Romanian Academy
Section of Medical Sciences

International Pharmaceutical Federation

Romanian Society for
Pharmaceutical Sciences

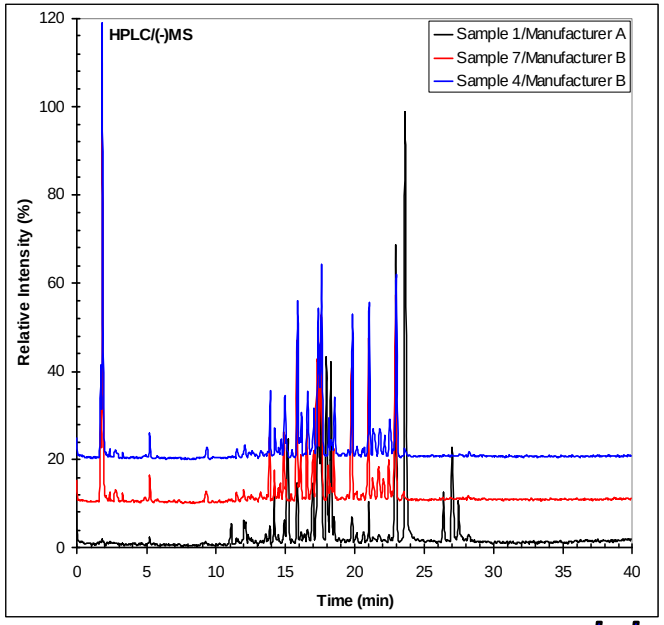
Chemometric approaches for characterization of complex mixtures of natural origins used as active ingredients in dietary supplements

Andrei MEDVEDOVICI (1), Victor A. VOICU (2)

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

(2)University of Medicine and Pharmacy "Carol Davila", Department of Pharmacology, Toxicology and Clinical Psychopharmacology, #8 Floreasca St., Bucharest - 014461 , Romania.

International Conference
"From Science to Guidance and Practice"
June 6-7th, 2016

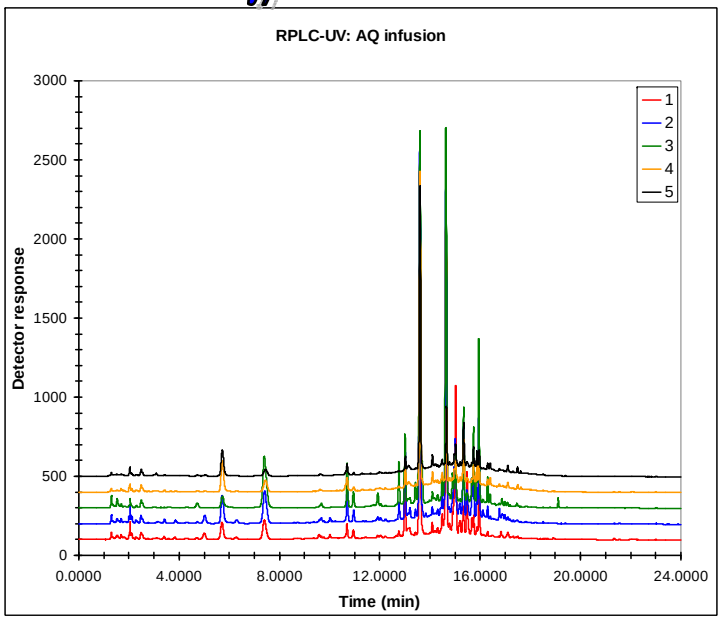


Mixtures of natural origins

COMPLEXITY

Need for Clasification!

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



Chemometric Approaches applied to:

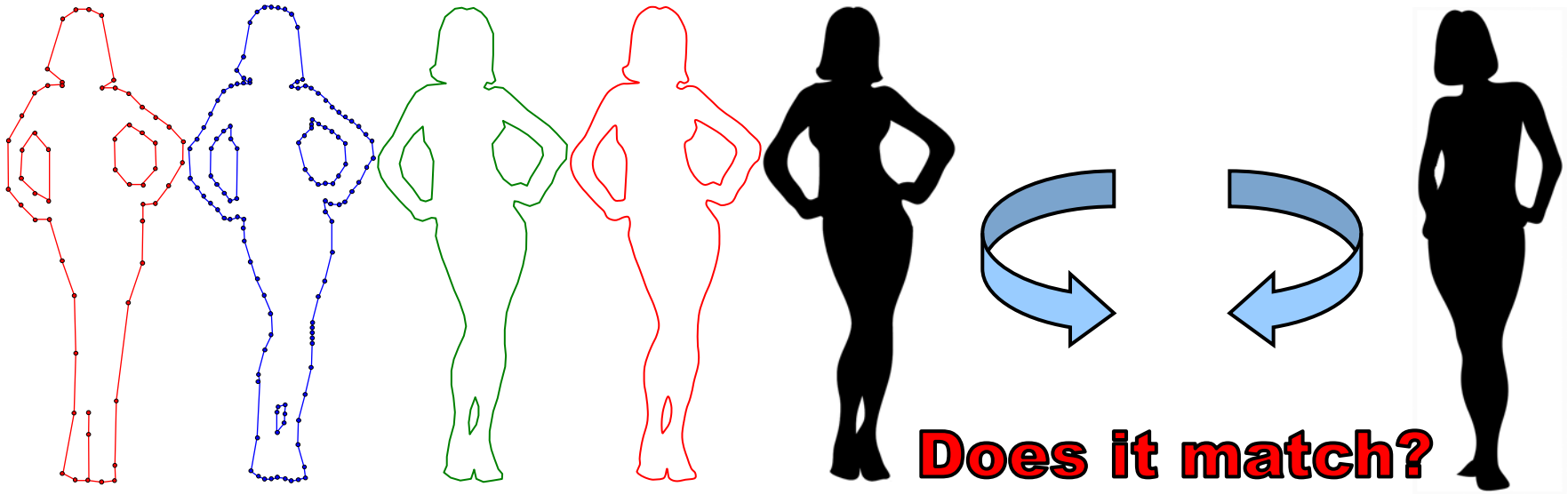
Holistic approach

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	-0.4	-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

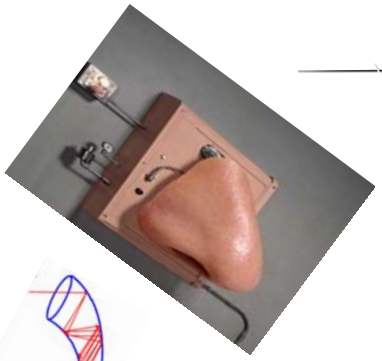
6900 data series for a 23 minutes chromatogram!

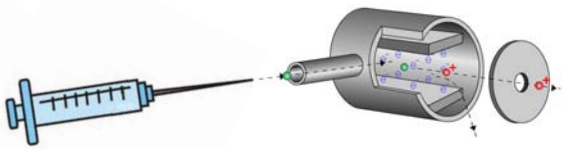
(one data every 0.2 sec) means

5 Hz acquisition frequency

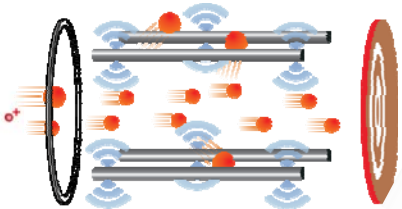
Usual PCA/CA softwares barely accept 1000 data in series.

Remember the Electronic Nose?

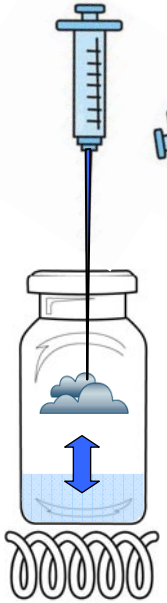




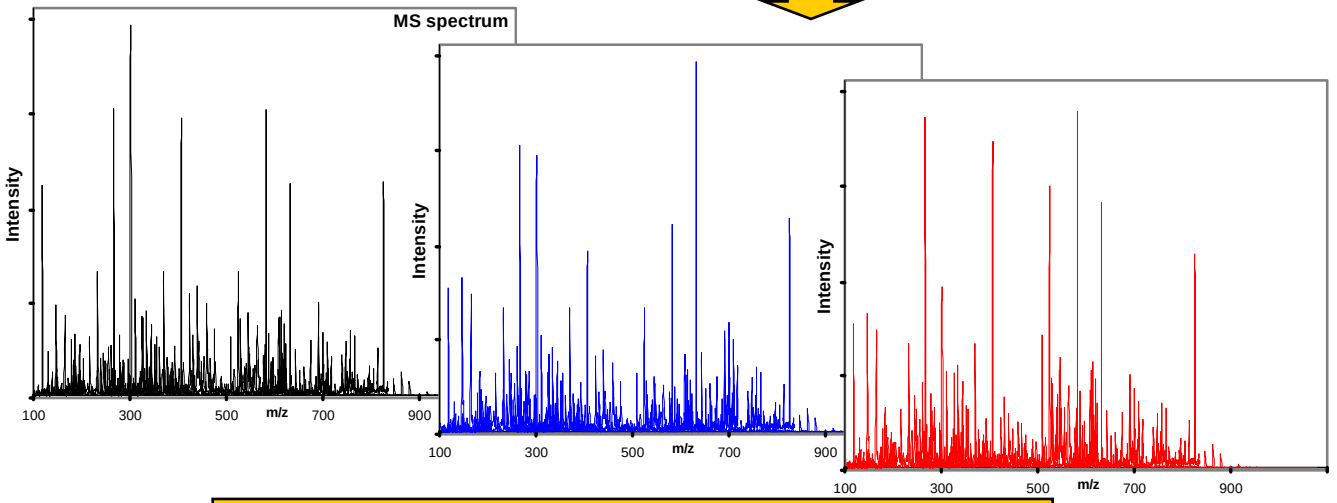
EI source



MS detector



Head-Space (HS)

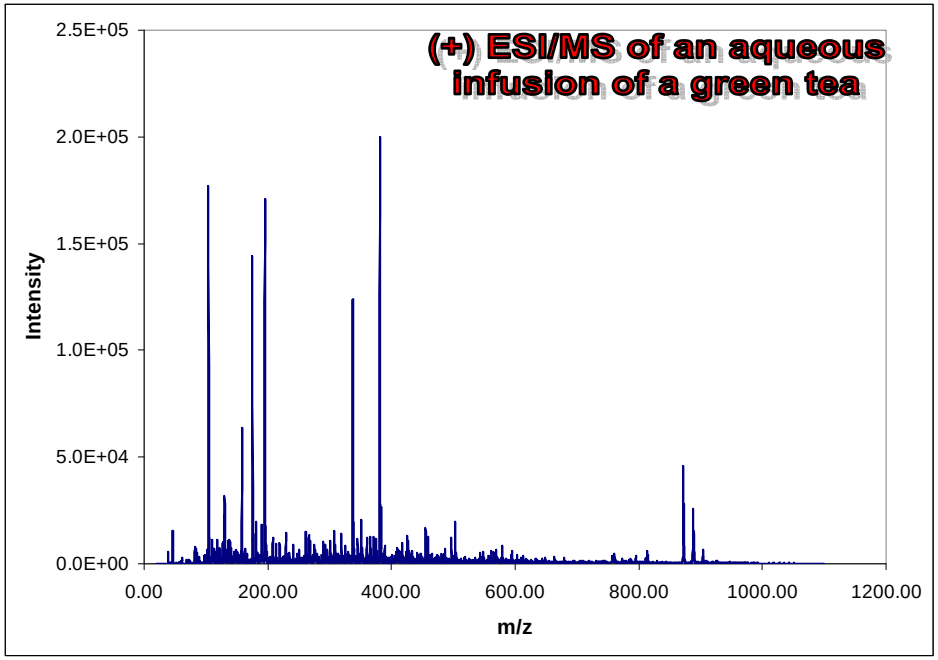


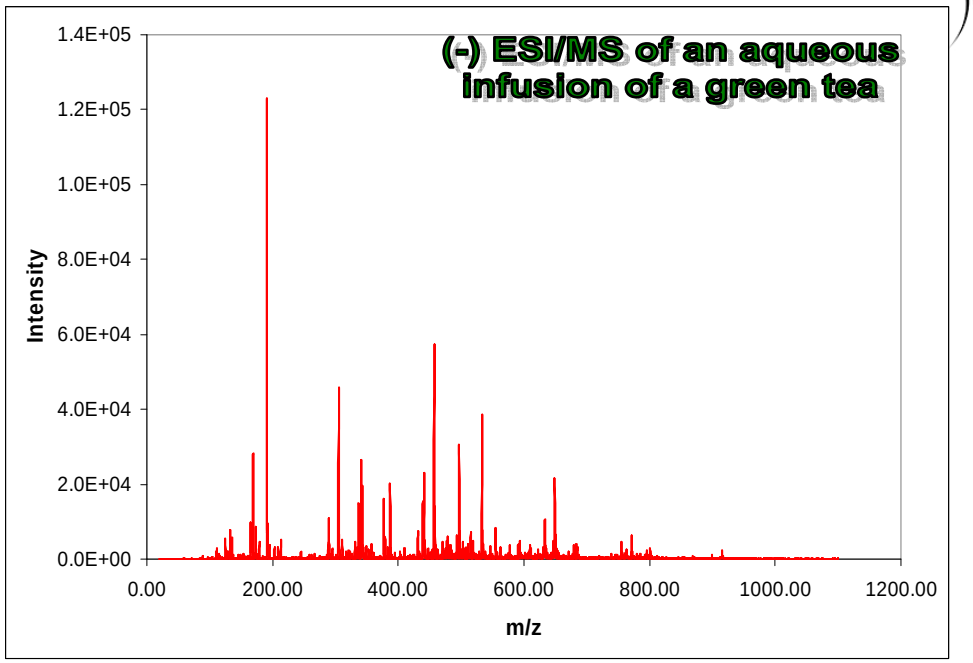
CHEMOMETRIC APPROACHES

Classification of Samples

Advantages of the AP ion sources:

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

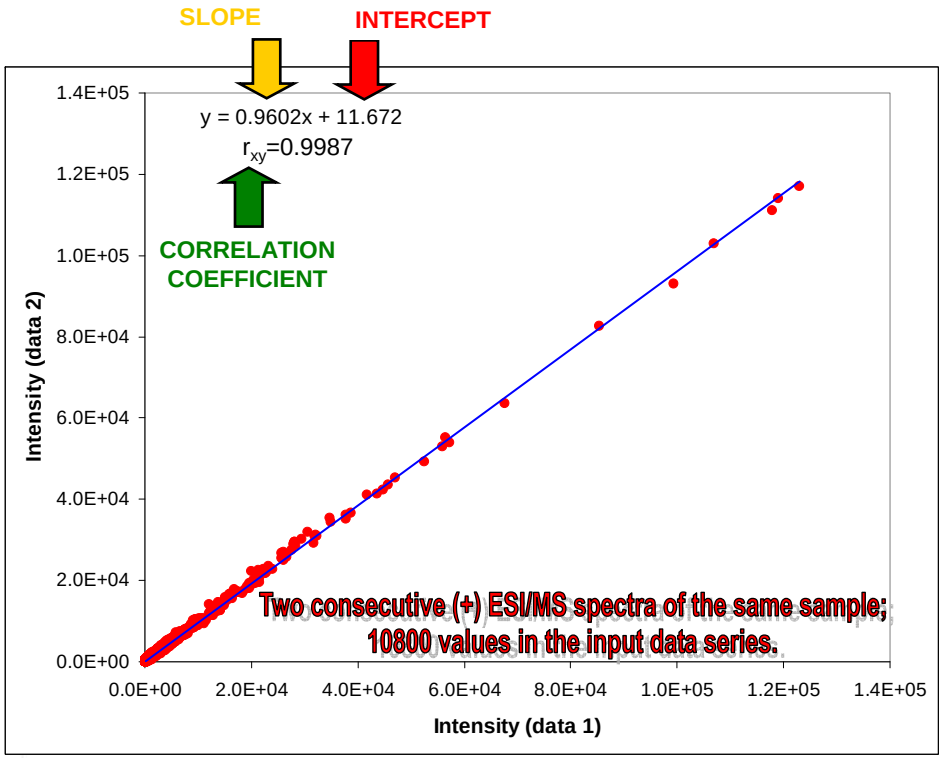


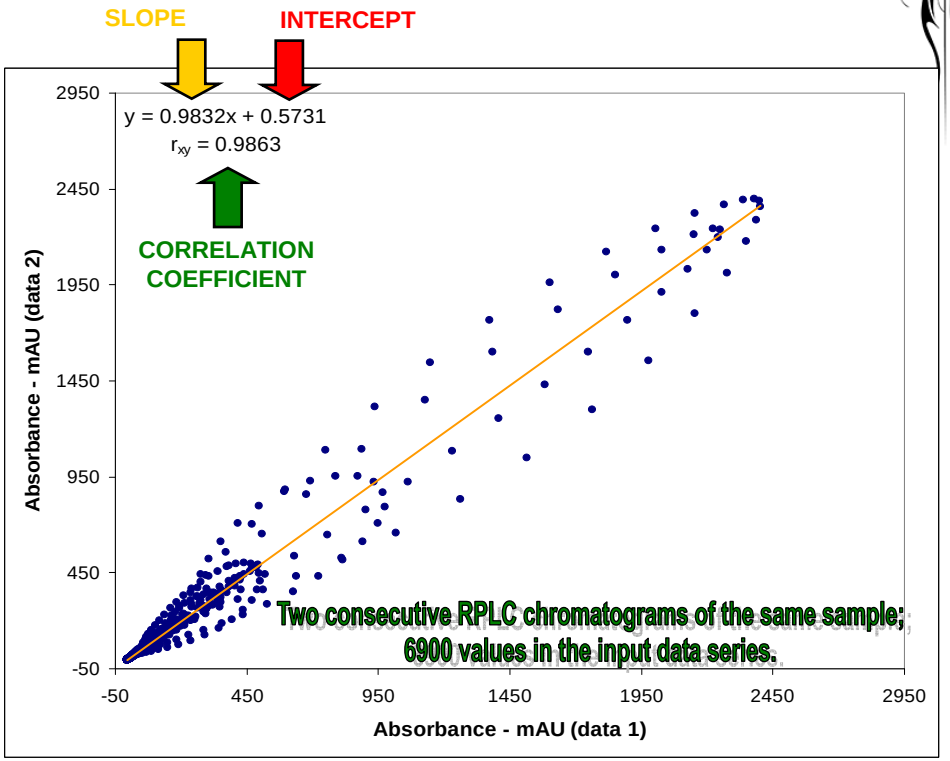


Are Direct Infusion (+/-) AP-IS/MS more informative than LC/Detection?

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)?

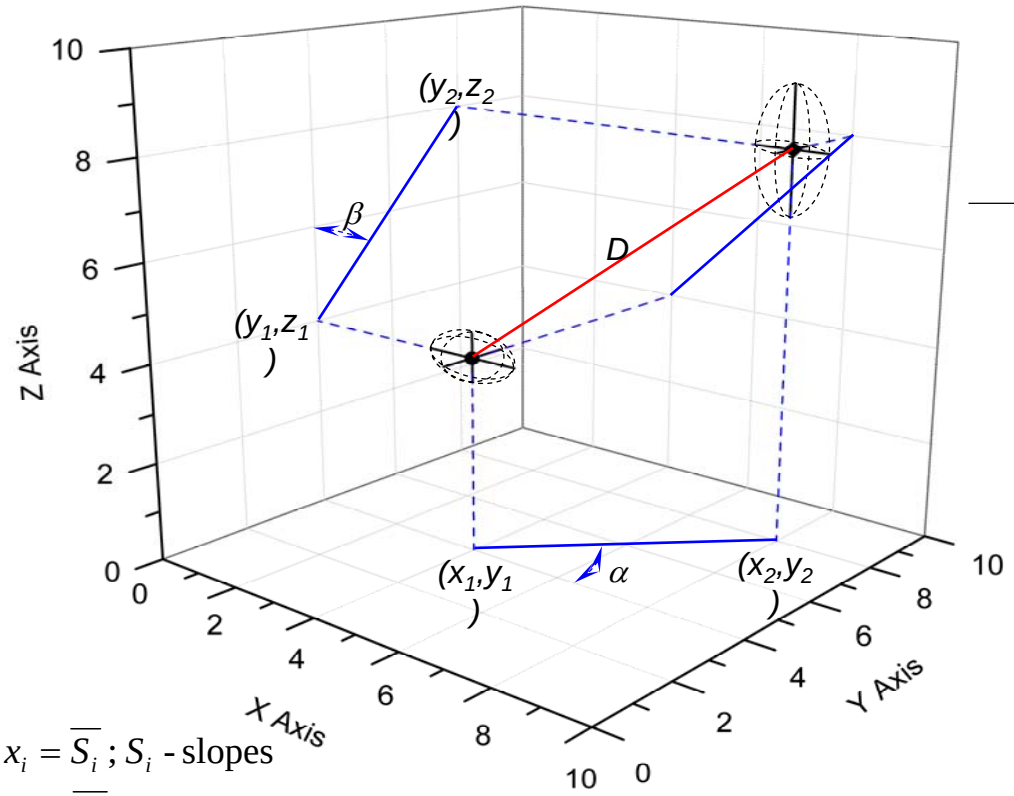


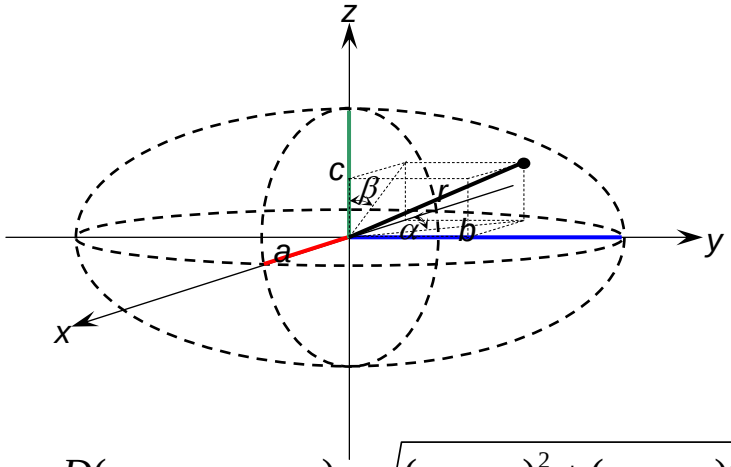


Two large sets of data may be reduced to 3 variables through expressing the Slope, Intercept and Correlation Coefficient of their reciprocal Linear Regression representation.

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

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

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

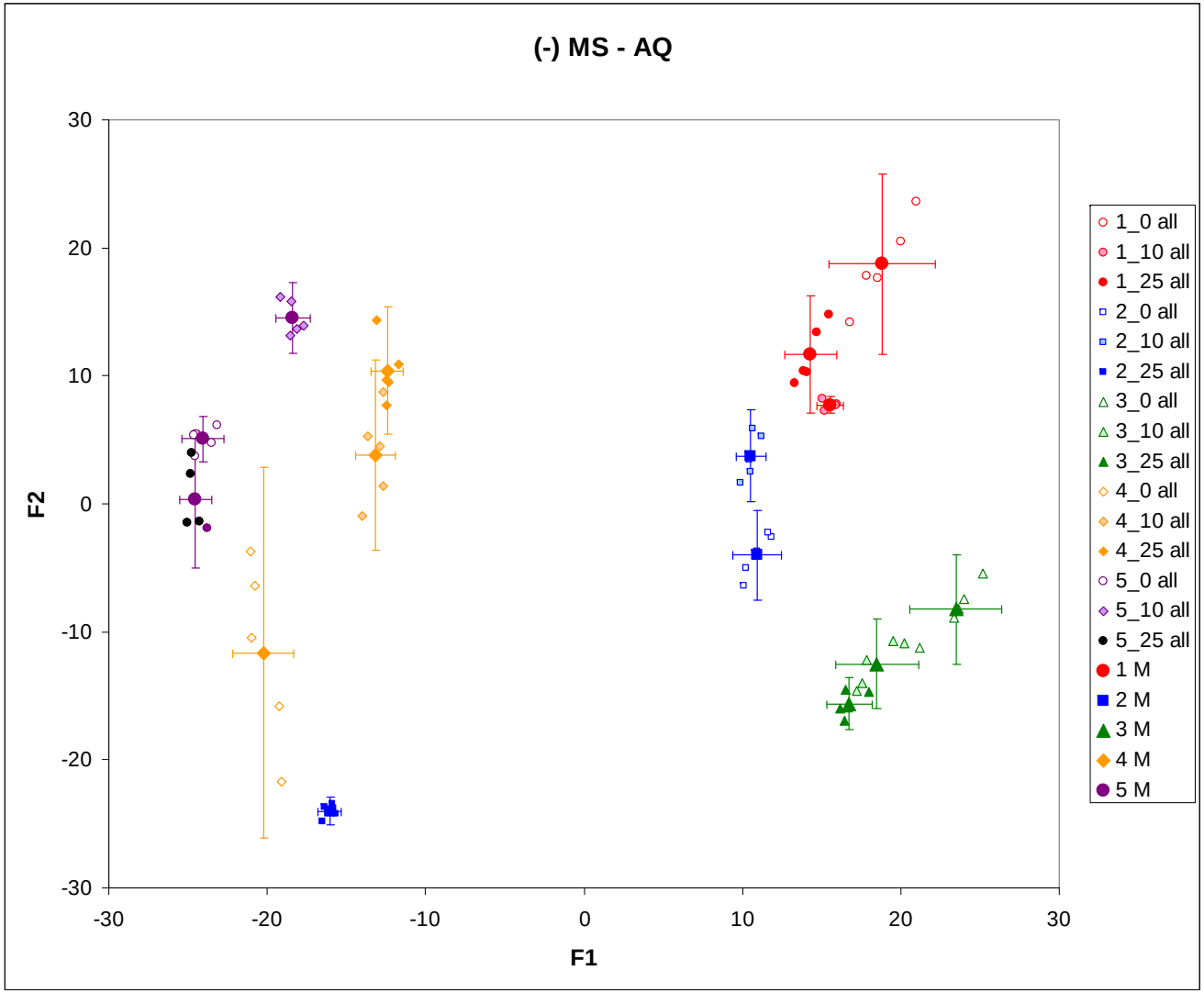
$$d_{1,2} = d(x_1y_1z_1, x_2y_2z_2) - r_1 - r_2$$

An application: the five green teas were gamma irradiated with 10 and 25 kGy doses

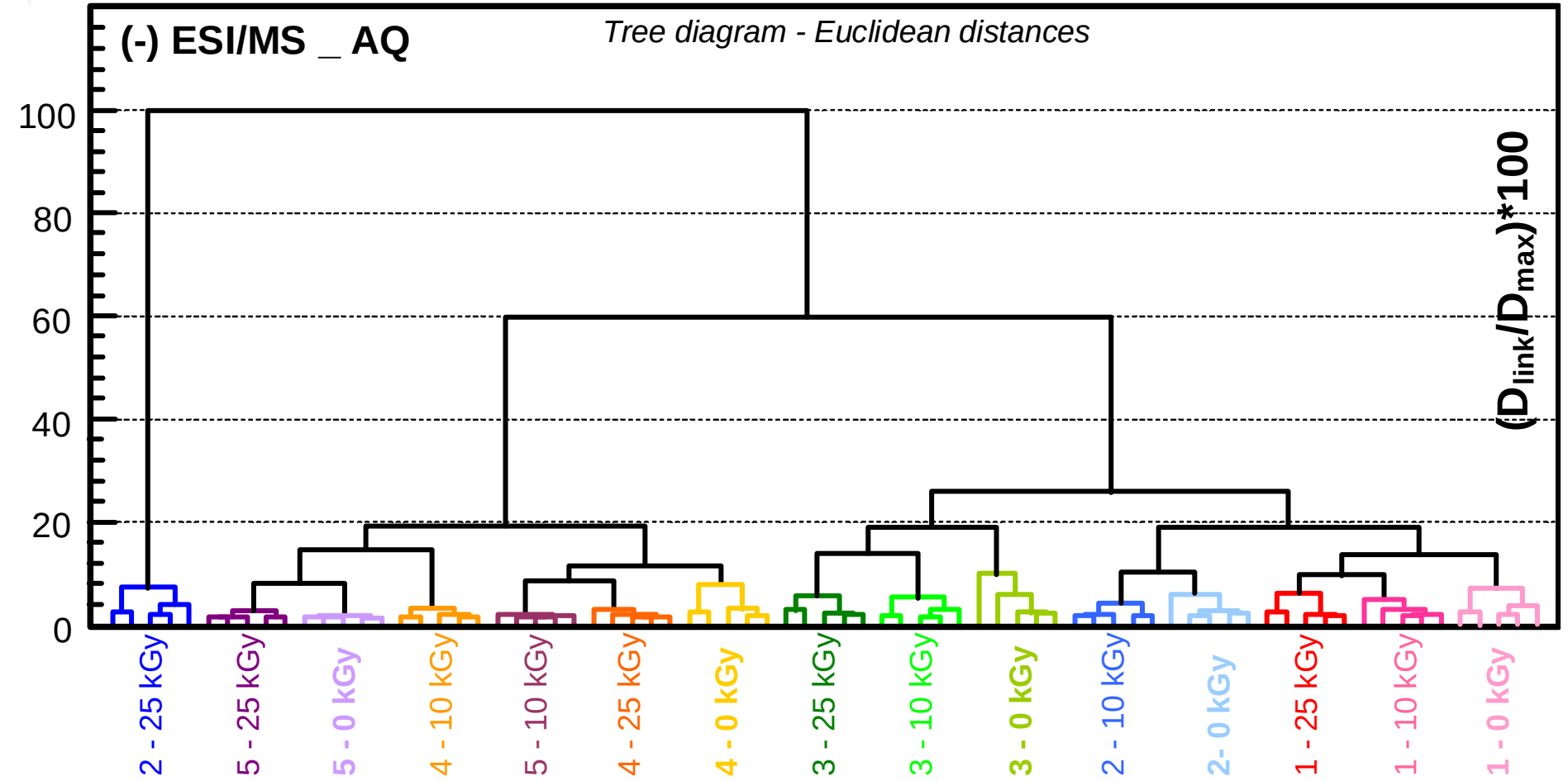
Same experimental approach: RPLC/UV chromatograms and (+/-)ESI/MS spectra

Same chemometric approaches: PCA, CA and LRA

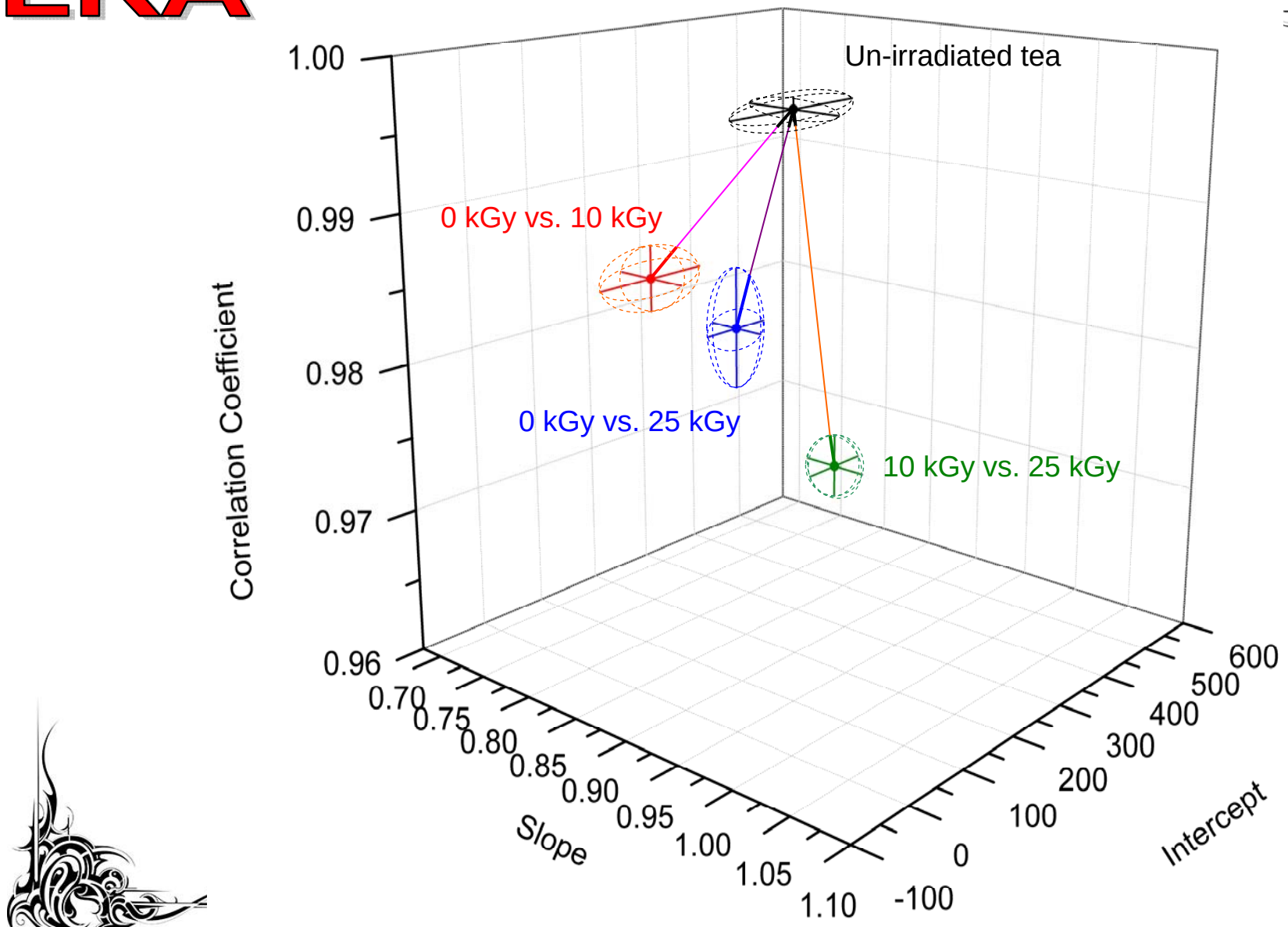
PCA



CA


(-) ESI/MS _ AQ
Tree diagram - Euclidean distances
100
80
60
40
20
0
(Dlink/Dmax)*100
2 - 25 kGy
5 - 25 kGy
5 - 0 kGy
4 - 10 kGy
5 - 10 kGy
4 - 25 kGy
4 - 0 kGy
3 - 25 kGy
3 - 10 kGy
3 - 0 kGy
2 - 10 kGy
2- 0 kGy
1 - 25 kGy
1 - 10 kGy
1 - 0 kGy

LRA


Un-irradiated tea
0 kGy vs. 10 kGy
0 kGy vs. 25 kGy
10 kGy vs. 25 kGy
Correlation Coefficient
1.00
0.99
0.98
0.97
0.96
Slope
0.70
0.75
0.80
0.85
0.90
0.95
1.00
1.05
1.10
Intercept
600
500
400
300
200
100
0
-100

Measured distances: comparison between PCA and LRA (I)

Experimental technique		RPLC/UV _ AQ				(+) ESI/MS _ AQ				(-) ESI/MS _ AQ			
Eigenvalues (%)*		F1	F2	LRA		F1	F2	LRA		F1	F2	LRA	
#	Tea Sample _ Compared Doses	45.53	28.66			35.11	17.75			34.52	15.74		
		d	Qual.	d	Qual.	d	Qual.	d	Qual.	d	Qual.	d	Qual.
1	1_0 vs 10 kGy	6.2	S	-0.30	NS	24.6	CS	24.4	S	4.6	S	56.4	CS
2	1_0 vs 25 kGy	-3.5	NS	0.20	S	23.3	CS	10.3	S	0.8	S	26.9	S
3	1_10 vs 25 kGy	10.4	CS	1.00	CS	-0.7	NS	20.3	S	-0.2	NS	7.6	S
4	2_ 0 vs 10 kGy	1.1	S	0.24	S	4.4	S	43.0	S	0.8	S	15.7	S
5	2_ 0 vs 25 kGy	22.6	CS	1.09	CS	16.3	CS	575.7	CS	30.9	CS	9.4	S
6	2_ 10 vs 25 kGy	21.5	CS	0.73	CS	11.8	CS	705.4	CS	36.2	CS	25.9	S
7	3_ 0 vs 10 kGy	-1.0	NS	0.10	S	6.5	S	266.0	CS	0.4	S	12.6	S
8	3_ 0 vs 25 kGy	3.9	S	1.70	CS	7.9	CS	199.2	S	4.8	S	67.9	CS
9	3_ 10 vs 25 kGy	5.6	S	1.33	CS	-1.0	NS	5.2	S	-1.5	NS	38.3	S
10	4_ 0 vs 10 kGy	8.4	CS	0.58	CS	0.4	S	-50.1	NS	9.7	S	115.6	CS
11	4_ 0 vs 25 kGy	4.1	S	0.26	S	-2.7	NS	-62.0	NS	15.4	CS	55.5	CS
12	4_ 10 vs 25 kGy	14.9	CS	1.44	CS	3.1	S	87.1	S	-3.8	NS	13.7	S
13	5_ 0 vs 10 kGy	0.9	S	-0.49	NS	1.2	S	14.7	S	7.6	S	163.5	CS
14	5_ 0 vs 25 kGy	5.7	S	0.15	S	-1.4	NS	17.9	S	-1.7	NS	5.1	S
15	5_ 10 vs 25 kGy	6.0	S	0.50	S	-3.2	NS	63.1	S	11.2	CS	84.9	CS

Measured distances: comparison between PCA and LRA (II)

Experimental technique		RPLC/UV _ EtOH				(+) ESI/MS _ EtOH				(-) ESI/MS _ EtOH			
Eigenvalues (%)*		LRA		F1	F2	LRA		F1	F2	LRA		F1	F2
				62.52	22.49			66.61	10.15			49.58	10.60
#	Tea Sample _ Compared Doses	d	Qual.	d	Qual.	d	Qual.	d	Qual.	d	Qual.	d	Qual.
1	1_0 vs 10 kGy	0.10	S	0.2	S	3.5	S	1.3	S	42.3	CS	2.6	S
2	1_0 vs 25 kGy	0.50	S	-25.7	NS	417.2	CS	7.1	CS	93.0	CS	1.6	S
3	1_10 vs 25 kGy	0.60	S	5.8	S	474.4	CS	3.4	S	135.7	CS	5.1	CS
4	2_ 0 vs 10 kGy	0.06	S	13.5	S	539.9	CS	7.0	CS	79.8	CS	9.4	CS
5	2_ 0 vs 25 kGy	2.66	CS	68.5	CS	785.4	CS	4.5	CS	35.7	S	12.5	CS
6	2_ 10 vs 25 kGy	2.87	CS	68.2	CS	427.0	CS	3.6	S	30.0	S	7.3	CS
7	3_ 0 vs 10 kGy	-0.001	NS	1.8	S	109.0	S	0.9	S	9.2	S	2.4	S
8	3_ 0 vs 25 kGy	0.05	S	0.6	S	27.1	S	7.1	CS	11.6	S	1.1	S
9	3_ 10 vs 25 kGy	0.03	S	4.8	S	59.3	S	9.3	CS	0.1	S	-4.1	NS
10	4_ 0 vs 10 kGy	0.11	S	3.7	S	10.3	S	7.2	CS	8.6	S	-5.6	NS
11	4_ 0 vs 25 kGy	0.22	S	10.4	S	61.9	CS	13.8	CS	1.4	S	-4.0	NS
12	4_ 10 vs 25 kGy	0.12	S	8.4	S	87.1	CS	3.1	S	11.8	S	-2.8	NS
13	5_ 0 vs 10 kGy	0.13	S	-0.5	NS	51.6	S	10.7	CS	4.5	S	2.4	S
14	5_ 0 vs 25 kGy	0.17	S	17.7	S	305.1	CS	10.7	CS	70.5	CS	-1.6	NS
15	5_ 10 vs 25 kGy	0.22	S	24.1	CS	191.1	S	1.7	S	98.5	CS	0.1	S

CONCLUSIONS:

A. (+/-)ESI/MS spectra are more informative compared to RPLC/UV chromatograms.

B. LRA has the advantage to consider large input data series, ensuring high definition of the profiles being compared.

C. LRA has a discrimination power at least equal to PCA and CA chemometric algorithms.

D. Data pre-processing techniques are not increasing the discrimination power.

E. Inter-day instrumental variability may affect the discrimination power of such a holistic approach.

F. The intra-day instrumental variability, if considered, add consistency to the discrimination, fixing the boundaries.

G. Holistic evaluation through chemometric approaches (including LRA) are successful in characterization of complex mixtures of natural origins.

Thank you for your attention!

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