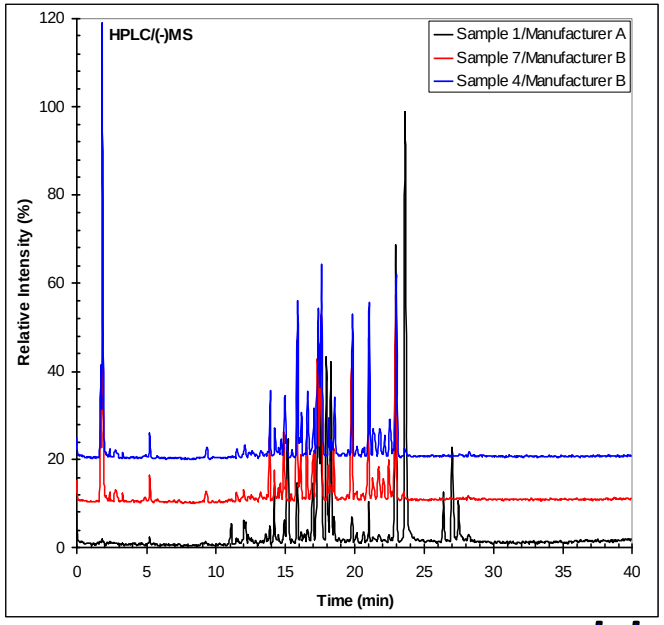
NEW CHEMOMETRIC APPROACH BASED ON LINEAR REGRESSION FOR HOLISTIC COMPARISON OF ANALYTICAL DATA

Andrei MEDVEDOVICI

University of Bucharest, Faculty of Chemistry, Department of Analytical Chemistry,
Panduri Ave. # 90-92, Bucharest, 050663, Romania

INTERNATIONAL CONFERENCE
CHIMIA 2016
NEW TRENDS IN APPLIED CHEMISTRY
May 26-28th, 2016
Constanta, ROMANIA

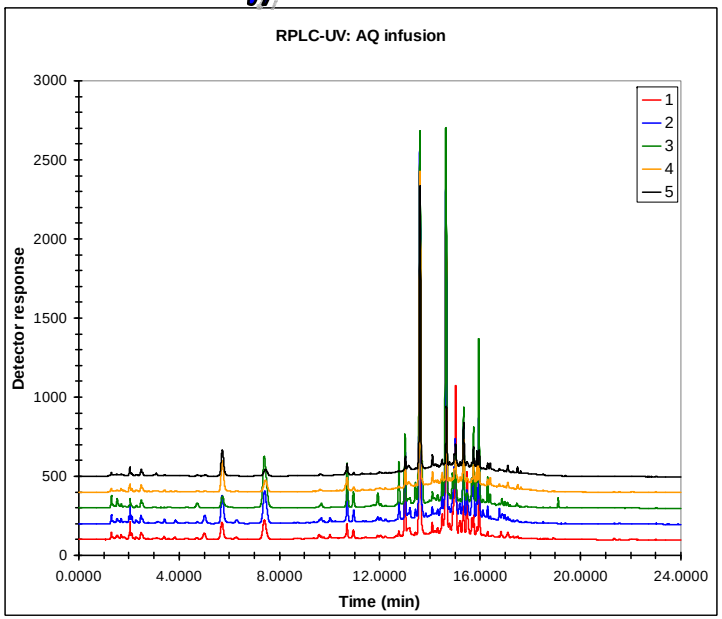


Mixtures of natural origins

COMPLEXITY

Need for Clasification!

(i.e. botanical/geographical origins, chemotype, industrial processing, authentication, batch to batch variability)



Chemometric Approaches applied to:

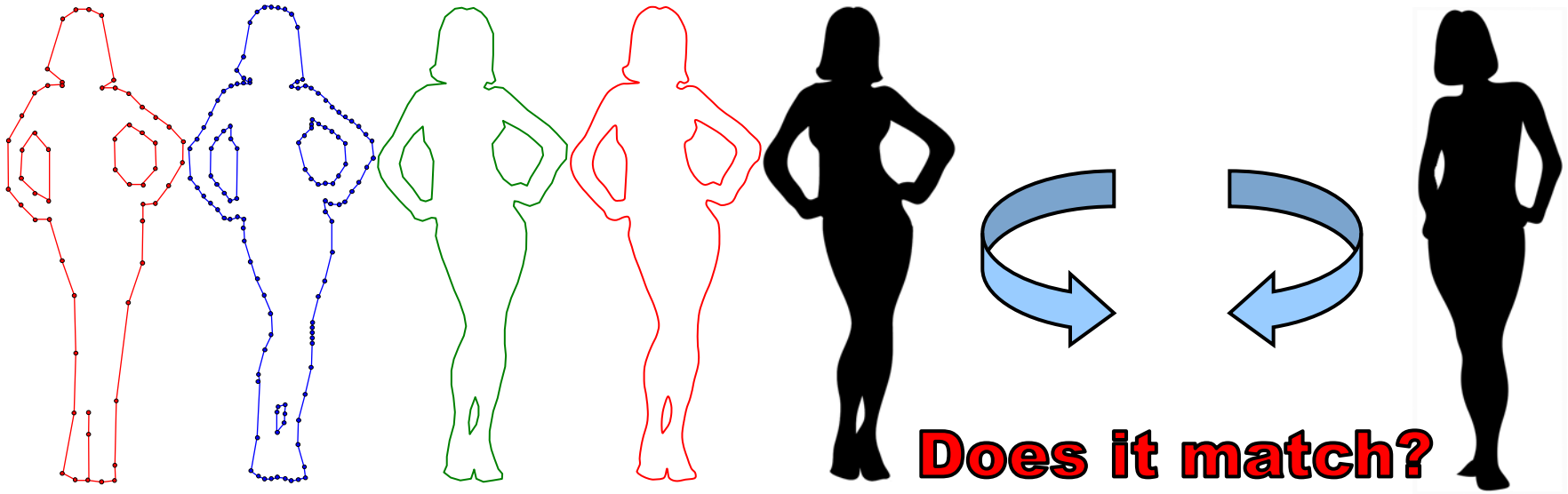
Holistic approach

UV-Vis; NIR; IR; ESR; NMR;
DART-MS; GC; LC;
Electrochemistry; Antioxidant Activity;
Chelating Capacity; Olfactory properties;
Texture properies (immaging);
High Resolution Melting Assay

Qualitative/Quantitative analytical data

Principal Component Analysis (PCA);
Cluster Analysis (CA);
Hierarchical Fuzzy Clustering (HFC);
K Nearest Neighbours (KNN);
Classification and Regression Tree (CART);
Soft Independent Modelling Cluster Analysis (SIMCA);
Partial Least Squares Discriminant Analysis (PLCDA);
Succesive Projections Algorithm (SPA);
Local Binary Patterns (LBP);
Kennard-Stone Uniform Sampling Algorithm (KS);
Supported Vector Machine (SVM)

Holistic approach - Comparison of Shapes



A better definition of the compared shapes

means

Higher Data Acquisition Rates / Higher Spectral Resolution

High definition means large data series:

#	Time	Abs. 1	Abs. 2	Abs. 3	Abs. 4	Abs. 5
1	0.000000	-0.371456	-0.399590	-0.409603	-0.414848	-0.278473
2	0.003333	-0.368118	-0.402451	-0.405312	-0.411034	-0.277519
3	0.006667	-0.366688	-0.404835	-0.401974	-0.407696	-0.276089
4	0.010000	-0.366688	-0.405312	-0.399113	-0.405312	-0.275135
5	0.013333	-0.367165	-0.403404	-0.399113	-0.403404	-0.277519
6	0.016667	-0.367165	-0.401020	-0.401974	-0.404358	-0.282288
7	0.020000	-0.367165	-0.398636	-0.405788	-0.408173	-0.286102
8	0.023333	-0.366688	-0.395775	-0.406742	-0.413418	-0.288963
9	0.026667	-0.367165	-0.391960	-0.406265	-0.418186	-0.291348
10	0.030000	-0.366688	-0.387192	-0.407696	-0.421524	-0.295162
11	0.033333	-0.366688	-0.382423	-0.409126	-0.423431	-0.298977
12	0.036667	-0.366688	-0.378609	-0.409126	-0.423431	-0.300407
13	0.040000	-0.368118	-0.376225	-0.407696	-0.422001	-0.298977
14	0.043333	-0.370979	-0.374317	-0.4	-0.420570	-0.295639
15	0.046667	-0.373840	-0.373363	-0.4	-0.421047	-0.291824
16	0.050000	-0.376225	-0.375271	-0.40960	-0.420570	-0.2
17	0.053333	-0.377178	-0.380516	-0.408649	-0.	-0.2
.....						
6888	22.956667	-3.407478	-3.398895	-3.430367	-3.381729	-2.99215
6889	22.960000	-3.385544	-3.367901	-3.440380	-3.350258	-2.960682
6890	22.963333	-3.361225	-3.334999	-3.443718	-3.312588	-2.922535
6891	22.966667	-3.334045	-3.294945	-3.438473	-3.263474	-2.874374
6892	22.970000	-3.299713	-3.238678	-3.425598	-3.194809	-2.809048
6893	22.973333	-3.251553	-3.152370	-3.406048	-3.092766	-2.713203
6894	22.976667	-3.179550	-3.017902	-3.382206	-2.935410	-2.565384
6895	22.980000	-3.067493	-2.808094	-3.354549	-2.696037	-2.342701
6896	22.983333	-2.894402	-2.492428	-3.321171	-2.343655	-2.019405
6897	22.986667	-2.635479	-2.049923	-3.281116	-1.859188	-1.570225
6898	22.990000	-2.264977	-1.473427	-3.228664	-1.240730	-0.984669
6899	22.993333	-1.765728	-0.781059	-3.150463	-0.514507	-0.283241
6900	22.996667	-1.142025	-0.023842	-3.027439	0.257492	0.480652
6901	23.000000	-0.427723	0.719070	-2.836227	0.987053	1.226902

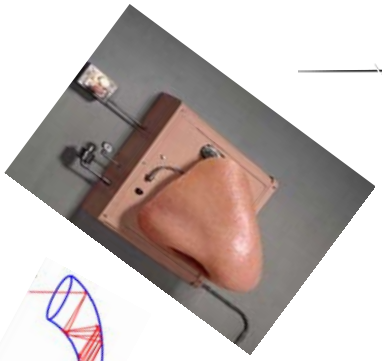
6900 data series for a 23 minutes chromatogram!

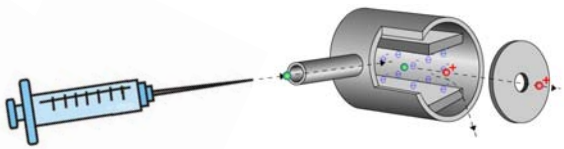
(one data every 0.2 sec) means

5 Hz acquisition frequency

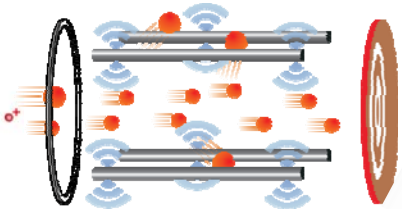
Usual PCA/CA softwares barely accept 1000 data in series.

Remember the Electronic Nose?

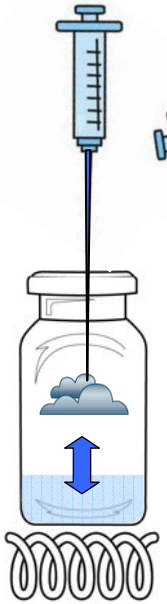




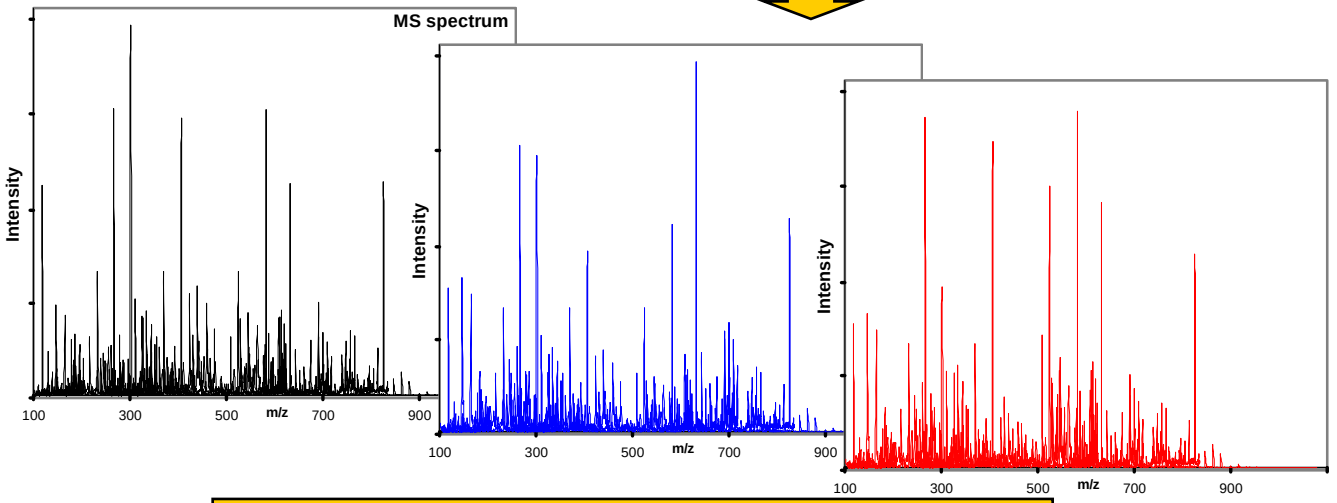
EI source



MS detector



Head-Space (HS)

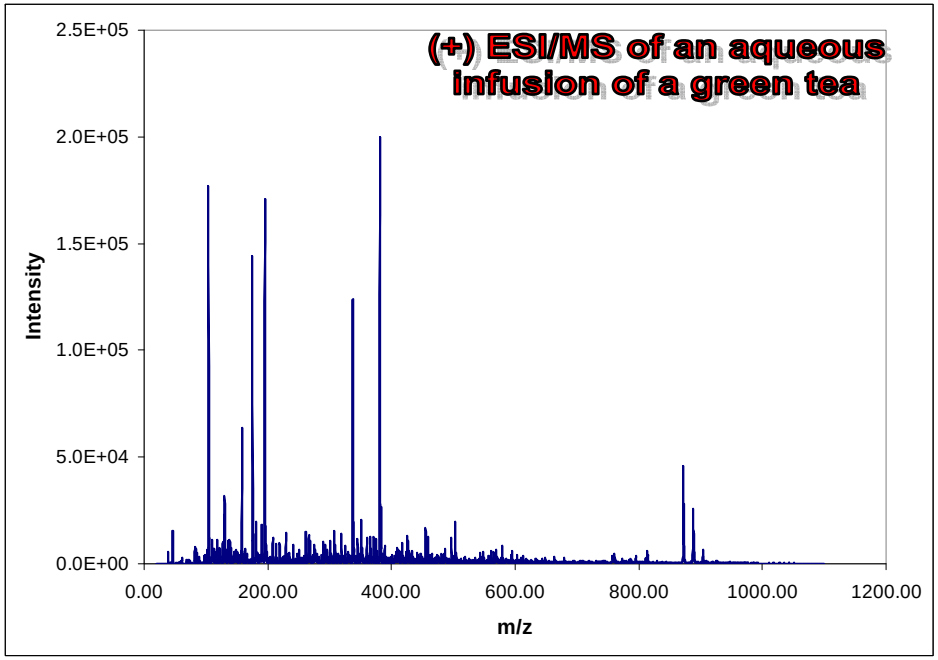


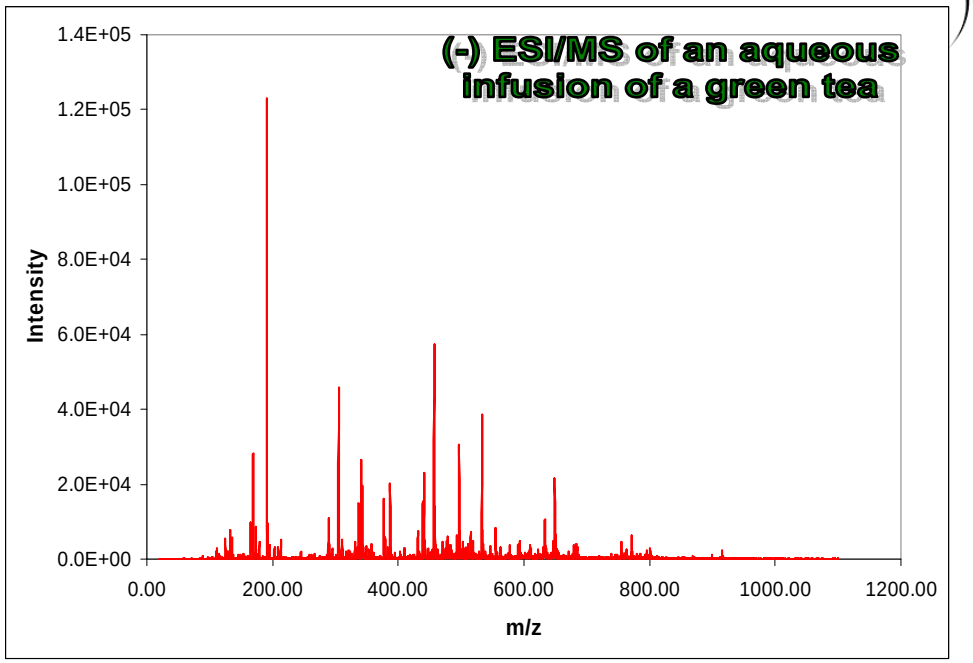
CHEMOMETRIC APPROACHES

Classification of Samples

Advantages of the AP ion sources:

In "mild" ionization sources most of the compounds produce only the corresponding pseudomolecular ions!

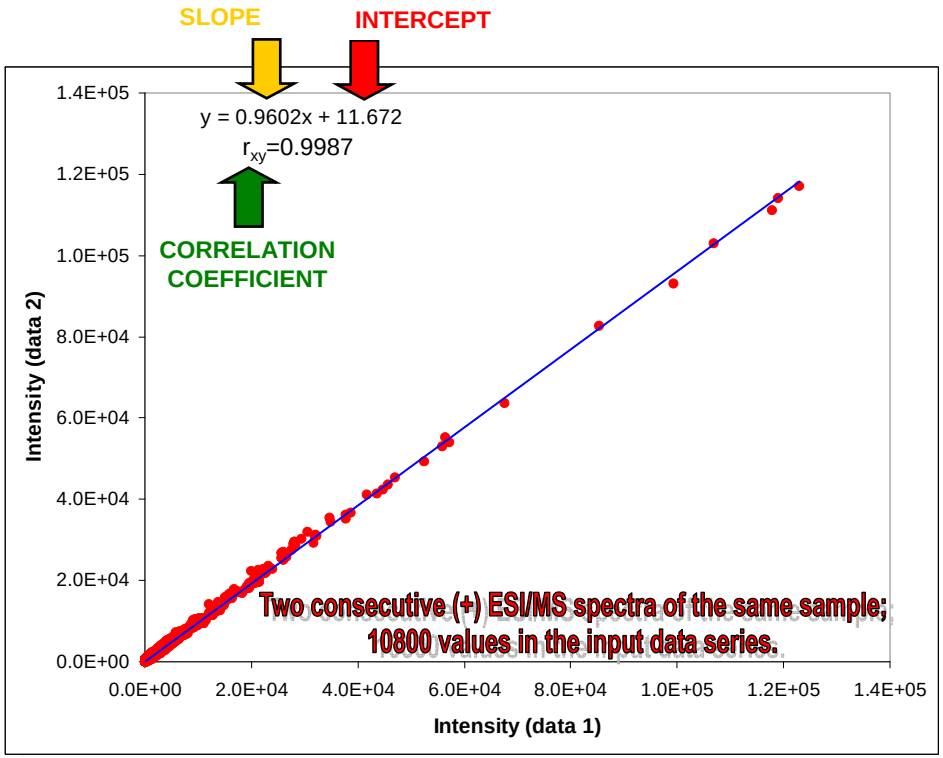


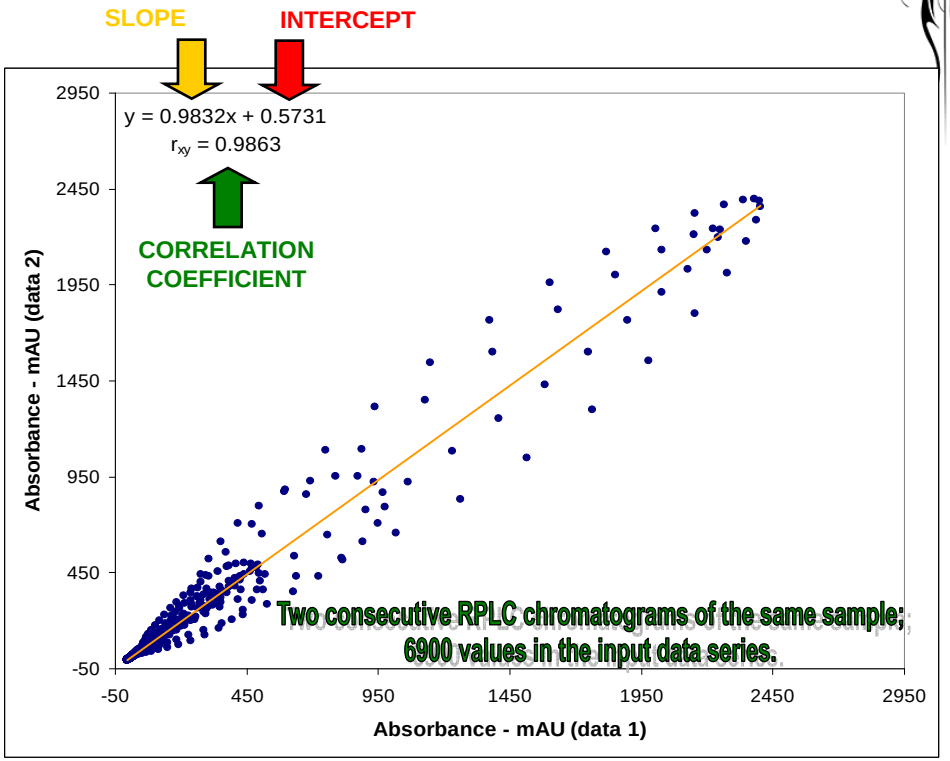


Are Direct Infusion (+/-) AP-IS/MS more informative than LC/Detection?

How does the analytical variability (sample prep, instrumental intra/inter-day) affect the discrimination power of the chemometric approaches?

Is this possible to reduce dimensionality of very large number of input raw data sets (increased definition of shapes submitted to chemometric processing)?



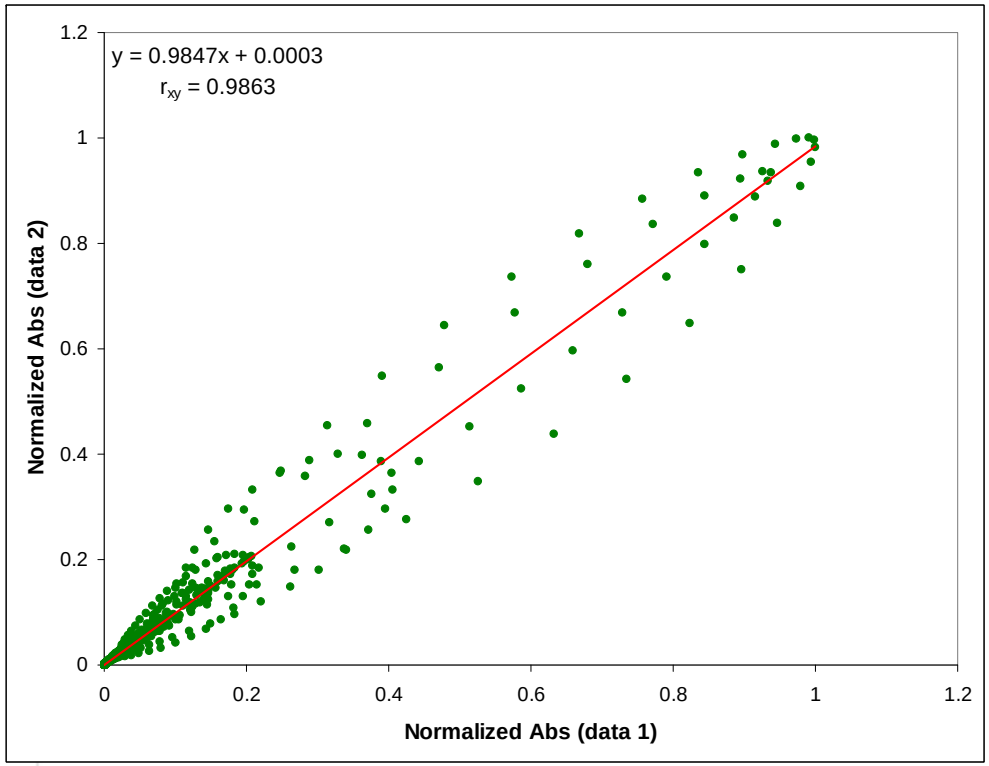


Two large sets of data may be reduced to 3 variables through expressing the Slope, Intercept and Correlation Coefficient of their reciprocal Linear Regression representation.

Perfect superposition means Slope=1; Intercept=0; Correlation Coefficient=1

Are data pre-processing tools affecting the Linear Regression approach?

i.e. Normalization



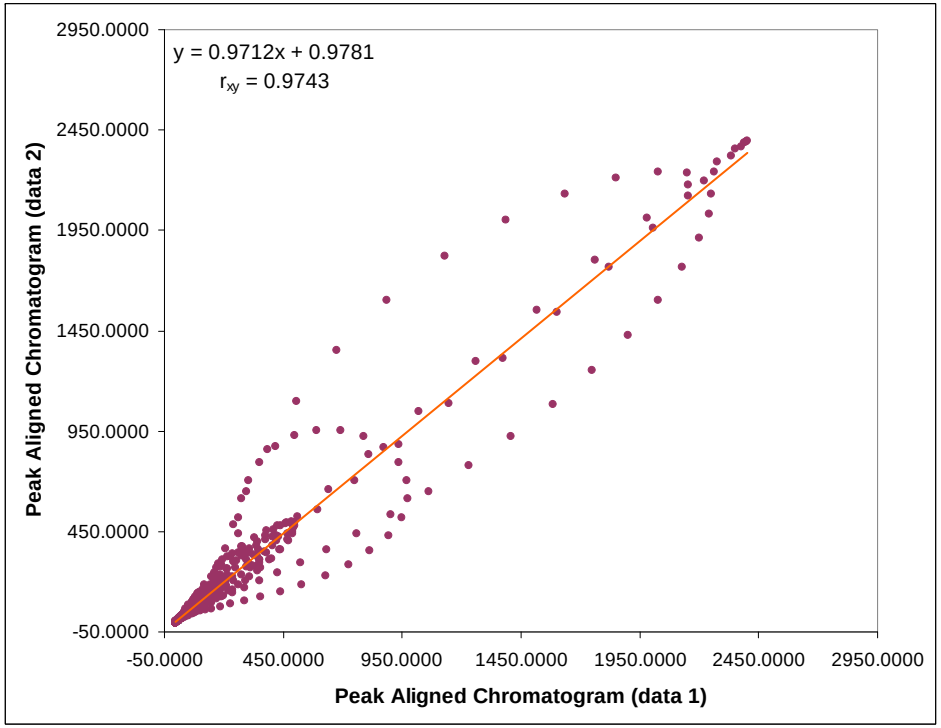
Two consecutive RPLC chromatograms of the same sample;
6900 normalized values in the input data series.

	Raw	Normalized
SLOPE	0.9832	0.9847
INTERCEPT	0.5731	0.0003
CORRELATION COEFFICIENT	0.9863	0.9863

NO effect on the Slope and Correlation Coefficient.
MAJOR effect on the Intercept.

Are data pre-processing tools affecting the Linear Regression approach?

i.e. Major Signal Alignment



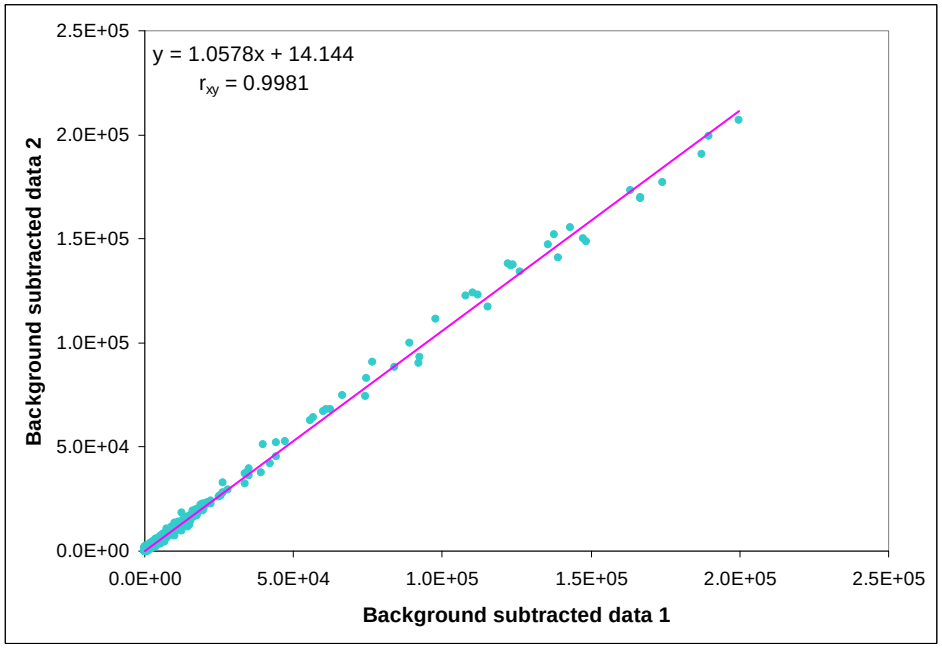
	Raw	Aligned
SLOPE	0.9832	0.9712
INTERCEPT	0.5731	0.9781
CORRELATION COEFFICIENT	0.9863	0.9743

All parameters seem negatively affected.

Two consecutive RPLC peak aligned chromatograms of the same sample; more than 6800 values in the input data series.

Are data pre-processing tools affecting the Linear Regression approach?

i.e. Background Subtraction



	Raw	Aligned
SLOPE	0.9602	1.0578
INTERCEPT	11.672	14.144
CORRELATION COEFFICIENT	0.9987	0.9981

NO effect on the Correlation Coefficient.
SIGNIFICANT effects on the Slope and the Intercept.

Two consecutive (+) ESI/MS spectra of the same sample;
10800 values after background subtraction in the input data series.

Design for an experimental application:

A. Five different brands of Green Teas.

B. Two sample preparation modes.

B1. Aqueous infusion (2hrs./ 80 oC).

B2. Ethanol extraction (RT; 7 days).

C. Three analytical procedures.

C1. RPLC/UV gradient elution.

C2. (+)ESI/MS direct infusion.

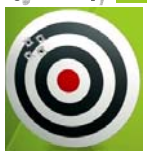
C3. (-)ESI/MS direct infusion.

D. Programming experiments.

D1. 5 X analysis of each sample (intraday var.)

D2. 5 X sample prep. processes for tea 5 (Sample prep. variability)

D3. Tea 5 samples analyzed 1 month latter (interday var.)

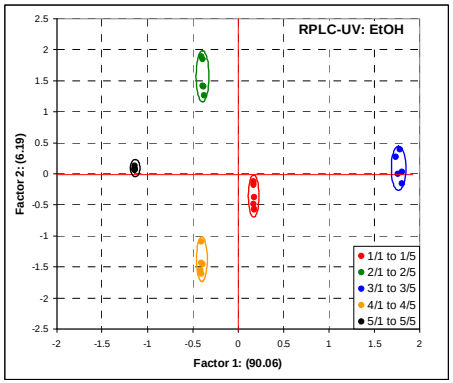


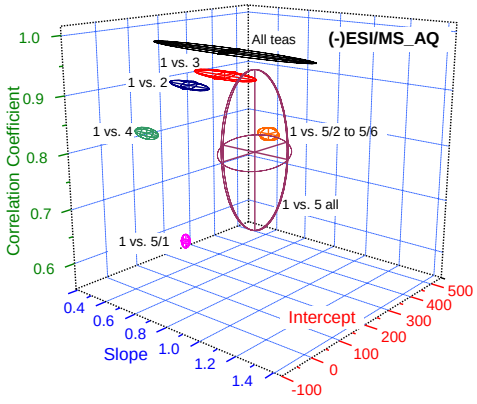
E. Chemometric approaches.

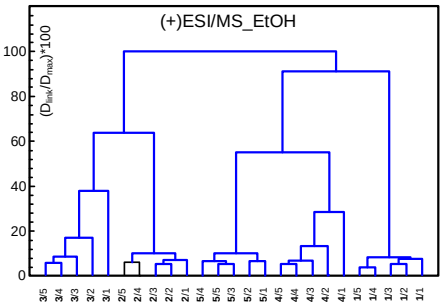
E1. Principal Component Analysis (PCA).

E2. Cluster Analysis (CA).

E3. Linear Regression based aproach (LRA).







PCA - refinement of the computed data:

A. Each 10 succesive raw input data are averaged to produce data series < 1000 values.

B. First two principal components (factors) are retained.

C. These factors retained > 50% of the whole information.

D. Each data series corresponds to one point (F1,F2 coordinates) in the space of the first two principal components.

E. For groups of data series (points) illustrating variability (5 points) one compute the Mean (M)and the Standard Deviation (s) for each of the factors.

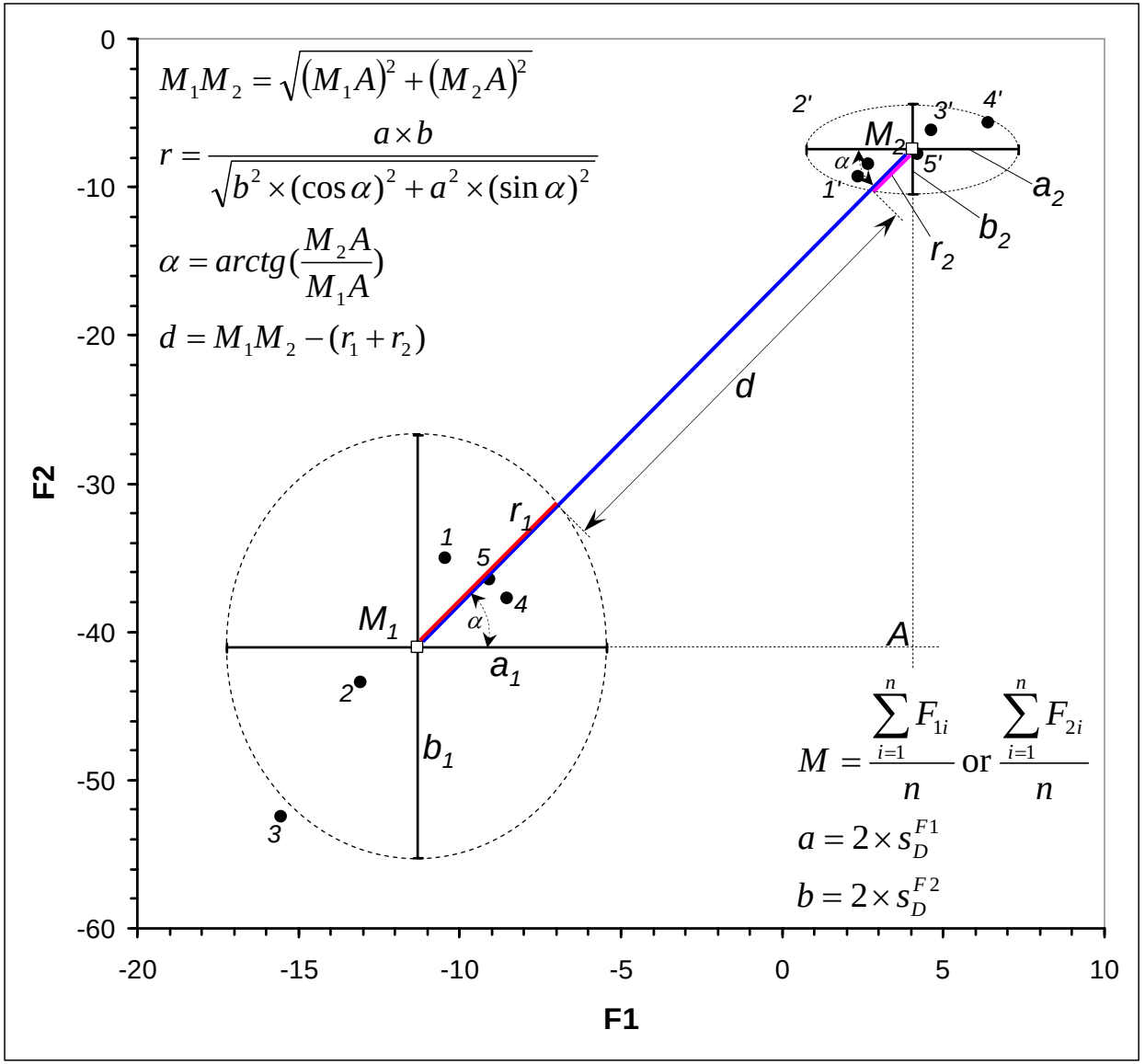
F. Each group is thus defined by an ellypse having the means as center and 2 x s as semi-diameters.

G. The true distance between two elliptic surfaces is d = D - (r1+r2) where

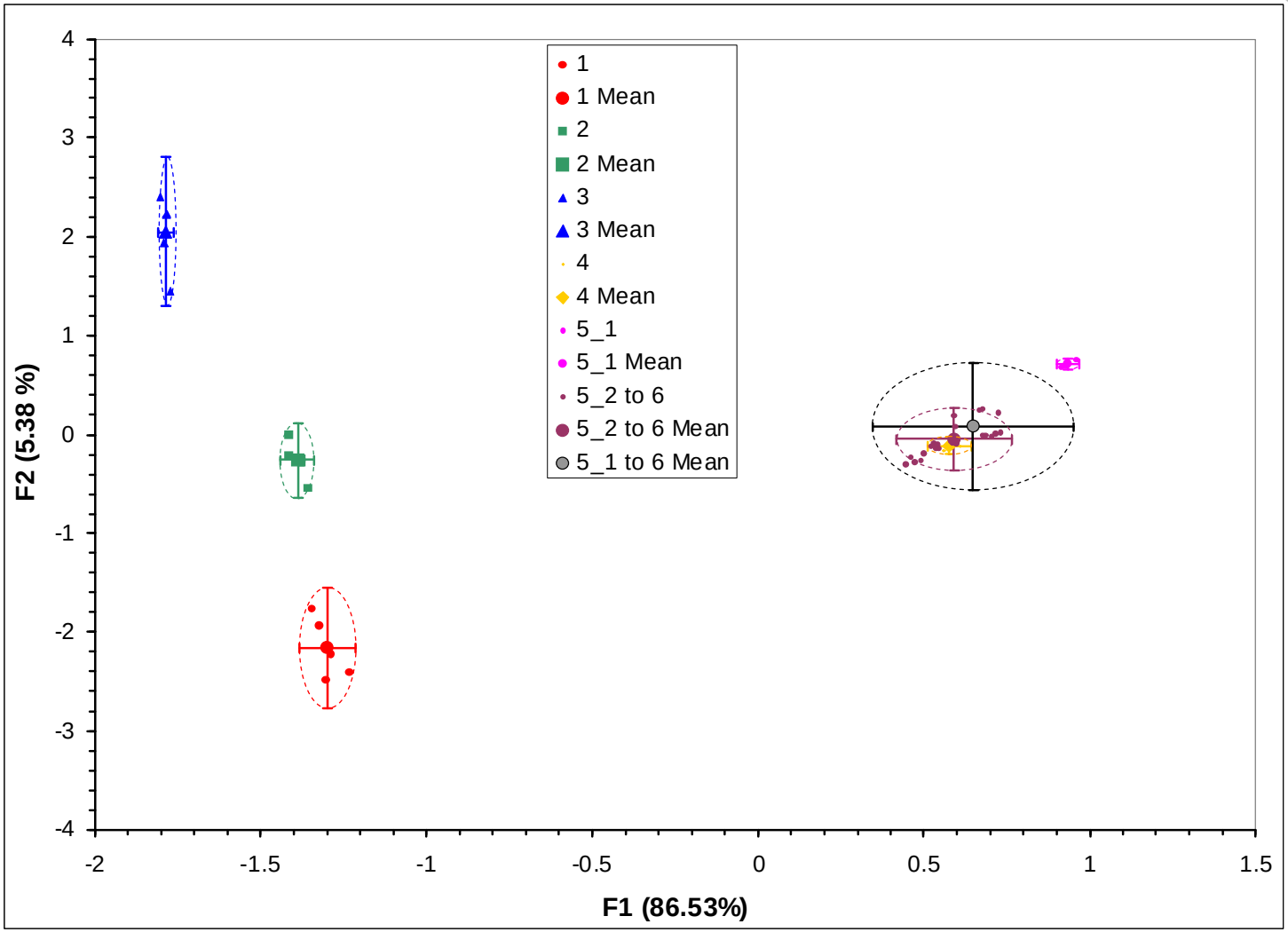
D is the distance between centers;

r1,2 are the radii of ellypses on the direction of their linkage;

Calculating the true distance between two elliptic surfaces:



PCA - RPLC/UV for aqueous infusions;



PCA results (1):

Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5-1	3 vs. 5/2 to 5/6	3 vs. 5 all	4 vs. 5-1	4 vs. 5-2 to 6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
RPLC /UV AQ	1.0	3.6	2.6	3.5	2.5	2.4	2.0	1.9	2.4	1.8	1.7	3.1	3.0	2.9	2.7	0.8	-0.3	-0.4	0.5
	S	ES	CS	ES	CS	CS	CS	CS	CS	S	S	CS	CS	CS	CS	S	NS	NS	S
RPLC /UV AQ_PA	9.0	29.9	18.3	36.9	18.2	18.9	46.2	31.8	48.1	32.1	33.1	36.4	58.6	37.9	39.4	19.6	-2.5	-1.6	-4.2
	S	CS	S	CS	S	S	CS	CS	CS	CS	CS	CS	ES	CS	CS	S	NS	NS	NS
RPLC /UV AQ_N	4.7	32.3	24.3	48.4	28.6	30.6	38.4	23.7	46.2	28.7	30.6	37.8	62.0	40.7	42.9	13.7	-4.4	-3.1	-19.6
	LS	CS	S	CS	S	S	CS	S	CS	S	S	CS	ES	CS	CS	S	NS	NS	NS
RPLC /UV EtOH	1.9	1.6	1.2	1.4	-	-	2.5	2.1	1.6	-	-	2.6	2.9	-	-	1.6	-	-	-
	CS	CS	S	S	-	-	CS	CS	CS	-	-	ES	ES	-	-	CS	-	-	-

Qualitative data interpretation:
NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.
Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S;
from 50 to 90% means CS; > 90% means ES.

PCA results (2):

Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5-1	3 vs. 5/2 to 5/6	3 vs. 5 all	4 vs. 5-1	4 vs. 5-2 to 6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
(+)ESI/MS	0.7	0.6	3.0	2.6	1.8	1.9	0.9	2.1	1.6	1.4	1.3	2.7	2.6	2.5	2.4	0.6	2.1	1.1	1.1
AQ	*S*	*S*	*ES*	*CS*	*CS*	*CS*	*S*	*CS*	*CS*	*S*	*S*	*CS*	*CS*	*CS*	*CS*	*S*	*CS*	*S*	*S*
(+)ESI/MS	11.7	13.2	29.9	30.7	38.8	18.7	-0.5	20.0	24.9	44.9	18.9	20.5	26.3	46.9	20.8	1.4	41.4	15.3	34.4
AQ_BS	*S*	*S*	*CS*	*CS*	*CS*	*S*	*NS*	*S*	*CS*	*ES*	*S*	*S*	*CS*	*ES*	*S*	*LS*	*CS*	*S*	*S*
(+)ESI/MS	12.6	25.2	22.2	27.2	29.5	16.3	10.7	18.5	32.9	43.6	29.2	19.7	39.0	53.7	38.9	-3.0	28.1	10.7	22.4
AQ_N	*S*	*S*	*S*	*CS*	*CS*	*S*	*S*	*S*	*CS*	*CS*	*CS*	*S*	*CS*	*ES*	*CS*	*NS*	*CS*	*S*	*S*
(+)ESI/MS	1.5	2.6	2.0	2.6	-	-	0.8	1.7	2.2	-	-	2.1	2.3	-	-	0.1	-	-	-
EtOH	*CS*	*ES*	*CS*	*ES*	-	-	*S*	*CS*	*CS*	-	-	*CS*	*ES*	-	-	*LS*	-	-	-

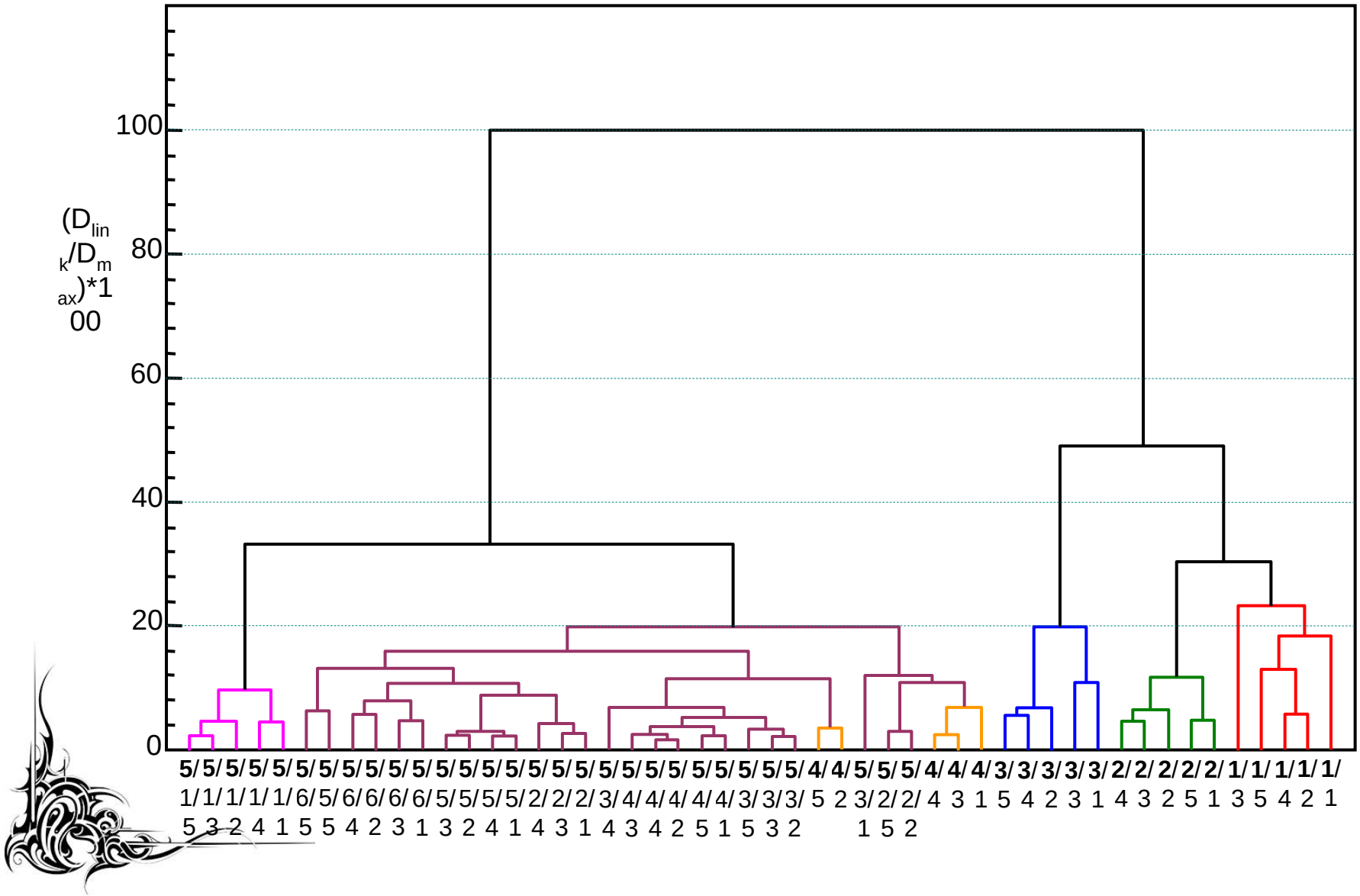
Qualitative data interpretation:
NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.
Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S;
from 50 to 90% means CS; > 90% means ES.

PCA results (3):

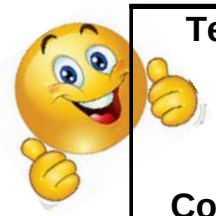
Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5-1	3 vs. 5/2 to 5/6	3 vs. 5 all	4 vs. 5-1	4 vs. 5-2 to 6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
(-)ESI/MS	0.4	-0.003	2.4	2.5	1.6	1.2	0.5	2.0	2.2	2.1	1.5	2.8	2.9	2.2	1.8	0.4	2.4	0.6	2.1
AQ	S	NS	CS	CS	CS	S	S	CS	CS	CS	CS	ES	ES	CS	CS	S	CS	S	CS
(-)ESI/MS	6.1	3.6	36.1	34.2	30.9	11.7	5.1	26.5	26.9	35.3	10.8	36.2	37.2	41.5	20.5	2.0	41.8	15.8	36.0
AQ_BS	S	LS	CS	CS	CS	S	S	CS	CS	CS	S	CS	CS	ES	S	LS	ES	S	CS
(-)ESI/MS	8.7	6.0	29.6	33.2	39.7	22.6	6.2	30.9	35.9	49.1	31.1	36.3	40.7	50.1	32.7	-3.2	29.7	9.1	28.7
AQ_N	S	S	CS	CS	CS	S	S	CS	CS	ES	CS	CS	CS	ES	CS	NS	CS	S	CS
(-)ESI/MS	2.6	2.2	2.1	2.5	-	-	0.7	1.9	2.0	-	-	2.0	2.3	-	-	0.3	-	-	-
EtOH	ES	CS	CS	ES	-	-	S	CS	CS	-	-	CS	CS	-	-	S	-	-	-

Qualitative data interpretation:
NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.
Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S;
from 50 to 90% means CS; > 90% means ES.

CA - RPLC/UV for aqueous infusions;



CA results:



Technique / Comparison	RPLC/UV_AQ	RPLC/UV_AQ_N	RPLC/UV_AQ_PA	RPLC/UV_EtOH	(+)ESI/MS_AQ	(+)ESI/MS_AQ_N	(+)ESI/MS_AQ_BS	(+)ESI/MS_EtOH	(-)ESI/MS_AQ	(-)ESI/MS_AQ_N	(-)ESI/MS_AQ_BS	(-)ESI/MS_EtOH
1 vs. 2	S	S	NS	S	ES	CS	CS	ES	S	S	S	S
1 vs. 3	S	S	S	ES	CS	CS	S	ES	S	S	S	S
1 vs. 4	ES	ES	ES	S	ES	ES	ES	ES	ES	CS	ES	ES
1 vs. 5/1	ES	ES	ES	S	ES	ES	ES	ES	ES	CS	ES	ES
1 vs. 5/2-5/6	ES	ES	ES	-	ES	ES	ES	-	ES	ES	ES	-
1 vs. 5 all	ES	ES	ES	-	ES	ES	ES	-	ES	CS	ES	-
2 vs. 3	S	S	S	ES	ES	CS	CS	CS	S	S	S	S
2 vs. 4	ES	ES	ES	S	CS	ES	ES	ES	ES	CS	ES	ES
2 vs. 5/1	ES	ES	ES	S	CS	ES	ES	ES	ES	CS	ES	ES
2 vs. 5/2-5/6	ES	ES	ES	-	CS	ES	ES	-	ES	ES	ES	-
2 vs. 5 all	ES	ES	ES	-	CS	ES	ES	-	ES	CS	ES	-
3 vs. 4	ES	ES	ES	ES	ES	ES	ES	ES	ES	CS	ES	ES
3 vs. 5/1	ES	ES	ES	ES	ES	ES	ES	ES	ES	CS	ES	ES
3 vs. 5/2-5/6	ES	ES	ES	-	ES	ES	ES	-	ES	ES	ES	-
3 vs. 5 all	ES	ES	ES	-	ES	ES	ES	-	ES	CS	ES	-
4 vs. 5/1	S	S	S	S	S	S	S	CS	S	CS	S	S
4 vs. 5/2-5/6	NS	NS	S	-	CS	CS	CS	-	CS	ES	CS	-
4 vs. 5 all	NS	NS	S	-	S	S	S	-	S	CS	S	-
5/1 vs. 5/2-5/6	S	S	S	-	CS	CS	CS	-	CS	ES	CS	-
5/2 vs. 5/3	NS	NS	S	-	S	S	NS	-	S	S	S	-
5/3 vs. 5/4	NS	NS	NS	-	NS	NS	NS	-	S	S	S	-
5/4 vs. 5/5	S	S	S	-	S	S	NS	-	S	S	S	-
5/5 vs. 5/6	NS	NS	NS	-	S	S	S	-	LS	S	LS	-

Qualitative data interpretation was based on the nearest linkage between compared clusters at a given level of $(D_{link}/D_{max})*100$: NS - not separated; LS - low separation < 10%; S - separation $\in$ [10-50)%; CS - clearly separated $\in$ [50-90]%; ES - extensively separated >90%.

LRA - refinement of the computed data:

A. No reduction of the raw input data series was achieved .

B. Each comparison between two series is characterized by the slope (S), the intercept (O) and the correlation coefficient (C).

C. Permutations of n items taken as 2 at a time produce comparison within repeated samples (20).

D. One calculate the means and standard deviations of S, O and C for a family of repeated samples .

E. Comparison between types of repeated samples was also made (25 LR for 5X repeated samples).

F. One calculate the means and standard deviations of S, O and C for a group of compared samples .

G. Graphic representation of the results in the Cartesian space (S on Ox, O on Oy and C on Oz).

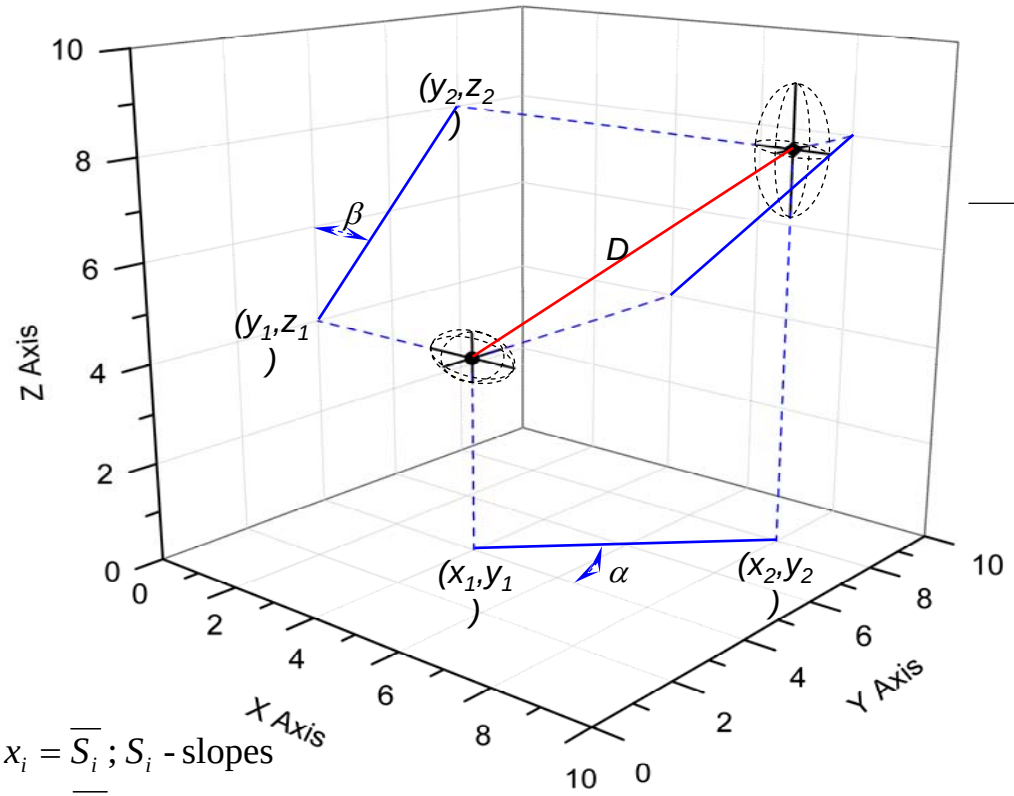
H. Each family or comparison between types of samples are characterized by an ellipsoid in the 3D space. having the means as center and the 2xs intervals as semi-diameters.

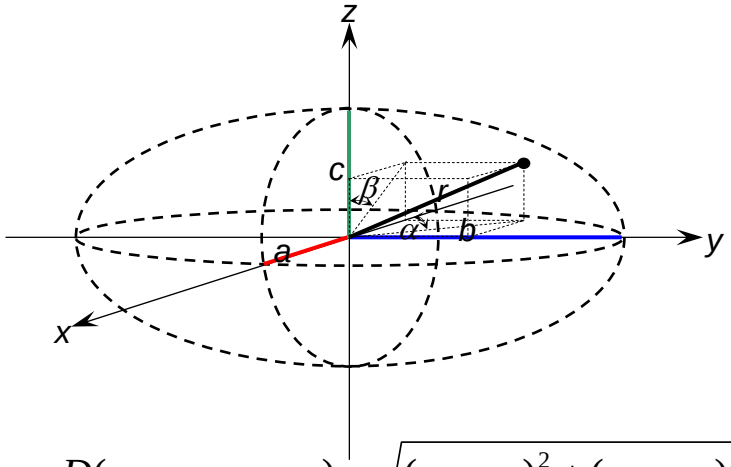
The distance between two ellipsoids is d = D - (r1 + r2) where:

D is the distance measured between the two centers;

r1,2 are the radii of ellypsoids on the direction of their center's linkage.

Calculating the true distance between two ellipsoidal volumes:





$$D(x_1y_1z_1, x_2y_2z_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

$$\alpha = arctg\left(\frac{y_2 - y_1}{x_2 - x_1}\right)$$

$$\beta = arctg\left(\frac{y_2 - y_1}{z_2 - z_1}\right)$$

$x_i = \overline{S_i}$; S_i - slopes

$y_i = \overline{O_i}$; O_i - interecept

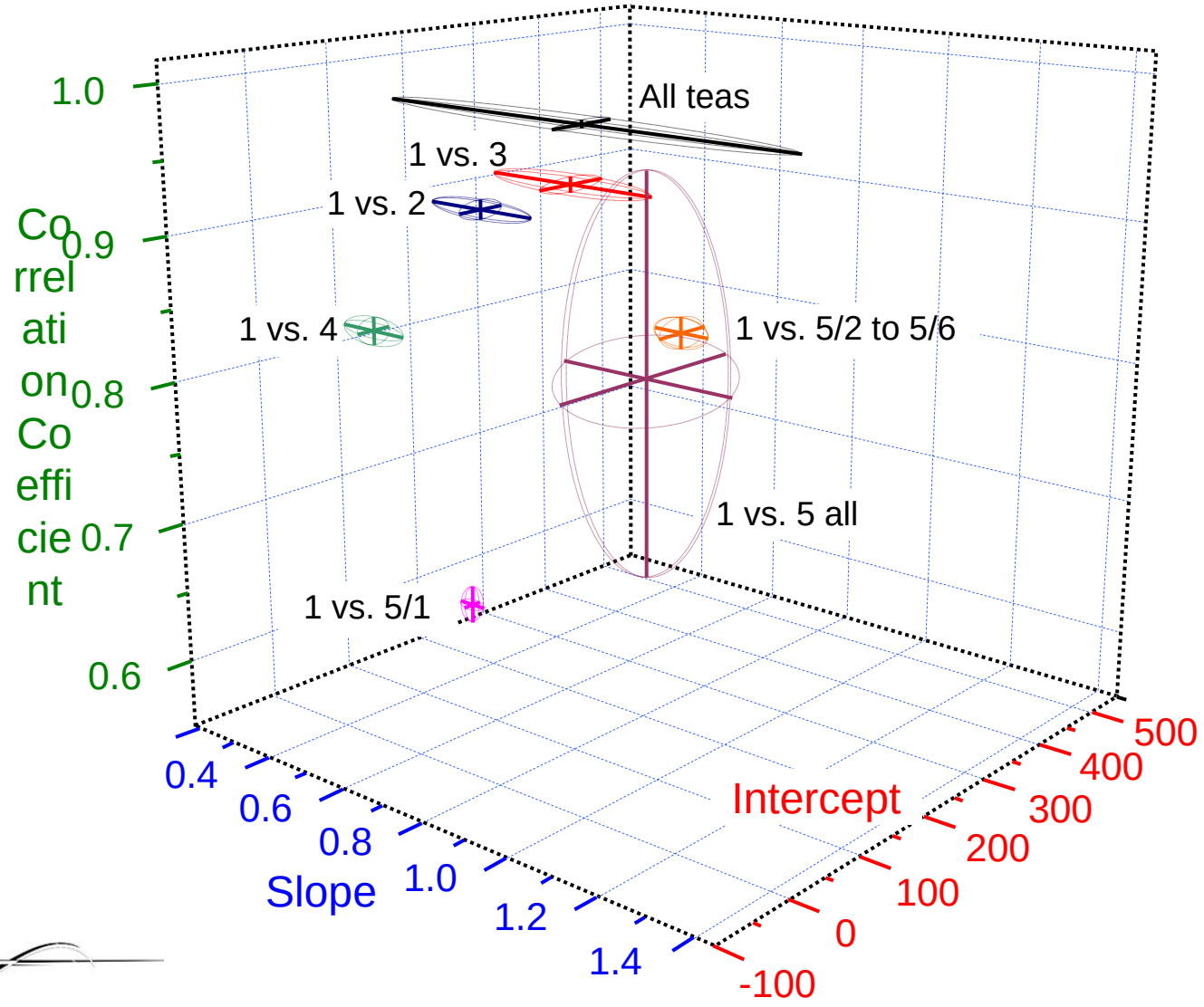
$z_i = \overline{C_i}$; C_i - correlation coefficient

$a = 2 \times s_S$; $b = 2 \times s_O$; $c = 2 \times s_C$; s - standard deviation

$$r = \frac{a \times b \times c}{\sqrt{b^2 \times c^2 \times \cos^2\alpha \times \sin^2\beta + a^2 \times c^2 \sin^2\alpha \times \sin^2\beta + a^2 \times b^2 \times \cos^2\beta}}$$

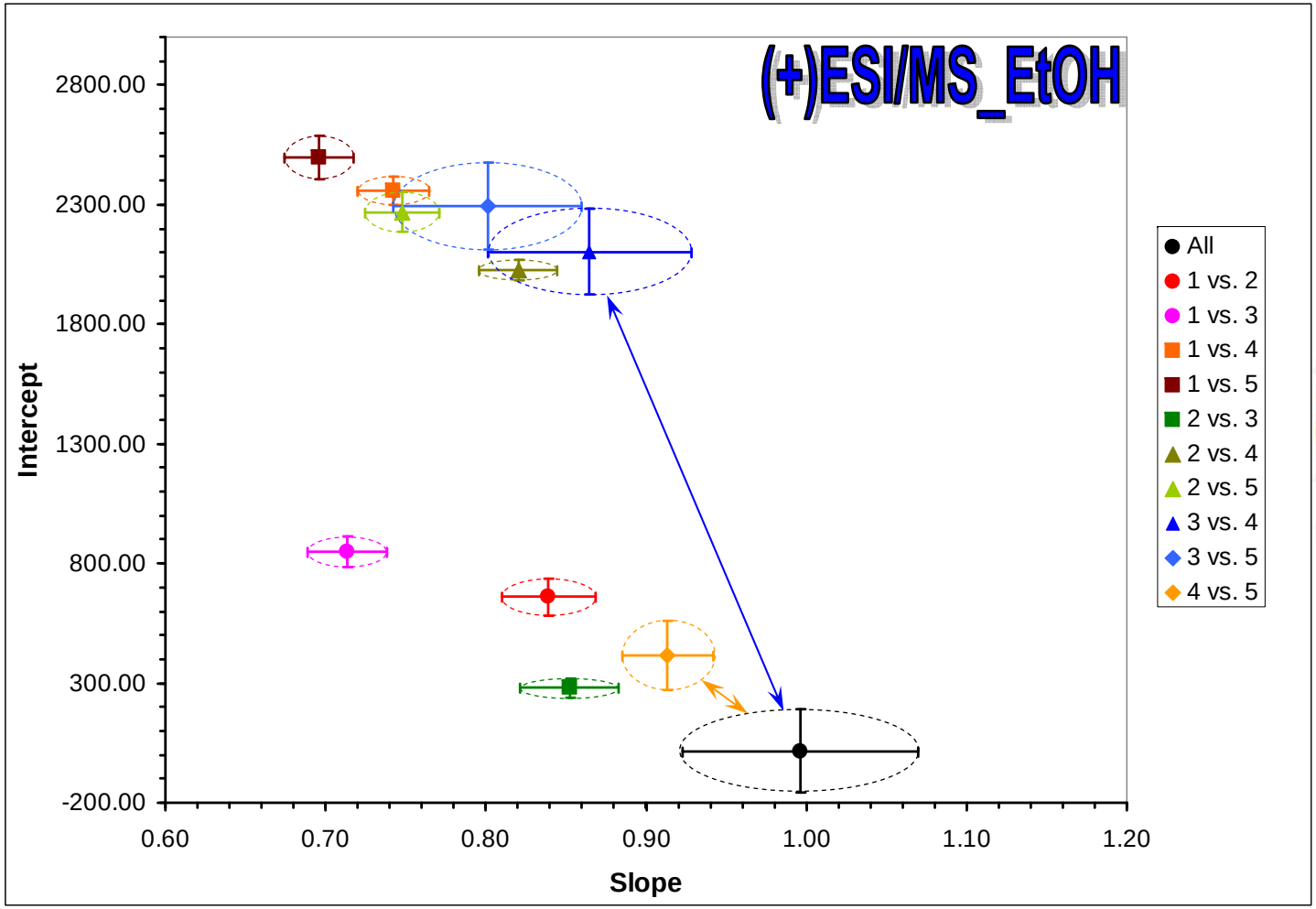
$$d_{1,2} = d(x_1y_1z_1, x_2y_2z_2) - r_1 - r_2$$

LRA - (-)ESI/MS for aqueous infusions;



Visual observation in the 3D space may be difficult!

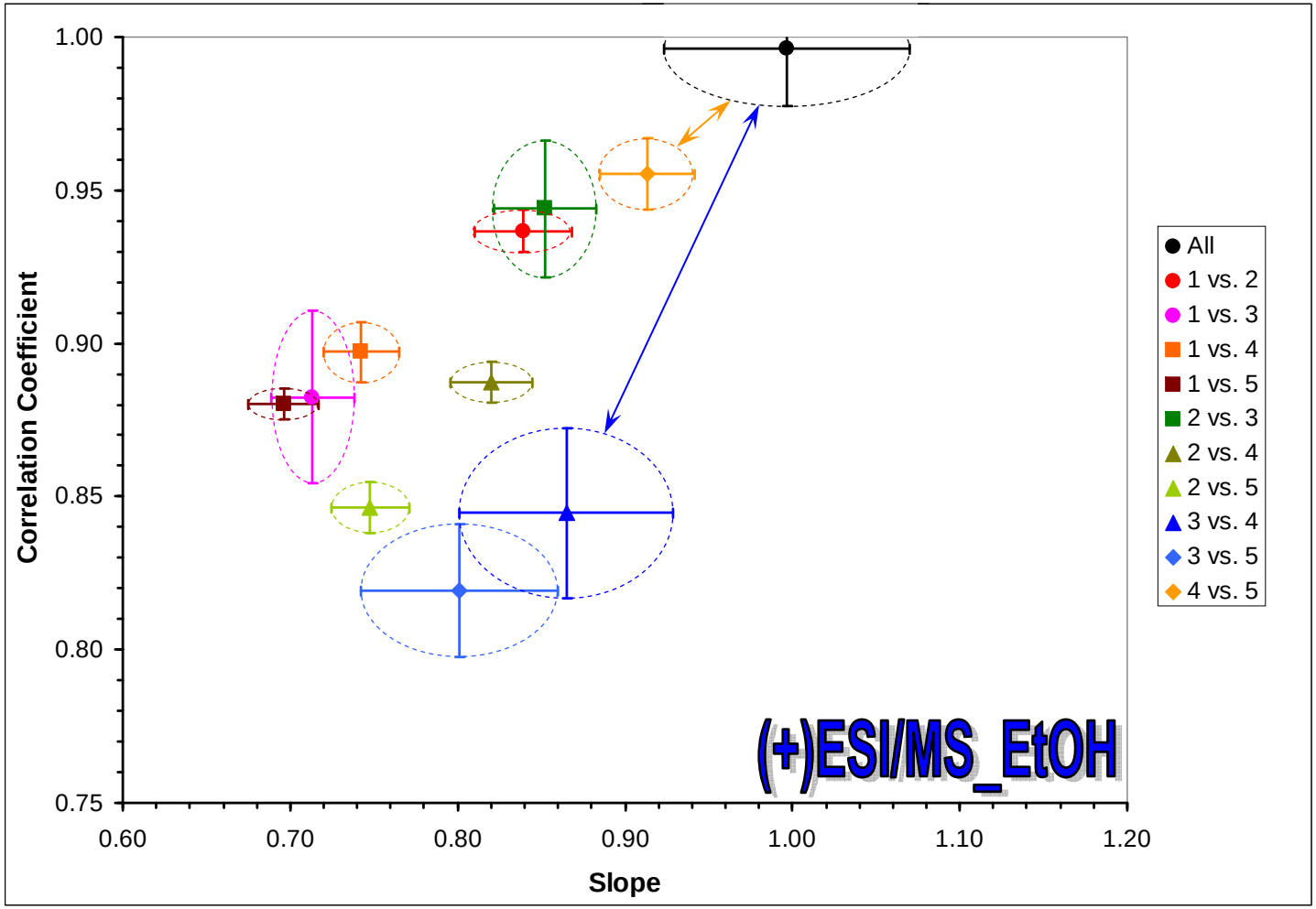
xOy



Projections in the xOy, xOz, and yOz planes may be used!

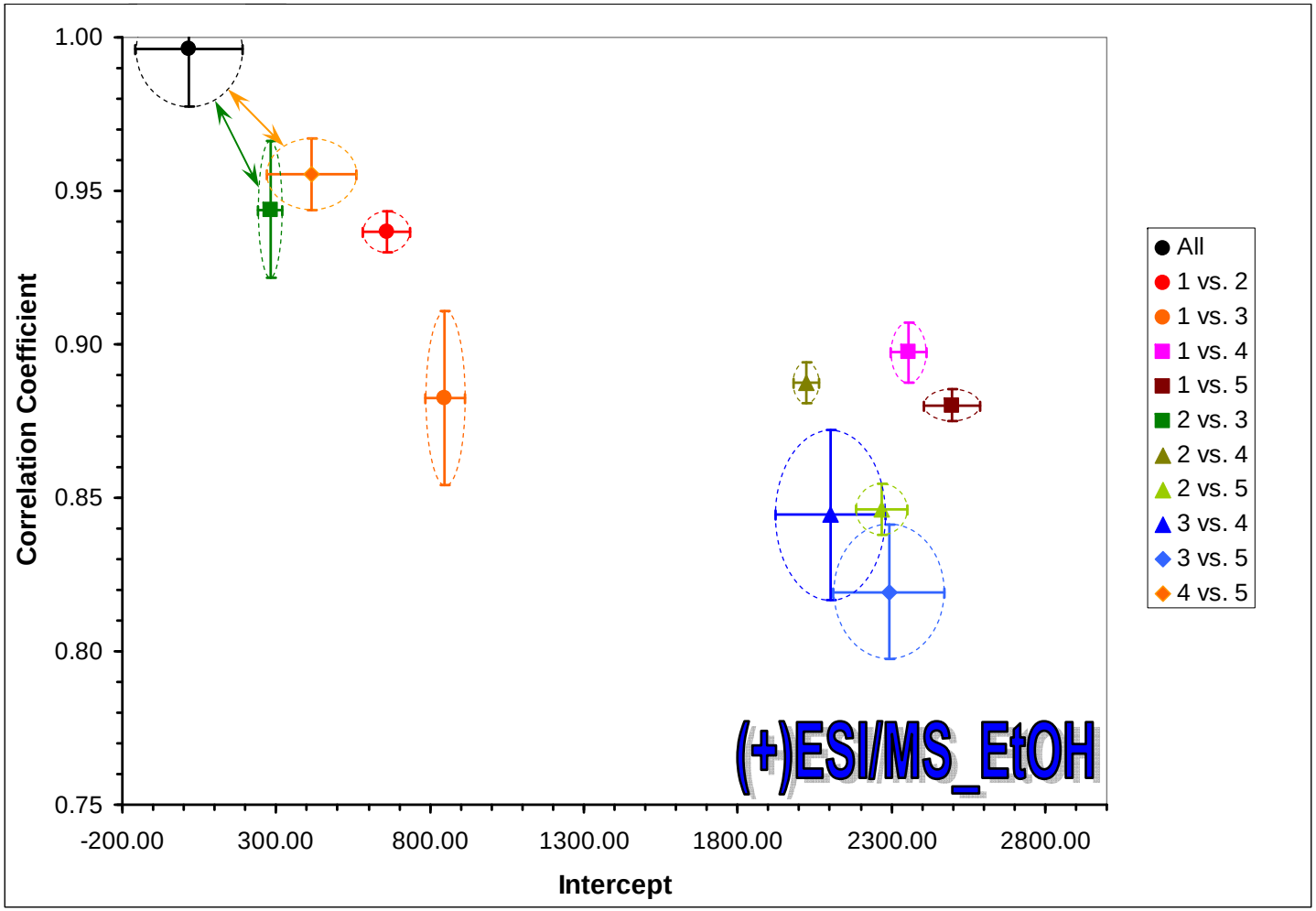
Superposition in all projections means "non-discrimination".

xOz



Computation of the true distance between ellipsoids is however a better choice!

yOz



Normalization reduces distances; Peak alignment negatively affects discrimination power!

Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5/1	3 vs. 5/2 to 5/ 6	3 vs. 5 all	4 vs. 5/1	4 vs. 5/2 to 5/6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
RPLC/UV	3.9	2.1	8.3	14.0	9.2	9.5	1.7	5.4	11.4	6.6	7.0	8.6	14.5	9.5	9.9	4.6	0.2	0.3	1.0
AQ	S	S	CS	ES	CS	CS	S	S	CS	S	S	CS	ES	CS	CS	S	LS	LS	LS
RPLC/UV	3.1	1.7	8.0	13.9	8.7	9.1	1.4	5.0	11.2	6.1	6.6	8.2	14.4	9.0	9.4	4.3	-0.3	-0.2	-2.5
AQ_PA	S	S	CS	ES	CS	CS	S	S	CS	S	S	CS	ES	CS	CS	S	NS	NS	NS
RPLC/UV	0.01	0.04	0.4	0.5	0.4	0.3	0.1	0.4	0.5	0.4	0.3	0.5	0.6	0.5	0.4	0.1	0.03	-0.02	0.1
AQ_N	LS	LS	CS	CS	CS	CS	LS	CS	CS	CS	S	CS	ES	CS	CS	S	LS	NS	S
RPLC/UV	0.3	0.9	0.2	0.4	-	-	1.5	0.4	0.2	-	-	0.5	0.8	-	-	0.5	-	-	-
EtOH	S	CS	S	S	-	-	ES	S	S	-	-	S	CS	-	-	S	-	-	-
RPLC/UV	0.04	0.2	0.1	0.1	-	-	0.1	0.1	0.2	-	-	0.3	0.3	-	-	0.03	-	-	-
EtOH_N	S	CS	S	S	-	-	S	S	CS	-	-	CS	ES	-	-	S	-	-	-

Qualitative data interpretation: NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.

Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S; from 50 to 90% means CS; > 90% means ES.

Normalization reduces distances; No significant effect on background subtraction!

Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5/1	3 vs. 5/2 to 5/ 6	3 vs. 5 all	4 vs. 5/1	4 vs. 5/2 to 5/6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
(+)ESI/MS AQ	58.2	51.9	90.9	221.8	363.8	260.1	36.3	114.7	252.6	458.0	338.8	115.1	260.2	451.1	341.3	70.5	280.5	221.5	184.5
	S	S	S	S	CS	CS	LS	S	CS	ES	CS	S	CS	ES	CS	S	CS	S	S
(+)ESI/MS AQ_BS	8.2	171.8	450.5	582.5	1022	863.7	131.2	501.7	647.8	1169	994.2	564.0	767.9	1268	1113	86.7	443.1	300.7	300
	LS	S	S	S	CS	CS	S	S	CS	ES	CS	S	CS	ES	CS	LS	S	S	S
(+)ESI/MS AQ_N	0.17	0.19	0.38	0.34	0.27	0.27	0.25	0.33	0.27	0.22	0.22	0.35	0.34	0.28	0.29	0.04	0.07	0.03	0.01
	S	CS	ES	CS	CS	CS	CS	CS	CS	CS	CS	ES	CS	CS	CS	S	S	LS	LS
(+)ESI/MS EtOH	474.4	688.6	2129	2263	-	-	162.9	1817	2042	-	-	1784	1989	-	-	206.2	-	-	-
	S	S	ES	ES	-	-	LS	CS	ES	-	-	CS	CS	-	-	LS	-	-	-
(+)ESI/MS EtOH_N	0.05	0.07	0.08	0.13	-	-	0.01	0.15	0.24	-	-	0.21	0.28	-	-	0.06	-	-	-
	S	S	S	S	-	-	LS	CS	CS	-	-	CS	ES	-	-	S	-	-	-

Qualitative data interpretation: NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.

Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S; from 50 to 90% means CS; > 90% means ES.

(+/-)ESI/MS data appear to be more informative than RPLC/UV ones!

Technique	1 vs. 2	1 vs. 3	1 vs. 4	1 vs. 5/1	1 vs. 5/2 to 5/6	1 vs. 5 all	2 vs. 3	2 vs. 4	2 vs. 5/1	2 vs. 5/2 to 5/6	2 vs. 5 all	3 vs. 4	3 vs. 5/1	3 vs. 5/2 to 5/ 6	3 vs. 5 all	4 vs. 5/1	4 vs. 5/2 to 5/6	4 vs. 5 all	5/1 vs. 5/2 to 5/6
(-)ESI/MS	58.2	51.9	90.9	221.8	363.8	260.1	36.3	114.7	252.6	458.0	338.8	115.1	260.2	451.1	341.3	70.5	280.5	221.5	292
AQ	S	S	S	S	CS	CS	LS	S	CS	ES	CS	S	CS	ES	CS	S	CS	S	CS
(-)ESI/MS	74.7	59.3	62.8	210.8	349.1	255.5	58.2	88.4	237.5	434.7	341.2	85.0	231.3	418.1	326.3	138.8	330.2	259.9	138
AQ_BS	S	S	S	S	CS	CS	S	S	CS	ES	CS	S	CS	ES	CS	S	CS	CS	S
(-)ESI/MS	0.13	0.03	0.31	0.55	0.20	0.07	0.03	0.20	0.48	0.27	0.16	0.28	0.58	0.24	0.13	0.17	0.27	0.22	0.20
AQ_N	S	LS	CS	ES	S	S	LS	S	CS	S	S	S	ES	S	S	S	S	S	S
(-)ESI/MS	47.6	8.8	324.0	362.8	-	-	24.9	336.9	374.7	-	-	362.7	402.4	-	-	76.4	-	-	-
EtOH	S	LS	CS	ES	-	-	LS	CS	ES	-	-	ES	ES	-	-	S	-	-	
(-)ESI/MS	0.08	0.04	0.18	0.20	-	-	0.06	0.11	0.14	-	-	0.12	0.16	-	-	0.03	-	-	-
EtOH_N	S	S	ES	ES	-	-	S	CS	CS	-	-	CS	CS	-	-	S	-	-	-

Qualitative data interpretation: NS - not separated; LS - low separation; S - separation; CS - clearly separated; ES - extensively separated.

Qualitative data interpretation refers to true distances (d) expressed as percentage from the maximum value in a set: < 0 means NS; <10% means LS; from 10 to 50% means S; from 50 to 90% means CS; > 90% means ES.

Thank you for your attention!

A.M. acknowledges the financial support given through the Romanian project PNII_ID_PCE_2011_3_0152.