of about 8mm * 8mm

Spectral matrix = 3 600 x 801



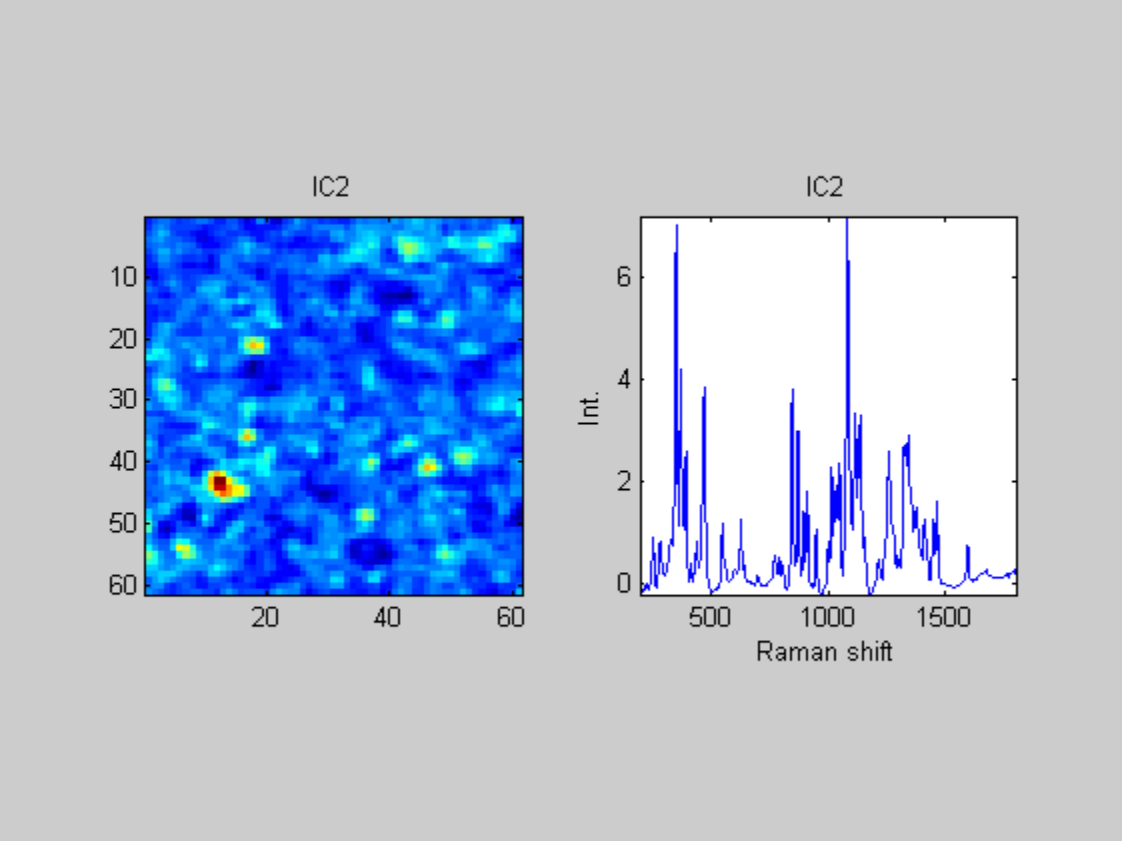
Refolded proportions

Source signal



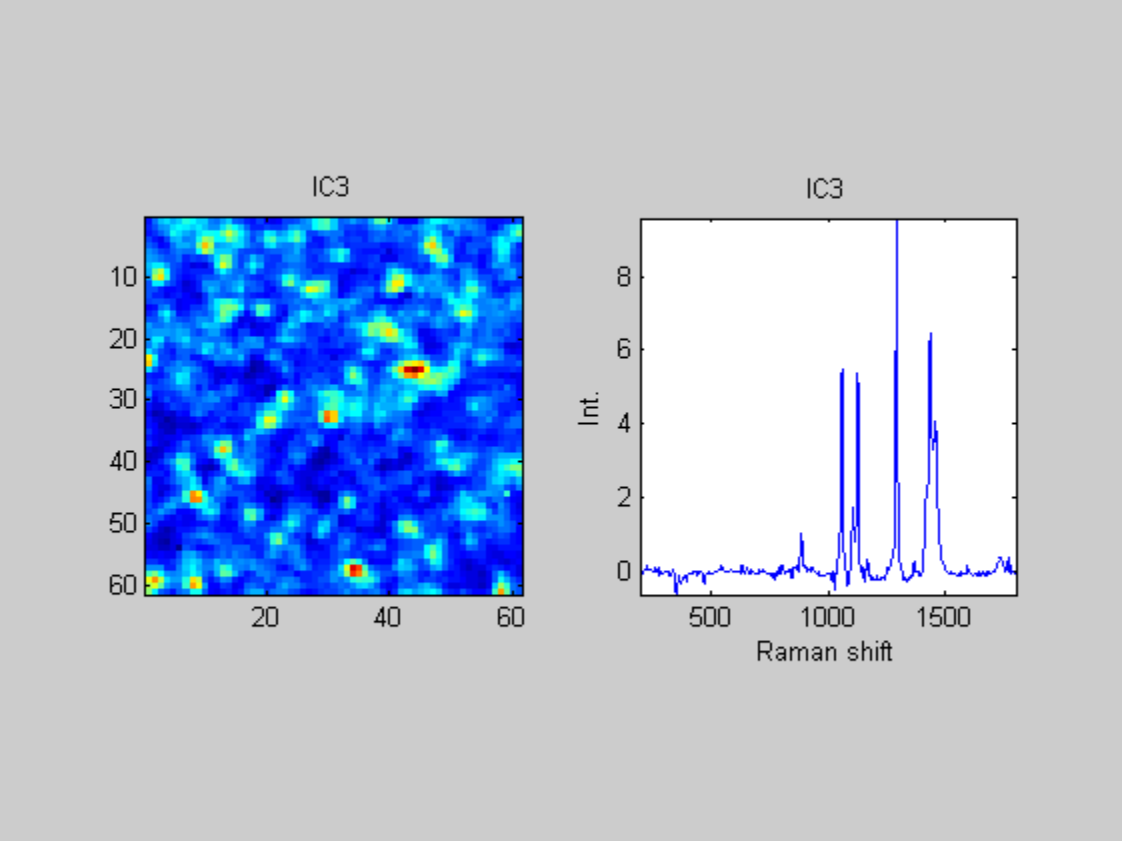
Refolded proportions

Source signal

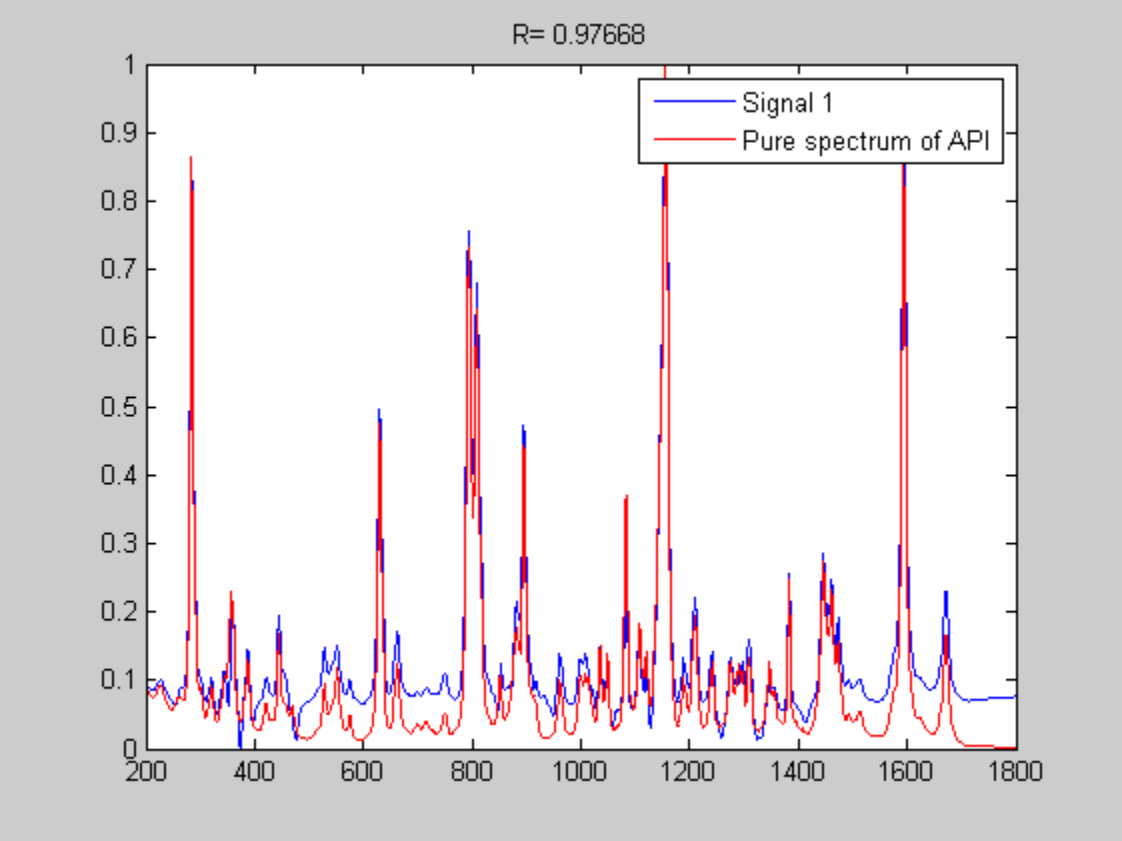


Refolded proportions

Source signal

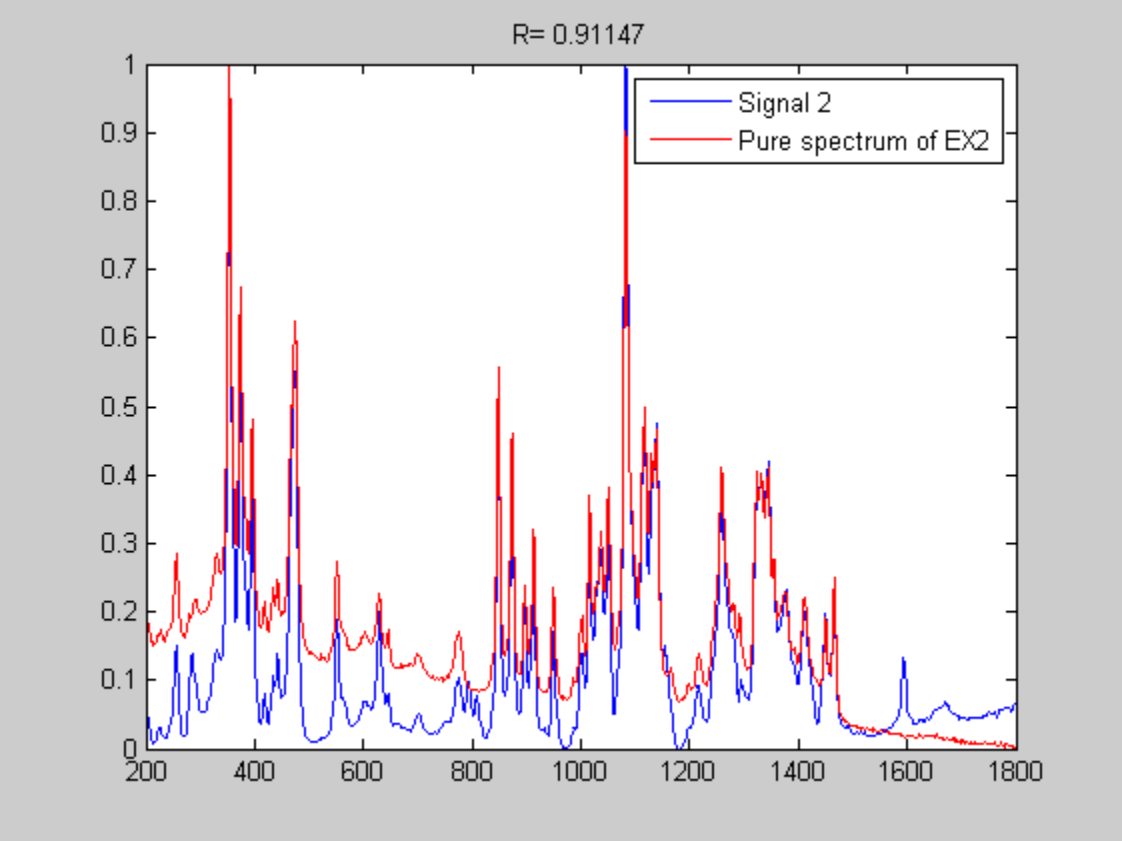


Library spectrum Source signal



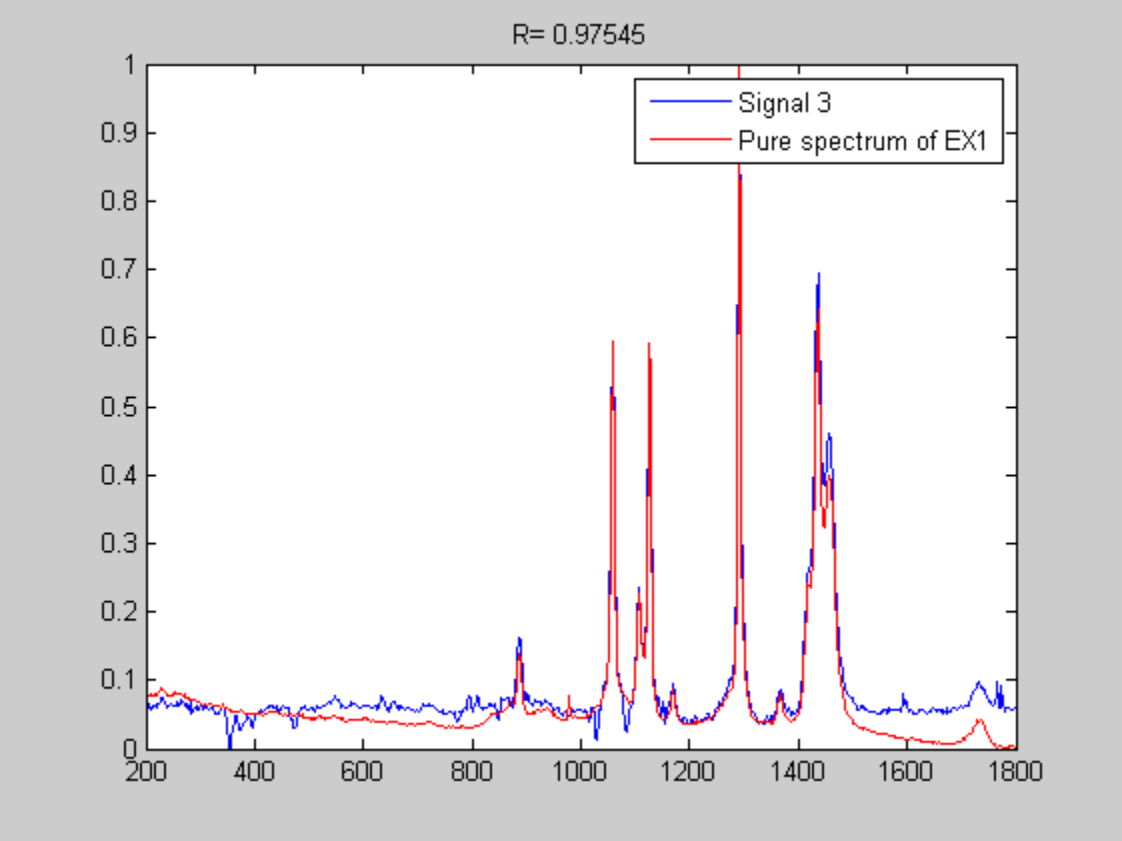
R=0.97668

Library spectrum Source signal



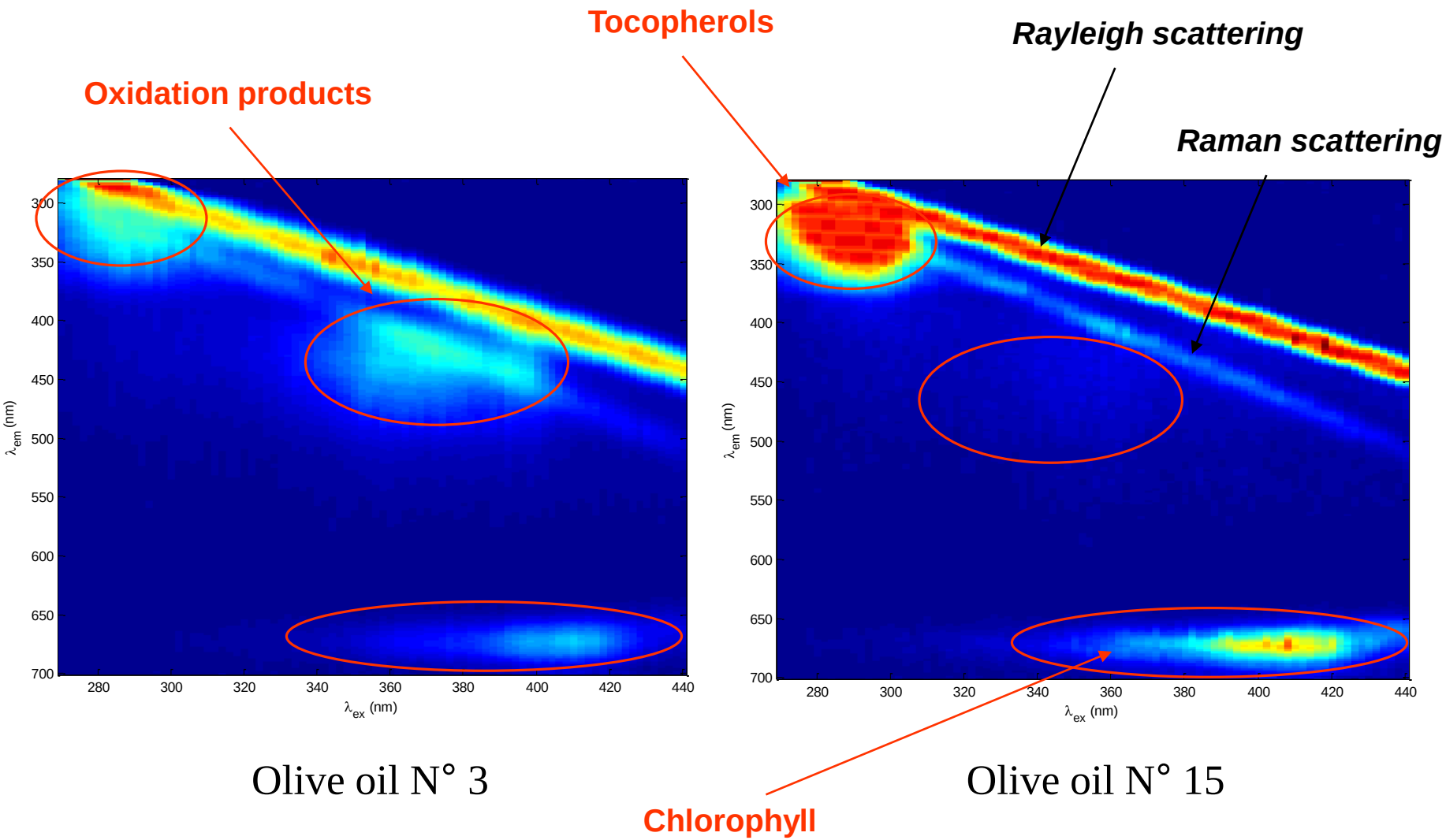
R=0.91147

Library spectrum Source signal

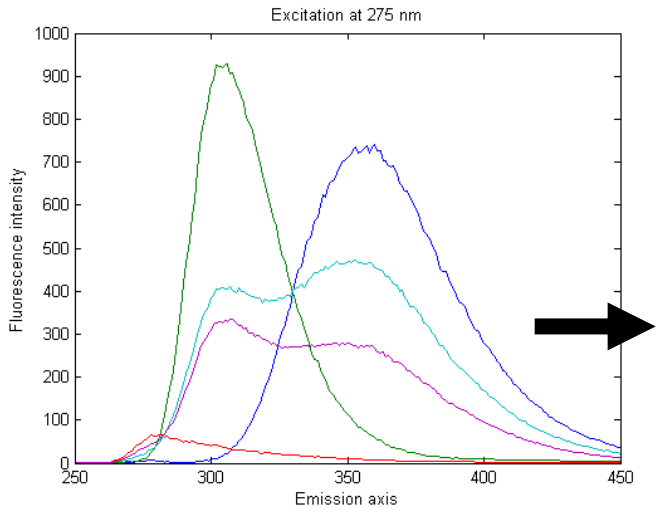


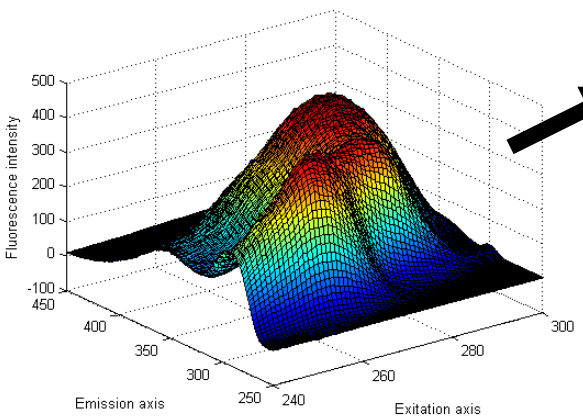
R=0.97545

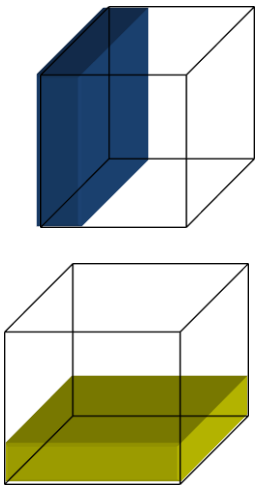
ICA applied to Multi-way data
3D-Fluorescence spectra of olive oils

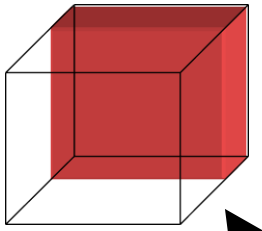


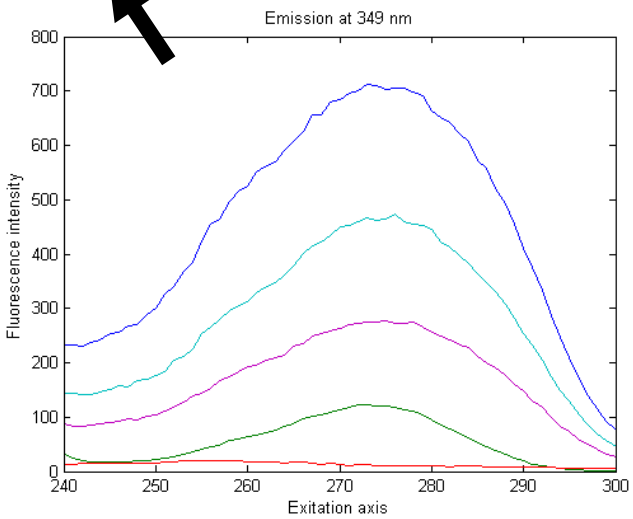
The 3 slabs of fluorescence data

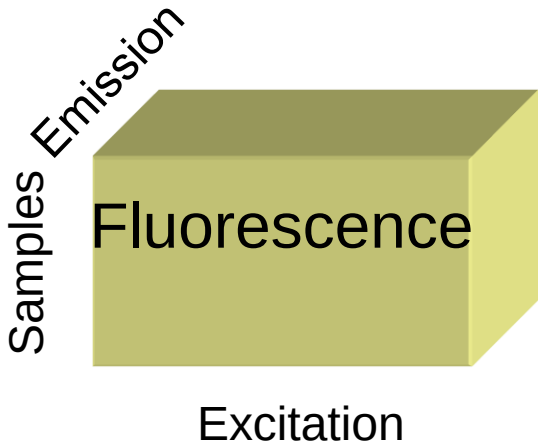






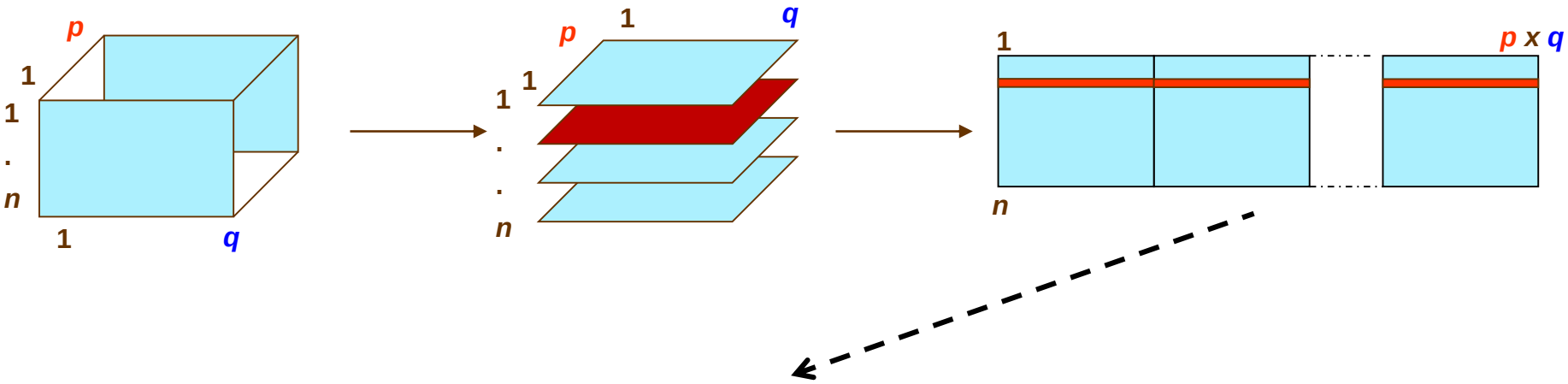






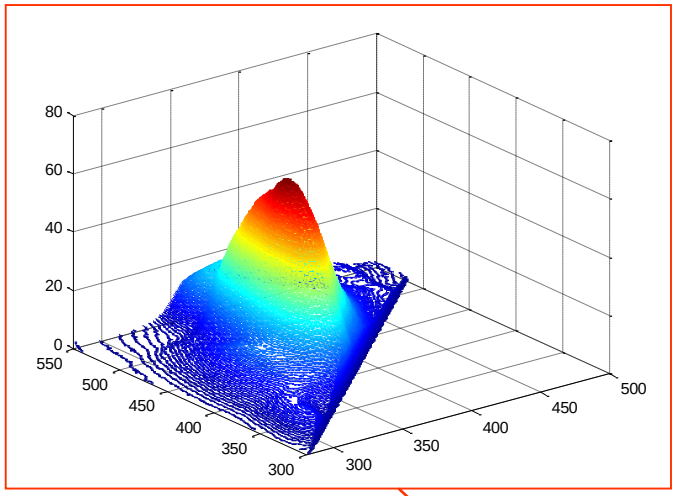
Apply ICA to the unfolded matrices

Unfold the cube to form a matrix ("matricise" the tensor)



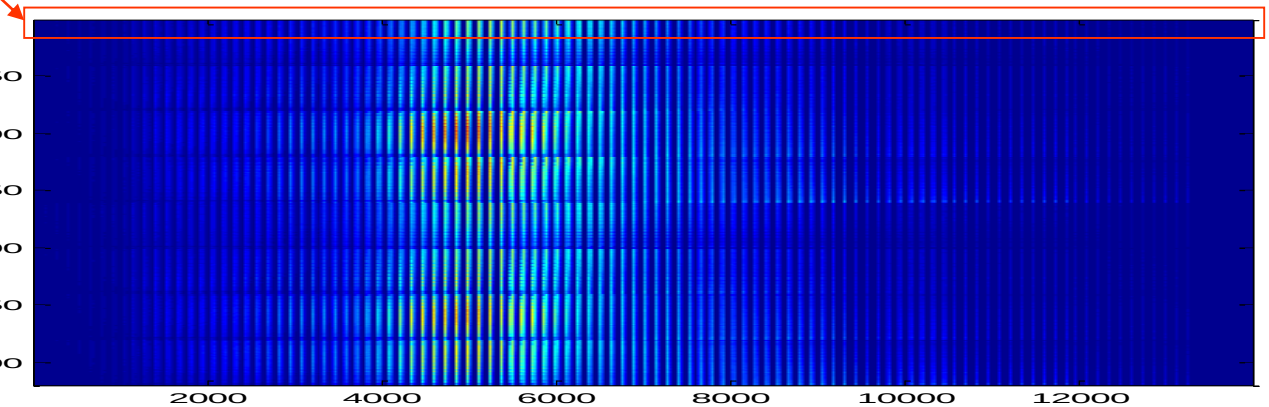
ICA

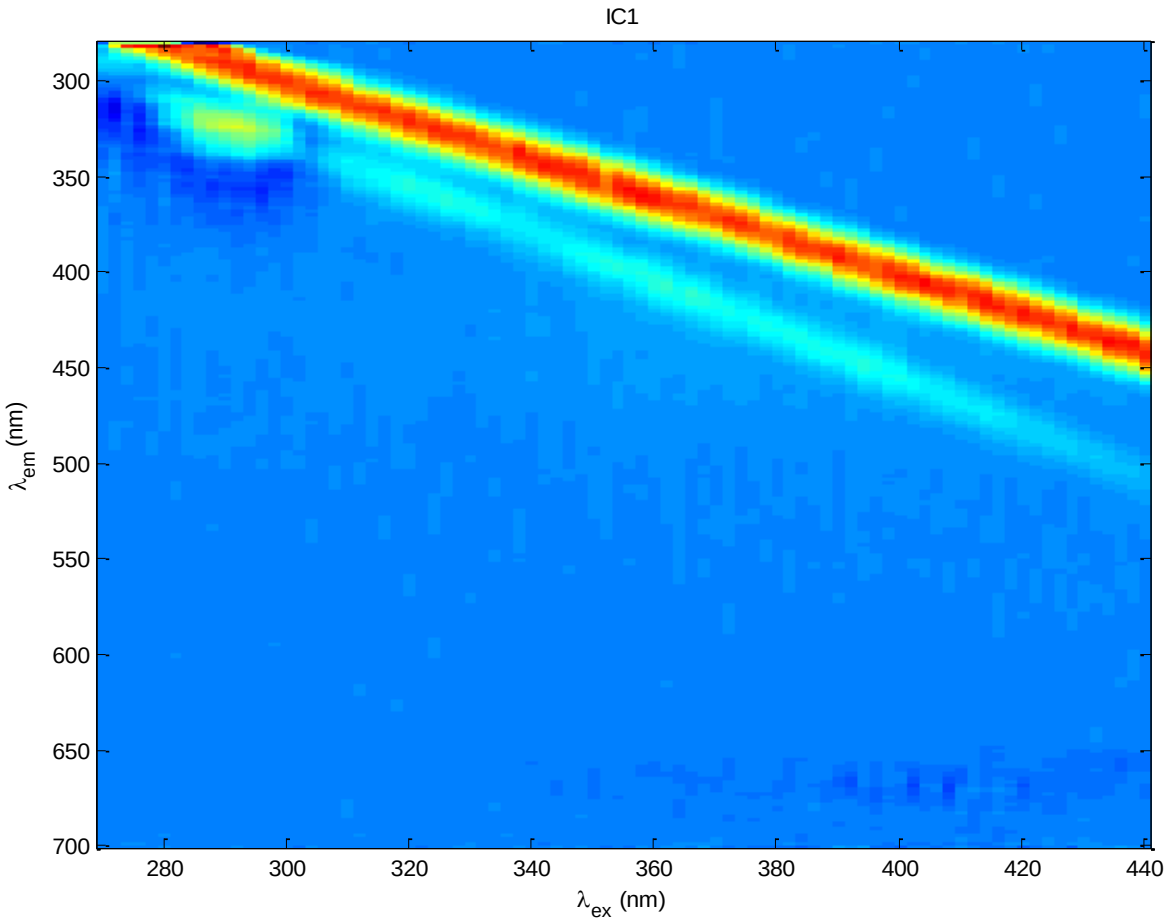
Unfolding an EEM matrix



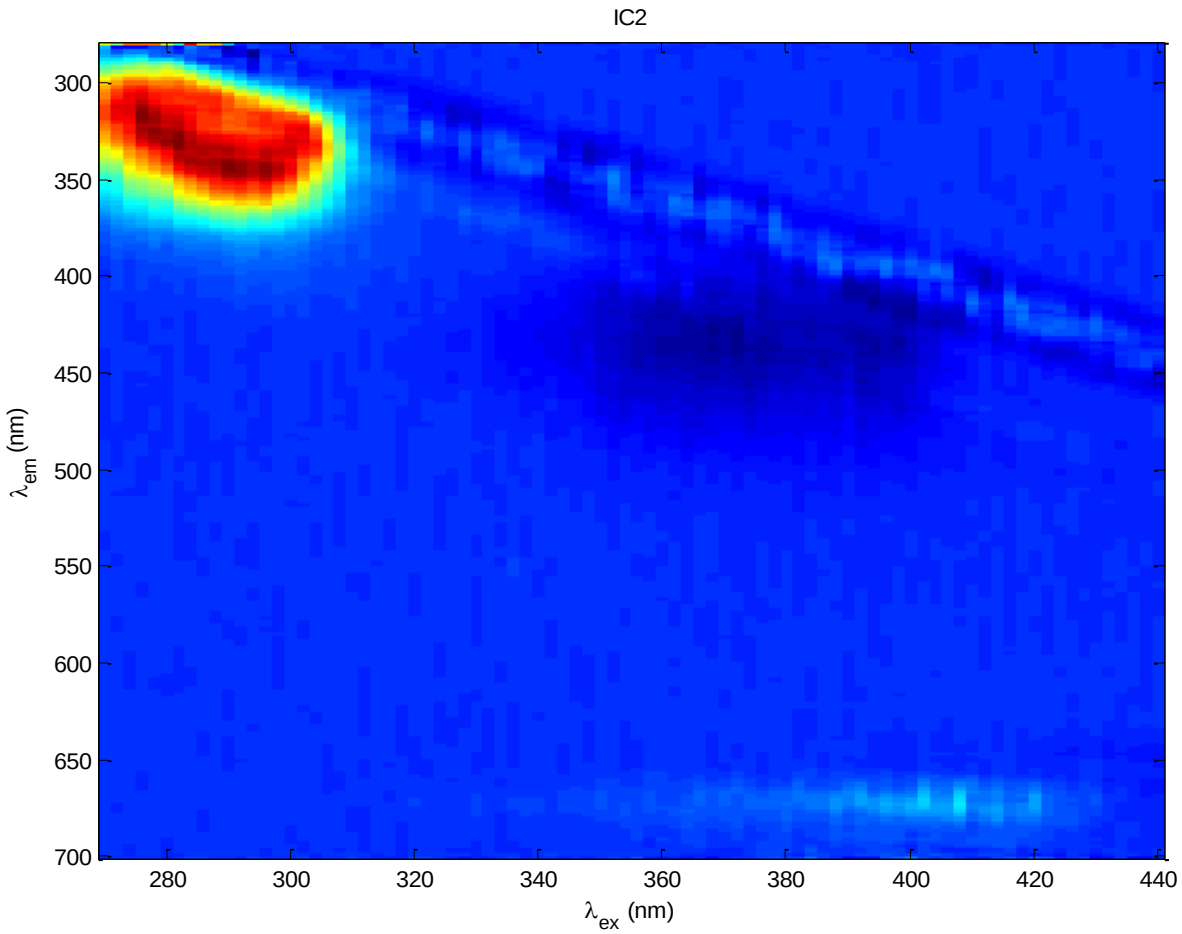
320 x 126 x 111 tensor
- 320 EEM spectra
- 111 excitation wavelengths
- 126 emission wavelengths

Each of the 320 matrices unfolded
- 320 x 13 986 matrix

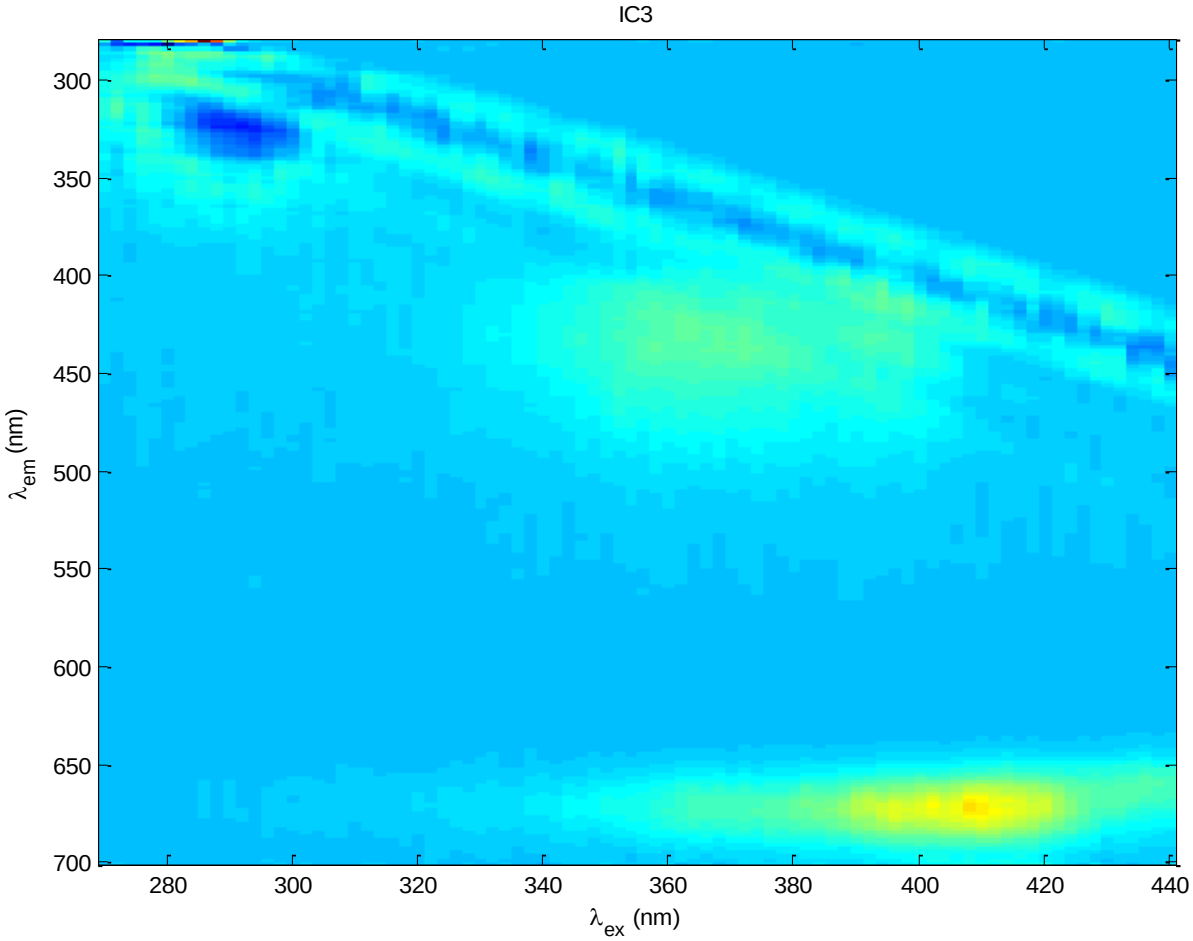




IC1 = Rayleigh + Raman



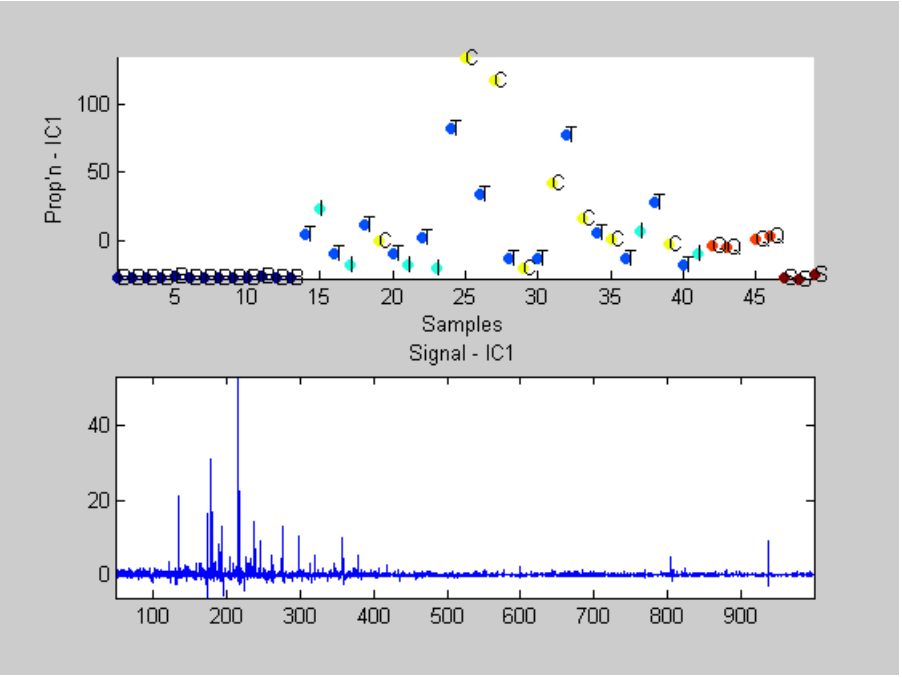
IC2 = Polyphenols

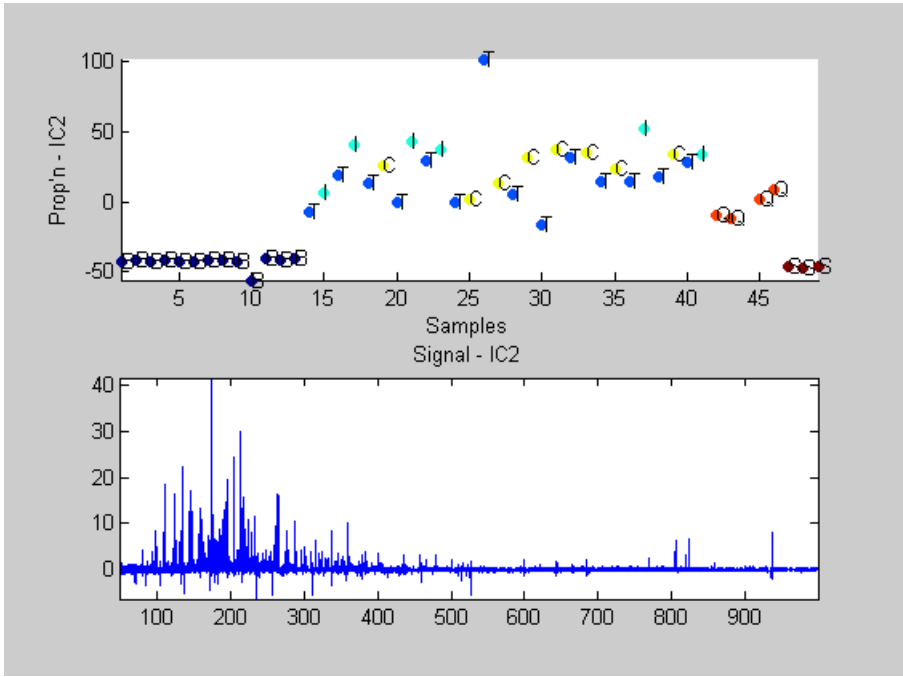


IC3 = Chlorophyll…

Non-Oriented ICA

HR Mass spectra of urine before and after exposure to pesticides





B. Habchi, S. Alves, A. Paris, D. Jouan-Rimbaud Bouveresse, Y. Lécluse, P. Gauduchon, P. Lebailly, D. N. Rutledge, E. Rathahao-Paris
Chemometric tools to highlight pesticide exposure status by analyzing metabolomic fingerprints produced by direct introduction using an FTMS instrument
Chimiométrie 2016, 17-20 Janvier 2016, Namur, Belgium

Oriented ICA

To orient extraction of ICs towards those that discriminate groups :

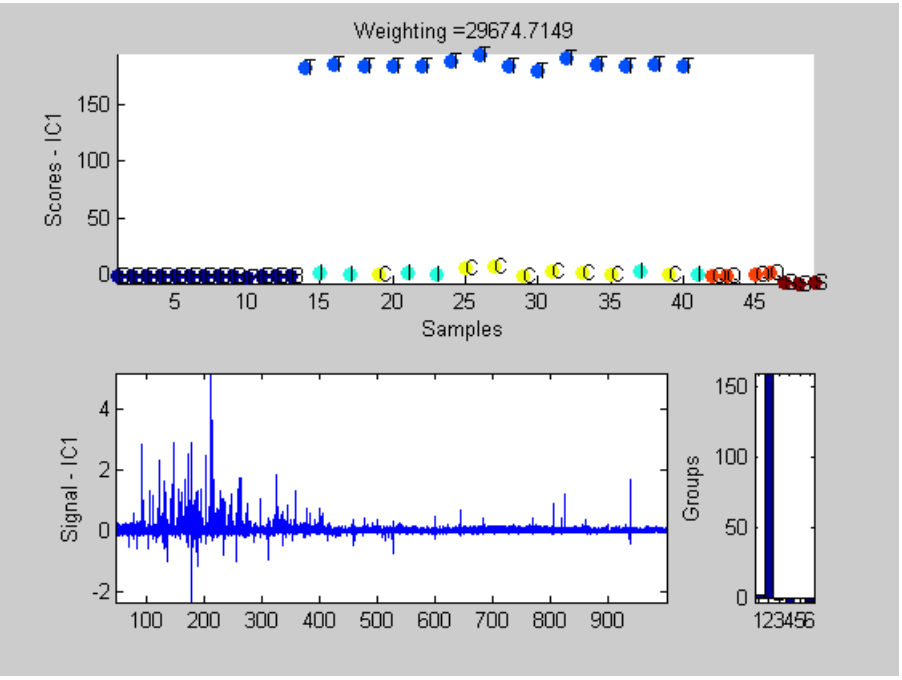
concatenate matrix of observed signals with group-membership matrix

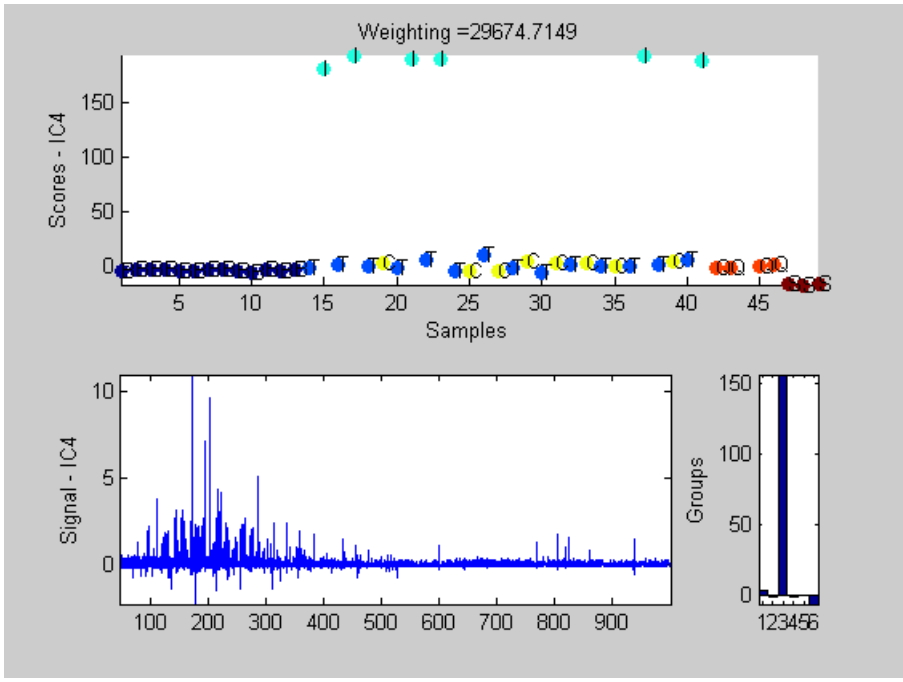
Xnew

	X	**Y**		
sample 1		1	0	0
sample 2		1	0	0
sample 3		0	1	0
sample 4		0	0	1
sample 5		1	0	0
sample 6		0	1	0
samples 7		0	0	1

Oriented ICA

HR Mass spectra of urine before and after exposure to pesticides

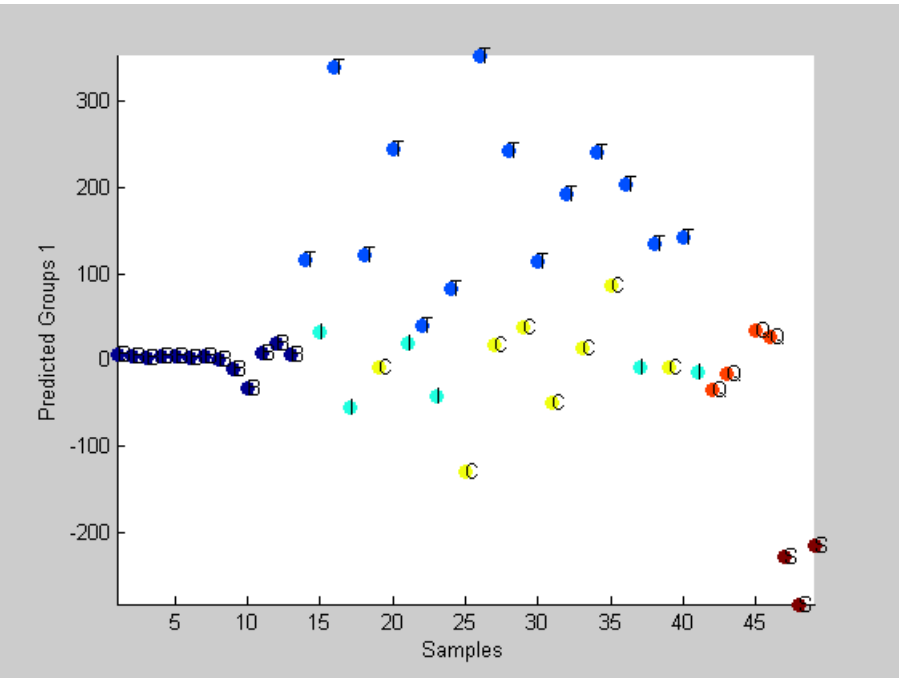


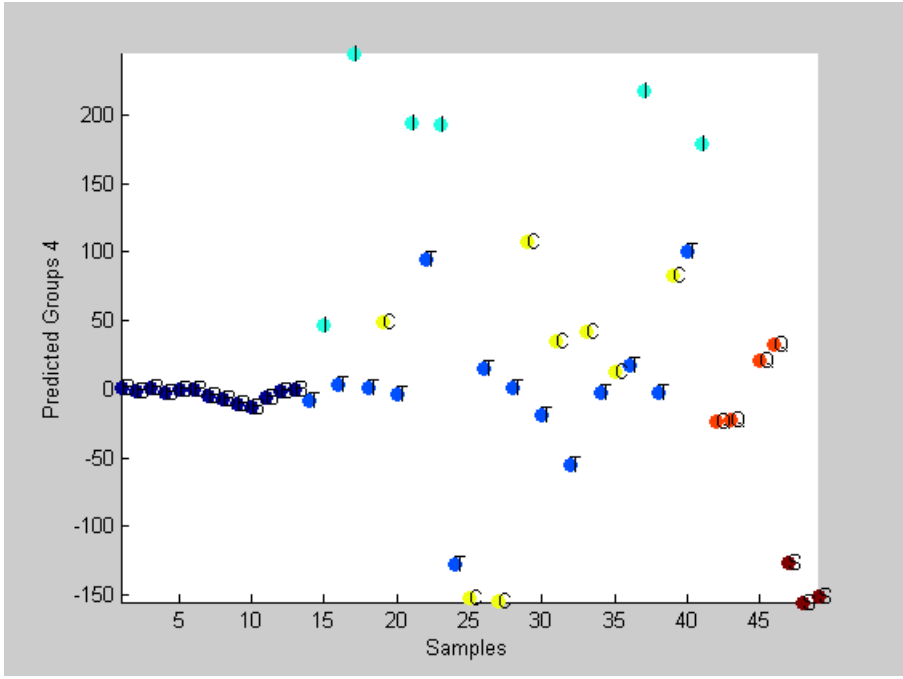


Proportions calculated using signal from MS and Groups

Oriented ICA

HR Mass spectra of urine before and after exposure to pesticides

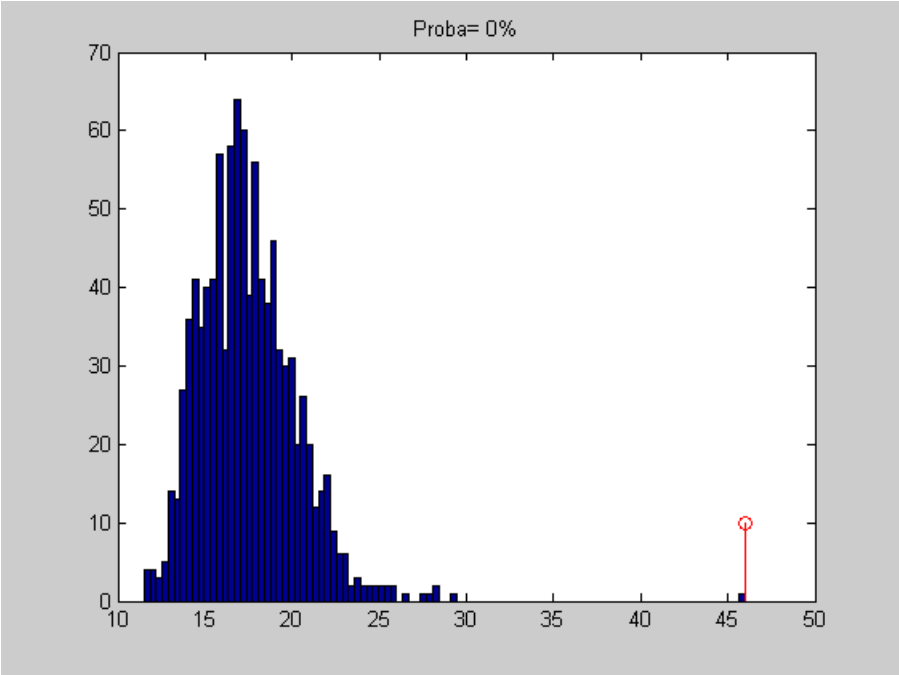




Proportions calculated using only signal from MS

Oriented ICA

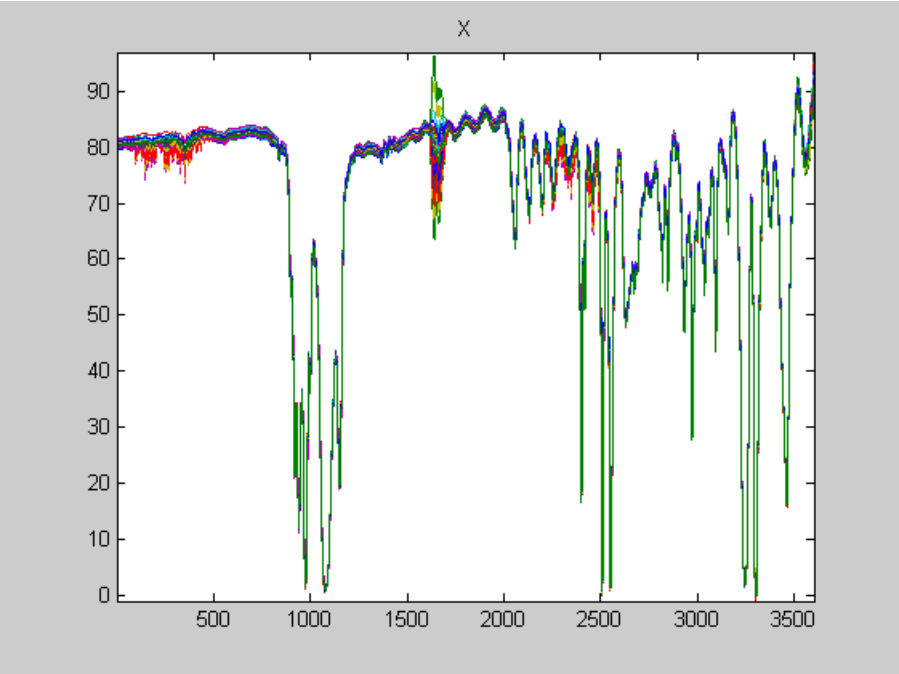
Permutation test



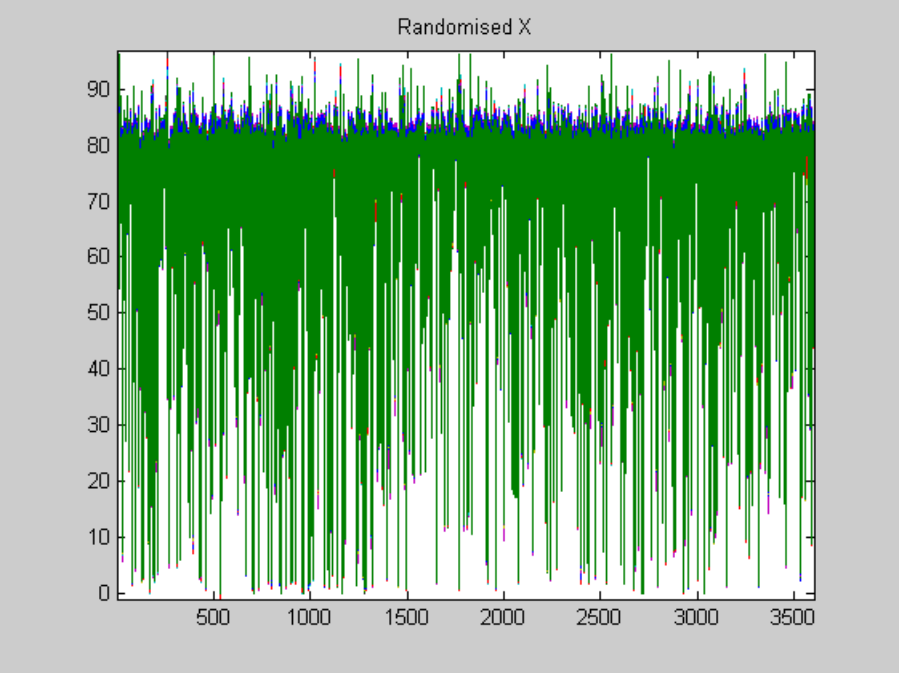
Fisher's F for Wilks' Lambda greater for "true" groups

ICA is based on *histograms* of signals

Original signals

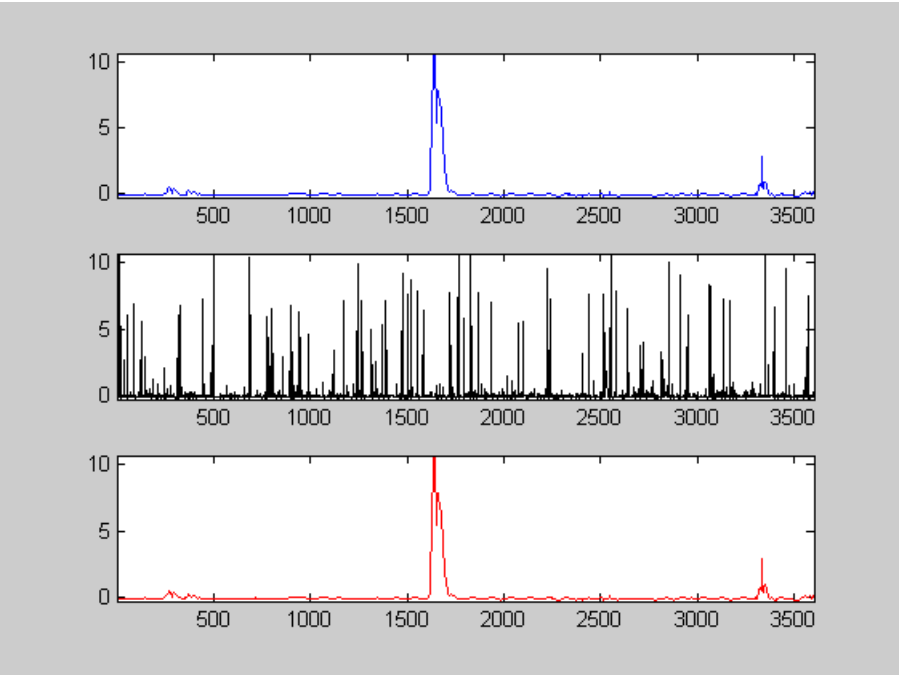


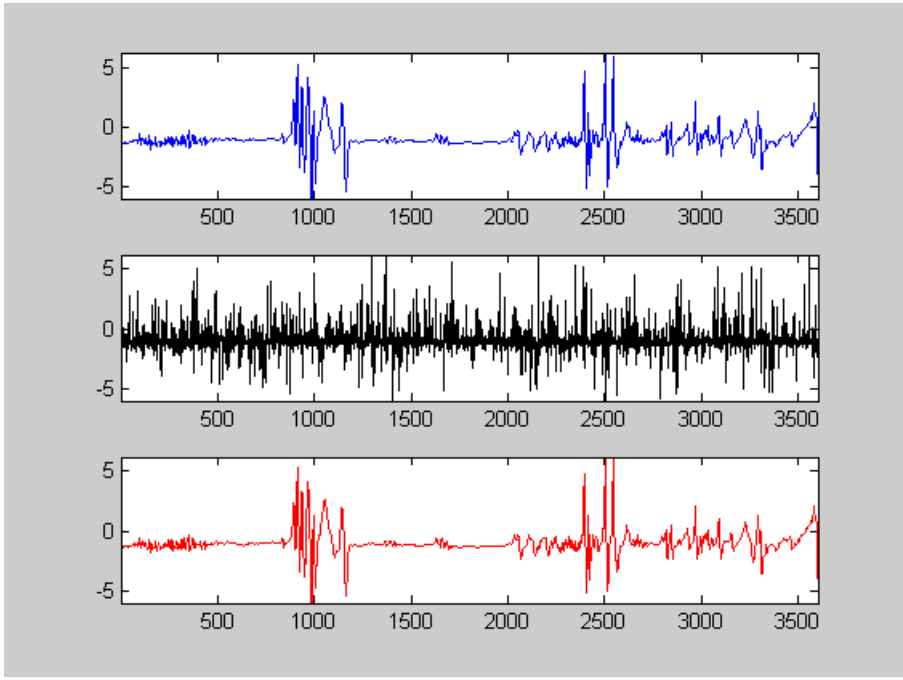
Randomised variables



ICA works on unstructured signals

Original signals

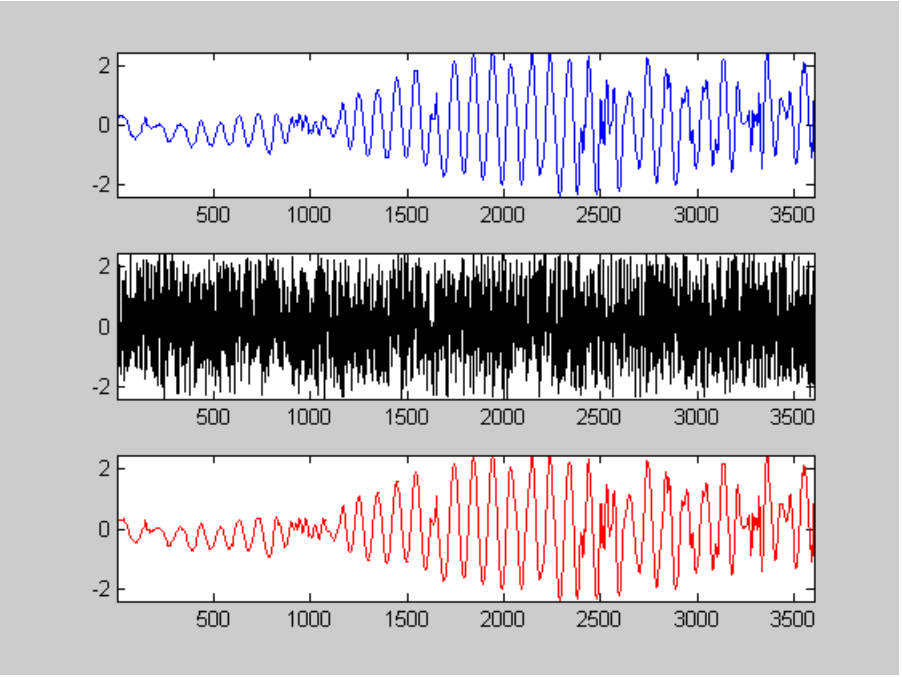


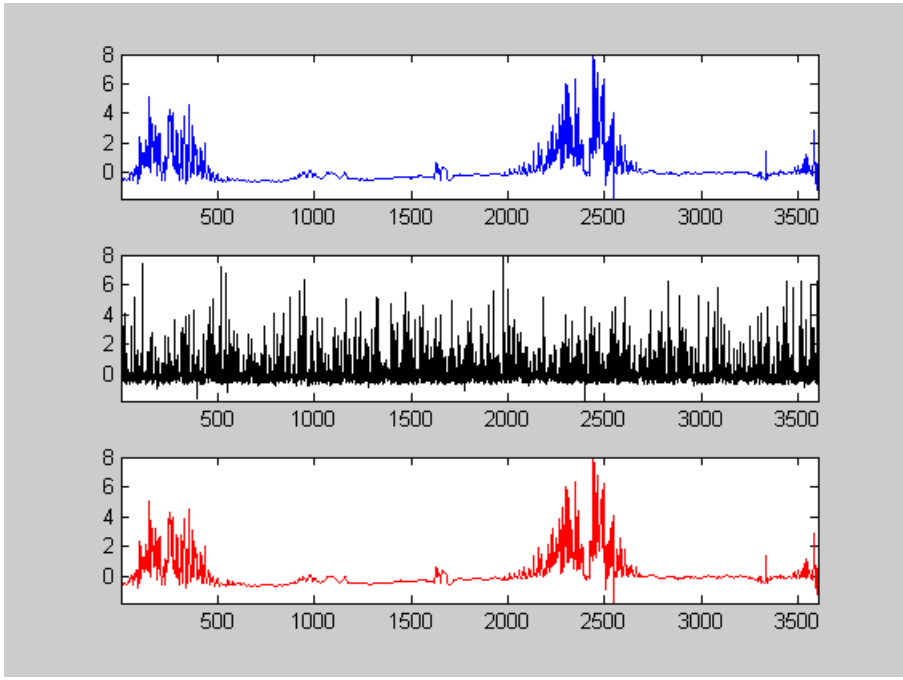


Reorderd randomised variables

ICA works on unstructured signals

Original signals

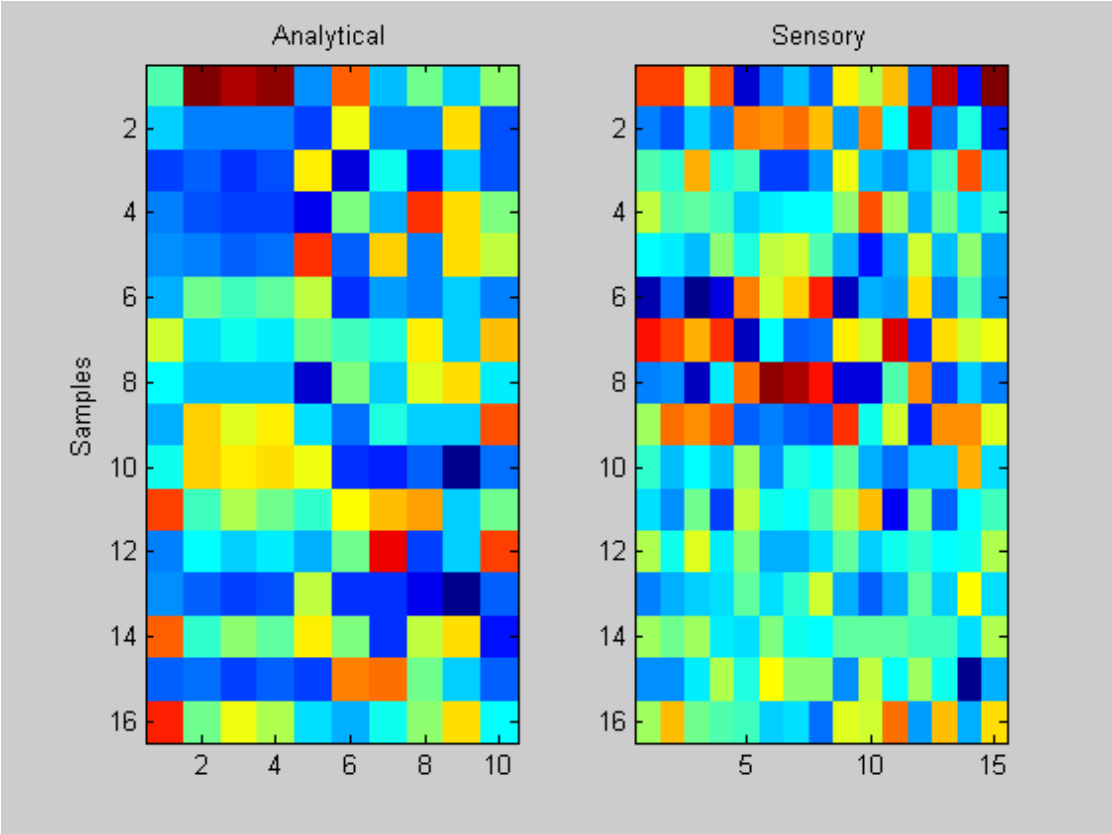




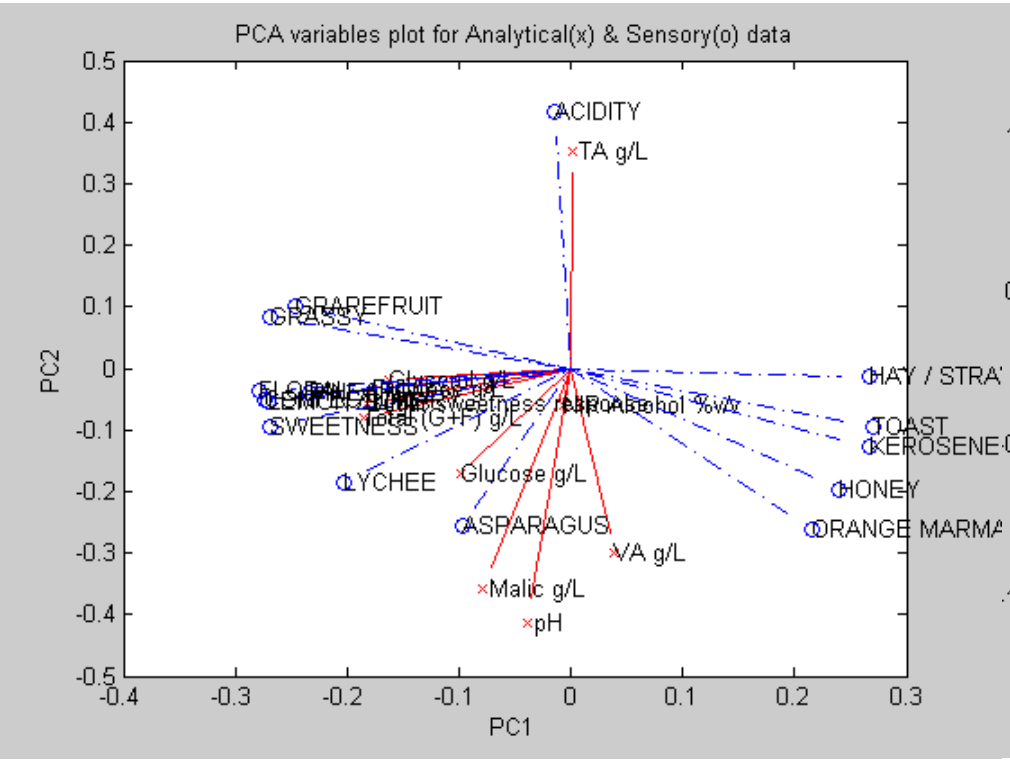
Reorderd randomised variables

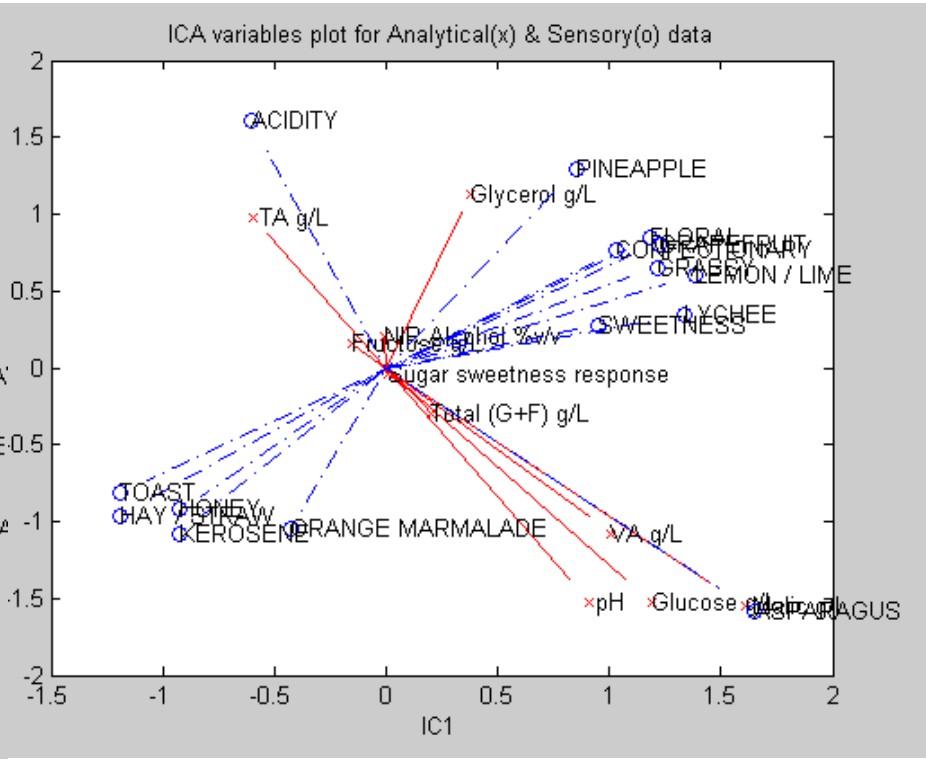
ICA applied to analytical and sensory data

Chemical and sensory analyses of 16 Australian wines



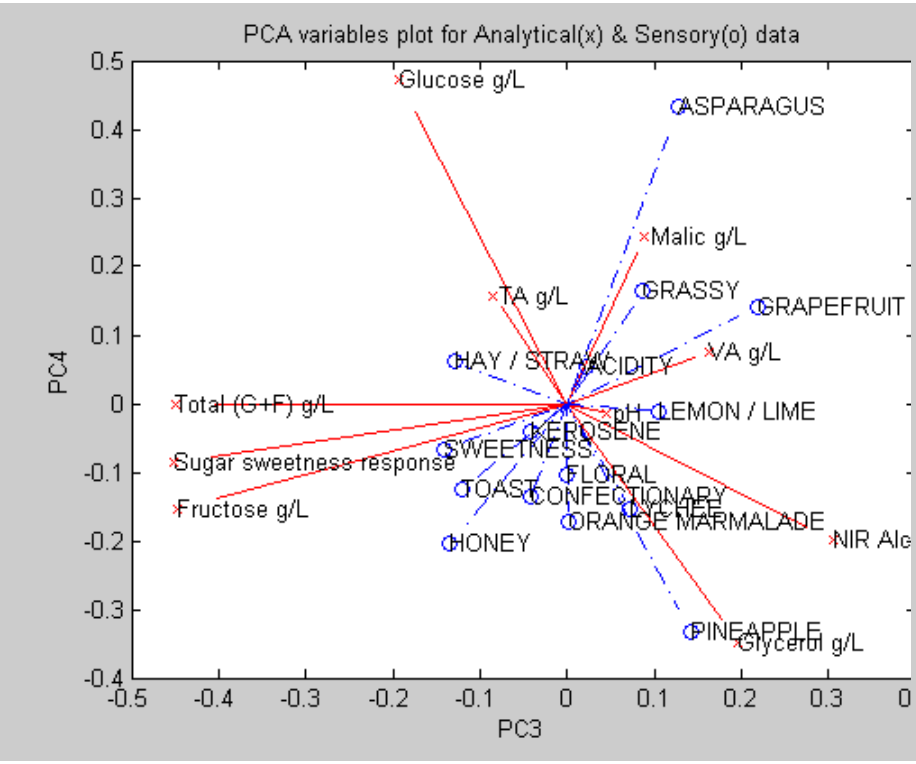
J. Blackman, D. N. Rutledge, D. Tesic, A. Saliba, G. R. Scollary
Examination of the potential for using chemical analysis as a surrogate for sensory analysis
Analytica Chimica Acta, (2010) 660, 2-7

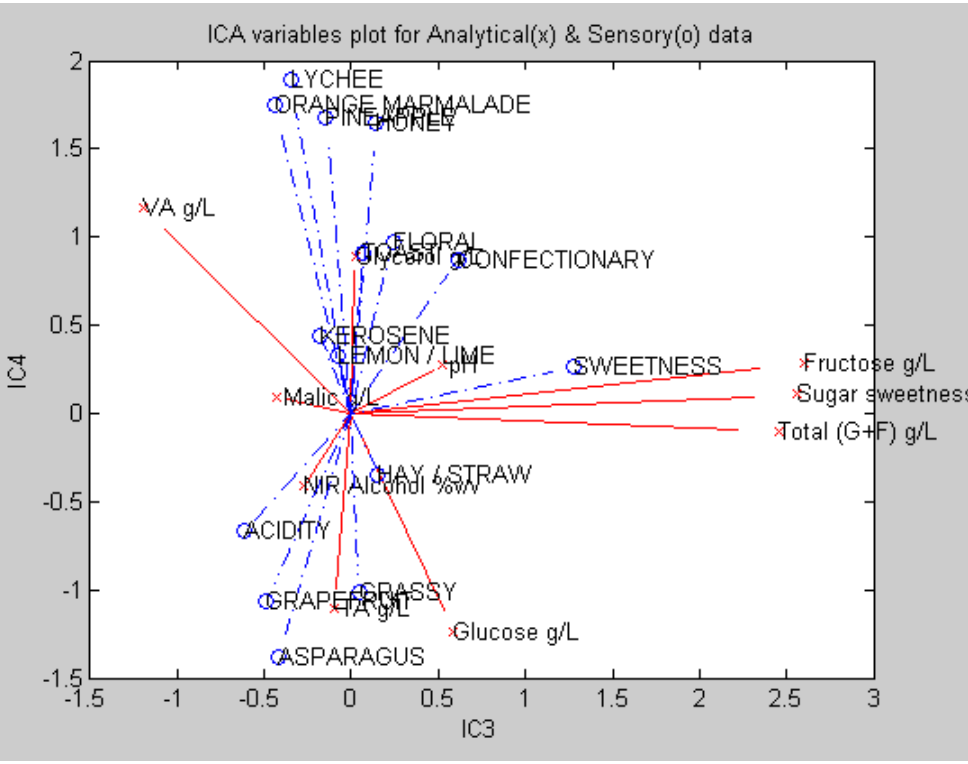




PCA Loadings

ICA Signals





PCA Loadings

ICA Signals

Advantages of ICA

- Extracted vectors (“Source signals”) have a physico-chemical meaning
- ICA can be applied to all kinds of signals, including multi-way signals
- Calculated “Proportions” are proportional !

  Therefore, useful for regression and discriminant analysis

Difficulties with ICA

- Different results may be obtained using different algorithms
- Different results may be produced as a function of the number of Independent Components (ICs) extracted
- Determination of the number of ICs

Conclusion

- PCA does not look for (*and usually does not find*) components with direct physical meaning

- ICA tries to recover the source signals by estimating a linear transformation, using a criterion that measures **statistical independence** among the sources

- This is done using higher-order information that can be extracted from the densities of the data

- ICA can be applied to **all types of data**, including multi-way data

D.N. Rutledge, D. Jouan-Rimbaud Bouveresse,
Independent Components Analysis with the JADE algorithm
Trends in Analytical Chemistry, 50, (2013) 22–32

D.N. Rutledge, D. Jouan-Rimbaud Bouveresse,
Corrigendum to “Independent Components Analysis with the JADE algorithm”
Trends in Analytical Chemistry, 67, (2015) 220