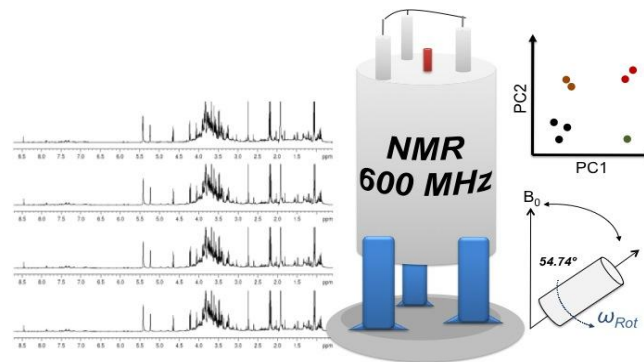


# Quimiometria aplicada à análise metabolômica

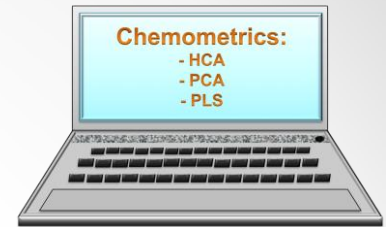
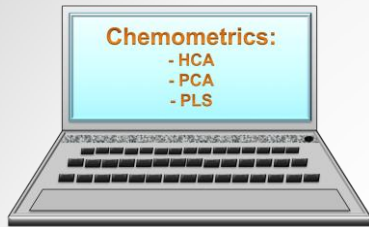


**Dr. Elenilson de Godoy Alves Filho**

[elenilson.godoy@yahoo.com.br](mailto:elenilson.godoy@yahoo.com.br)

**Dr Edy Sousa de Brito**

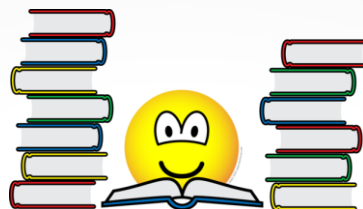
[edy.brito@embrapa.br](mailto:edy.brito@embrapa.br)



- 1 – Teoria para nivelamento do conhecimento metodológico e analítico
- 3 – Teoria sobre pré-tratamento de dados para uma análise quimiométrica
- 4 – Teoria básica sobre análises multivariadas exploratórias. Exemplos reais de aplicações
- 5 – Prática para construção de matrizes – Origin e Excel
- 6 – Criação de uma matriz na extensão para “Matlab-PLS\_Toolbox”. Prática sobre pré-tratamento dos dados
- 7 – Prática sobre análises multivariadas exploratórias - PLS\_Toolbox
- 8 – Teoria/prática sobre análises multivariadas por calibração e regressão - PLS\_Toolbox
- 9 – Teoria/prática sobre classificação binária - PLS\_Toolbox
- 10 – Teoria/prática sobre análises de agrupamento por *Heat Map*
- 11 – *Avaliação: trabalho escrito/apresentação oral sobre análises multivariadas de um alimento qualquer (dupla)*

**Programas:**

- *Origin*
- *Excel*
- *Matlab*
- *PLS Toolbox*



## BIBLIOGRAFIA SUGERIDA

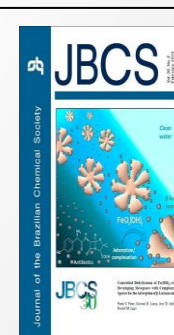
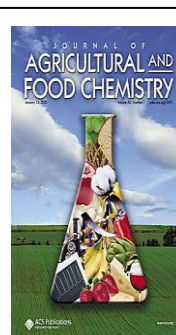
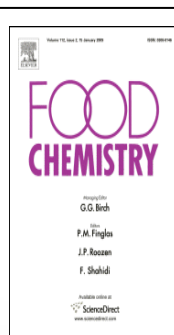
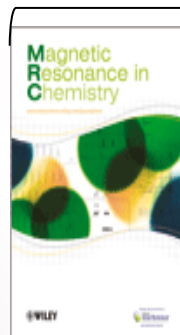
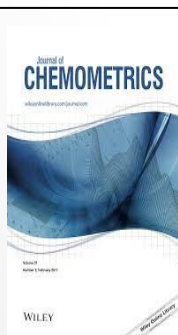
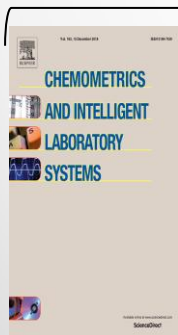
- 1) K. R. Beebe, R.J. Pell e M.B. Seasholtz, “*Chemometrics – A Pratical Guide*”, Wiley, N. York 1998.
- 2) H. Martens e T. Naes, “*Multivariate Calibration*”, Wiley, N. York, 2002.
- 3) D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte e L. Kaufman, “*Chemometrics: A Textbook*”, Elsevier, Amsterdam, 1988.
- 4) Matias Otto, “*Chemometrics – Statistics and Computer Application in Analytical Chemistry*”, 2<sup>nd</sup> ed, Wiley, N. York, 2007.
- 6) André Marcelo S., Ronei J. Poppi, *Quim. Nova*, Vol. 35, No. 1, 223-229, 2012.
- 7) André Marcelo S., Ronei J. Poppi, *Quim. Nova*, Vol. 36, No. 7, 1057-1065, 2013.
- 8) Albano P. Cunha Jr. e Vicente de Paula P. Cunha, *Quim. Nova*, Vol. 26, No. 6, 850-854, 2003
- 9) Márcia M. C. Ferreira, *Quimiometria, conceitos, métodos e aplicações*, ed. UNICAMP, Campinas, 2015 →

química nova



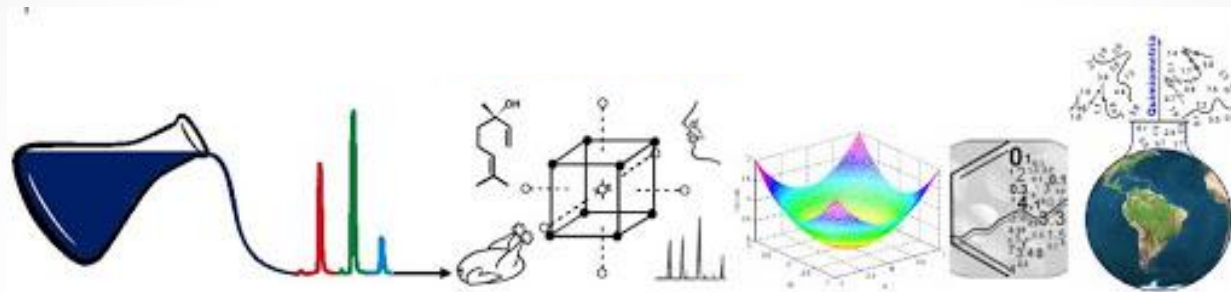
específica

aplicações



# QUIMIOMETRIA

*Aplicação de análises estatísticas multivariada em dados “químicos”*



Estatística Multivariada em outras áreas:

- Psicologia → Psicometria
- Economia → Econometria
- Biologia → Biometria

Conhecimentos básicos:

- *Informática/Programação;*
- *Estatística básica;*
- *Metodologias analíticas/Qualidade experimental*
- *Álgebra Linear (equações lineares);*



# QUIMIOMETRIA

Área da “química” a qual aplica metodologias estatísticas (algoritmos matemáticos) para extração dos maior número de resultados COM QUALIDADE.

obs.: Algoritmos Matemáticos → Etapas de um processamento para realização de uma determinada tarefa

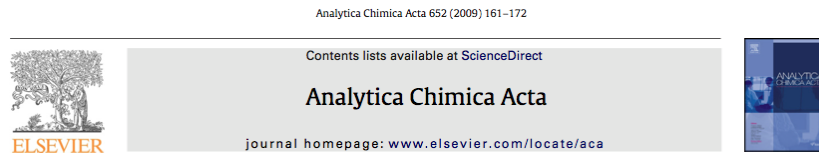
Análise Multivariada → abordagem analítica considerando o comportamento de duas ou mais variáveis, concomitantemente, envolvendo uma ampla variedade de conceitos estatísticos e matemáticos

Análise Univariada → não se detecta interações entre as variáveis presentes na matriz ou procedimento

Análise Multivariada → exploração de grandes conjuntos de dados químicos e dependência entre as variáveis

## Subáreas da Quimiometria:

1 – *Planejamento Experimental* → efeito de variações experimentais na resposta; busca maximizar a obtenção de resultados. Não se avalia uma variável por vez.



Experimental design in chemistry: A tutorial

Riccardo Leardi\*

Department of Pharmaceutical and Food Chemistry and Technology, Via Brigata Salerno (ponte), University of Genoa, I-16147 Genoa, Italy

2 – *Reconhecimento de Padrões* → conjunto complexo de dados → conjuntos simples → correlações amostras x variáveis.

3 – *Calibração Multivariada* → calibração de constituintes ou variáveis através de todo espectro (ou cromatograma)

# Quimiometria na Graduação



70 Universidades Públicas:

59 → Graduação em Química

20 → Quimiometria



34% dos Cursos de Química

- **Sudeste 43%**
- **Sul 35%**
- **Centro-Oeste 13%**
- **Nordeste 9%**
- **Norte 0%**

Aprofundar e ampliar os conhecimentos sobre o assunto

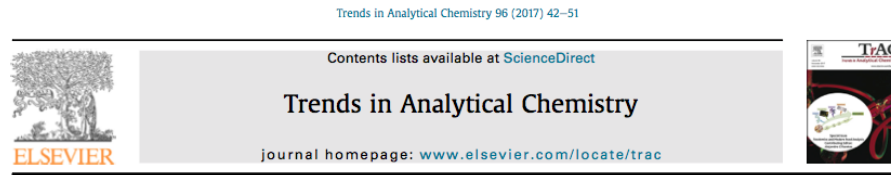
Análise mais completa e ampla dos resultados adquiridos

Despertar interesse dos alunos → Facilitar a compreensão da quimiometria

Incorporar aulas práticas envolvendo determinados assuntos

# Quimiometria aplicada às análises de alimentos

- **Metabolômica** → Estudo qualitativo e quantitativo de alterações de metabólitos em sistemas biológicos, relacionando variações dos metabólitos com alterações genéticas, perturbações das condições ambientais, dentre outros...
- **Metabonômica** → Investigação de um conjunto compostos através da amostra controle e amostras estudadas e, portanto, necessita-se de um conhecimento prévio acerca das características amostrais, ou seja, tem o objetivo de detectar quantitativamente mudanças metabólicas de sistemas alimentares frente a variações genéticas, ambientais...
- **Foodomics** → Estudos metabolômicos e metabonômicos desenvolvidos no âmbito alimentício, objetivando a o bem-estar e melhoria da saúde do consumidor.



## Recent chemometrics advances for foodomics

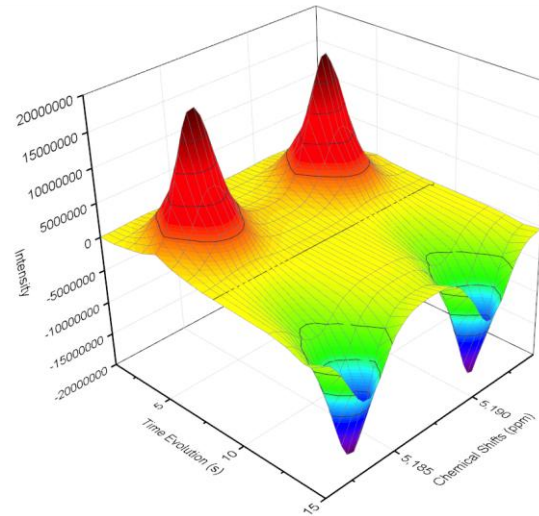
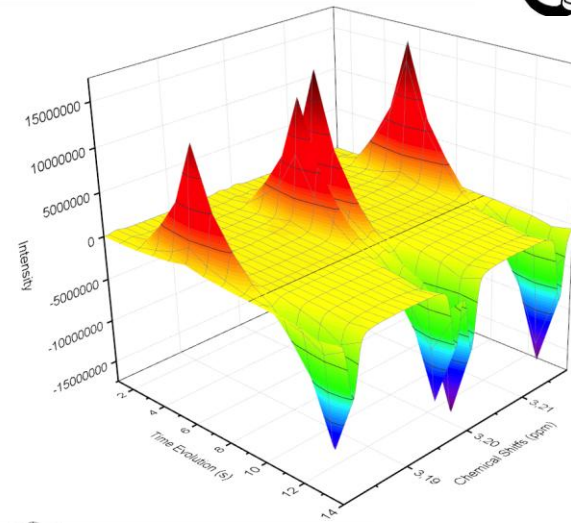
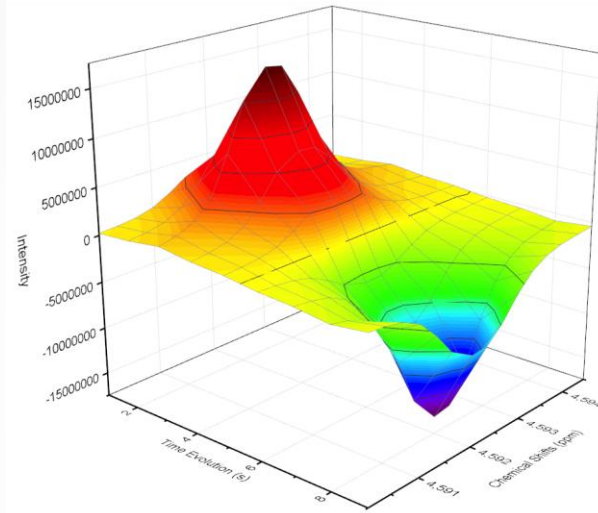
Marta Bevilacqua<sup>a</sup>, Rasmus Bro<sup>a</sup>, Federico Marini<sup>b</sup>, Åsmund Rinnan<sup>a</sup>, Morten Arendt Rasmussen<sup>a,c</sup>, Thomas Skov<sup>a,\*</sup>



## ***Estudos nas ciências ômica → Diversos tipos de análise:***

- Estudo de dinâmica molecular (mobilidade e flexibilidade de um sistema);
- Cinética de diferentes tratamentos ou reações;
- Caracterização de superfícies;
- Determinação da composição amostral em relação à origem;
- Adulteração de amostras de origem alimentar;
- Identificação de compostos orgânicos e organometálicos.

# POR QUE A QUIMIOMETRIA?



# Disposição dos dados para Quimiometria amostras e variáveis

## Análise Sensorial

<i>Amostras de Caju</i>	<i>Aroma de caju</i>	<i>Aroma doce</i>	<i>Aroma sulfuroso</i>	<i>Sabor de caju</i>	<i>Gosto doce</i>	<i>Gosto ácido</i>	<i>Maciez</i>	<i>Suculência</i>	<i>Fibrosidade</i>	<i>Adstrigência</i>	<i>Aspereza na boca</i>	<i>Irritação na garganta</i>
<b>A6.1</b>	6,0	6,5	2,4	6,9	6,8	2,5	4,4	4,6	4,4	2,6	4,6	4,0
<b>A6.2</b>	7,4	5,1	1,2	1,5	3,0	5,9	6,8	6,5	2,8	2,5	0,8	0,7
<b>A6.3</b>	7,8	0,9	0,6	8,1	6,2	6,0	4,5	6,9	7,1	0,5	0,7	0,9
<b>B9.1</b>	3,8	3,2	2,3	4,3	3,2	2,2	4,2	4,2	4,1	3,7	4,9	4,9
<b>B9.2</b>	5,1	3,1	3,1	3,4	2,5	5,0	2,2	2,2	2,6	2,2	1,6	4,6
<b>B9.3</b>	7,7	6,2	0,1	7,2	0,8	8,7	5,4	6,5	7,8	2,1	2,0	1,9
<b>C76.1</b>	5,6	4,4	2,0	5,6	6,5	1,1	5,9	5,9	2,3	2,2	2,1	2,0
<b>C76.2</b>	6,5	2,5	2,2	2,3	3,2	1,3	8,7	8,7	2,1	1,0	0,7	0,2
<b>C76.3</b>	8,6	2,6	2,8	8,5	4,7	6,3	6,7	6,5	7,0	2,2	1,4	0,7
<b>C10.1</b>	7,2	6,8	2,2	5,6	6,1	3,4	4,5	4,4	4,6	5,2	5,9	5,9
<b>C10.2</b>	5,5	6,8	2,2	3,7	5,3	4,4	4,5	4,5	3,5	4,2	1,4	3,4
<b>C10.3</b>	7,2	0,4	0,1	6,8	5,1	2,9	5,6	5,3	6,8	1,6	2,9	1,8
<b>B89.1</b>	5,6	5,4	1,6	6,5	5,5	2,3	5,5	5,5	3,5	5,0	5,3	6,5
<b>B89.2</b>	5,0	1,2	5,2	5,0	1,9	5,0	1,3	0,7	2,7	1,7	1,8	1,6
<b>B89.3</b>	7,2	2,5	0,1	6,5	5,7	6,8	6,4	5,7	7,2	1,4	1,4	0,7
<b>B65.1</b>	5,8	5,0	2,3	6,0	6,3	2,4	4,9	4,7	3,4	2,1	3,3	3,2
<b>B65.2</b>	1,7	1,8	0,0	1,7	1,8	4,3	2,0	2,0	4,4	3,0	3,0	1,6
<b>B65.3</b>	7,3	2,4	0,1	8,4	8,3	0,2	7,7	8,3	5,4	0,1	0,2	0,1
<b>B26.1</b>	7,2	7,1	1,3	7,0	6,5	7,2	7,5	7,5	7,5	7,7	7,6	6,3
<b>B26.2</b>	7,0	6,7	6,9	9,0	7,9	1,0	8,1	8,4	5,7	1,5	1,5	1,5
<b>B26.3</b>	8,0	7,4	1,5	6,8	5,0	6,6	5,1	6,6	7,1	2,9	3,7	1,6
<b>B74.1</b>	8,5	8,6	0,8	5,5	1,6	6,3	4,5	2,7	7,5	7,5	7,5	8,2
<b>B74.2</b>	7,1	7,0	2,4	5,1	3,7	8,8	3,6	3,9	3,5	8,6	6,5	7,2
<b>B74.3</b>	6,0	4,1	1,9	6,5	1,8	8,1	1,8	7,0	7,1	2,4	3,4	6,7
<b>B75.1</b>	8,4	8,5	2,7	8,1	7,5	5,1	7,5	7,5	5,5	6,8	4,4	4,1
<b>B75.2</b>	4,8	3,3	5,5	7,8	7,1	6,4	7,9	8,3	2,4	2,0	1,2	2,2
<b>B75.3</b>	3,9	7,8	6,5	7,6	6,6	3,2	3,3	7,4	6,7	2,1	1,6	1,2
<b>E50.1</b>	7,9	8,1	1,9	3,9	4,0	4,5	7,7	7,7	6,6	7,5	7,4	7,5
<b>E50.2</b>	5,6	3,3	6,5	4,6	1,0	6,6	3,1	6,5	6,8	6,5	5,9	6,1
<b>E50.3</b>	7,6	6,6	5,6	6,5	2,5	7,2	2,0	2,1	7,3	4,0	2,6	1,4

# Disposição dos dados para Quimiometria amostras e variáveis

## Análise Física

<i>Amostras de Caju</i>	<i>Peso</i>	<i>Comprimento</i>	<i>Diâmetro Basal</i>	<i>Diâmetro Apical</i>	<i>L* = luminosidade</i>	<i>C* = saturação</i>	<i>h* = tonalidade</i>	<i>Firmeza</i>
<b>A6.1</b>	94,54	71,20	52,92	37,10	69,57	62,16	90,9	3,38
<b>A6.2</b>	90,78	63,69	49,79	35,17	68,27	61,88	73,58	1,75
<b>A6.3</b>	86,65	61,48	50,50	29,72	69,05	61,17	93,67	2,25
<b>B9.1</b>	104,51	56,69	53,11	43,95	65,10	61,24	75,78	6,13
<b>B9.2</b>	91,70	57,05	53,44	39,05	69,86	63,09	73,24	6,50
<b>B9.3</b>	110,93	54,83	51,31	46,06	59,60	59,55	60,77	5,25
<b>C76.1</b>	125,45	64,65	61,34	48,35	66,55	64,78	67,76	9,50
<b>C76.2</b>	117,69	59,13	61,07	44,66	59,34	61,65	54,58	4,88
<b>C76.3</b>	134,02	71,52	58,96	44,75	61,95	60,82	61,62	3,50
<b>C10.1</b>	60,38	53,45	44,99	53,45	58,90	64,83	53,35	4,52
<b>C10.2</b>	58,59	50,04	44,29	50,04	60,43	60,60	42,02	6,63
<b>C10.3</b>	51,15	44,47	44,83	44,47	62,08	56,03	54,02	3,75
<b>B89.1</b>	175,86	76,26	66,81	49,37	46,77	55,09	45,05	4,13
<b>B89.2</b>	134,05	75,96	61,61	49,48	44,26	52,65	37,85	3,25
<b>B89.3</b>	151,87	69,68	59,94	41,95	42,94	50,92	39,65	3,50
<b>B65.1</b>	147,40	62,35	57,81	48,23	64,19	45,47	68,4	5,25
<b>B65.2</b>	122,06	59,41	54,08	43,85	55,73	59,01	50,08	4,25
<b>B65.3</b>	88,88	58,91	51,90	35,27	53,01	47,08	58,66	8,25
<b>B26.1</b>	114,74	68,68	49,44	35,53	65,34	57,10	74,87	6,50
<b>B26.2</b>	80,61	76,18	64,15	44,15	60,45	56,56	74,42	2,50
<b>B26.3</b>	96,52	60,53	52,50	35,37	60,59	54,29	61,70	3,88
<b>B74.1</b>	73,29	65,75	54,41	33,75	53,29	53,02	55,32	8,13
<b>B74.2</b>	125,02	58,74	47,35	32,53	57,58	49,56	59,10	6,13
<b>B74.3</b>	99,73	60,12	51,97	35,63	63,54	50,12	65,94	6,50
<b>B75.1</b>	122,31	66,26	49,75	46,00	62,88	52,87	75,73	6,50
<b>B75.2</b>	145,25	69,79	60,07	50,86	48,16	37,36	66,59	7,75
<b>B75.3</b>	135,83	72,34	53,07	42,34	63,08	57,10	47,43	4,00
<b>E50.1</b>	126,80	80,54	60,62	38,42	66,01	62,42	74,87	2,88
<b>E50.2</b>	84,81	69,73	53,62	41,13	68,68	62,10	74,42	3,13
<b>E50.3</b>	131,43	68,66	58,93	40,91	60,06	60,05	61,07	3,00

# Disposição dos dados para Quimiometria amostras e variáveis

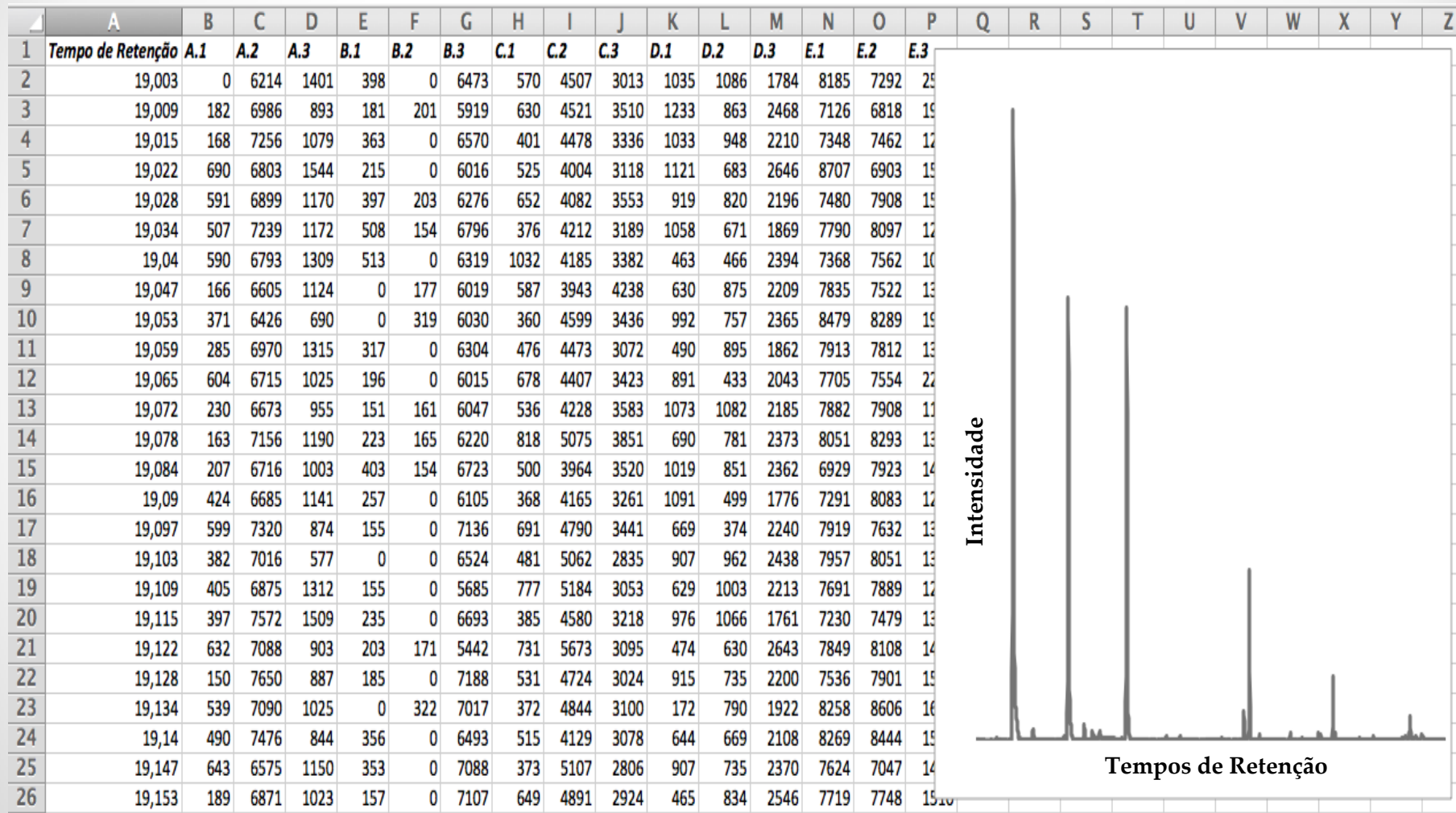
## *Análise Físico-química*

<i>Amostras de Caju</i>	<i>pH</i>	<i>Fenóis Totais</i>	<i>Acidez</i>	<i>Grau Brix</i>
<b>B9.1</b>	4,43	172,9432	0,273606989	13
<b>B9.2</b>	4,41	178,6518	0,274781503	13
<b>B9.3</b>	4,4	175,7975	0,274754074	13
<b>C76.1</b>	4,32	95,6666	0,310087921	11,6
<b>C76.2</b>	4,31	99,8312	0,273498242	11,7
<b>C76.3</b>	4,31	97,7489	0,291992581	11,6
<b>C10.1</b>	4,38	64,294	0,256667722	13,8
<b>C10.2</b>	4,38	60,1653	0,220044874	13,9
<b>C10.3</b>	4,36	62,22965	0,237149631	13,8
<b>B89.1</b>	4	151,6019	0,365973449	11,4
<b>B89.2</b>	3,99	148,6173	0,366485096	11,5
<b>B89.3</b>	3,98	145,1504	0,329441809	11,4
<b>B26.1</b>	4	278,1778	0,38293559	12,9
<b>B26.2</b>	3,99	270,8335	0,401809488	12,8
<b>B26.3</b>	3,96	274,50565	0,403012871	12,9
<b>B74.1</b>	4,02	191,7495	0,403536578	13,4
<b>B74.2</b>	4	194,4135	0,440001808	13,5
<b>B74.3</b>	3,99	196,9543	0,438599552	13,3
<b>B75.1</b>	3,9	174,5473	0,439694299	14
<b>B75.2</b>	3,87	178,0313	0,439255745	14,1
<b>B75.3</b>	3,87	179,5128	0,438730634	14
<b>E50.1</b>	4,26	215,9789	0,292852272	11,7
<b>E50.2</b>	4,23	204,4347	0,293217357	11,6
<b>E50.3</b>	4,23	210,2068	0,292254171	11,8
<b>E51.1</b>	4,05	331,9116	0,348439203	13,1
<b>E51.2</b>	4,04	336,4397	0,456965553	13,2
<b>E51.3</b>	4,03	334,17565	0,456556249	13,1



# Disposição dos dados para Quimiometria amostras e variáveis

*Cromatogramas: LC-UV, LC-MS, GC-MS...*

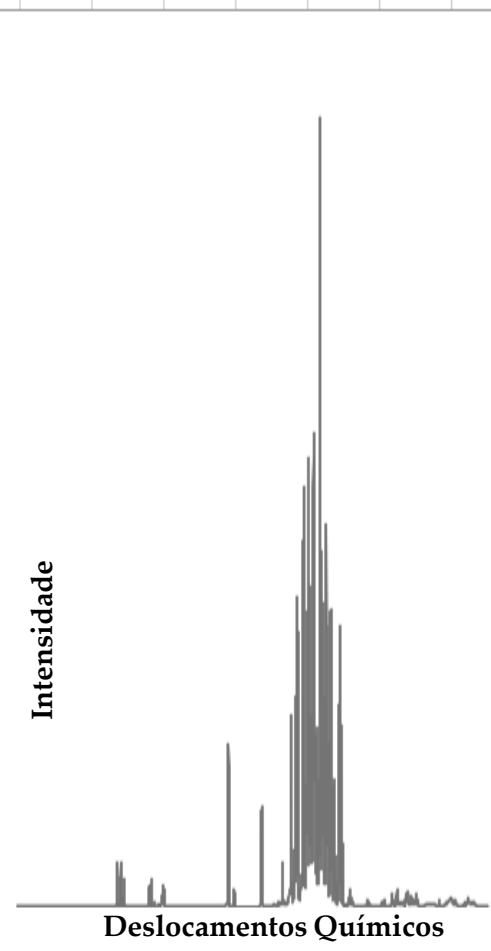




# Disposição dos dados para Quimiometria amostras e variáveis

*Espectro de RMN de  $^1\text{H}$*

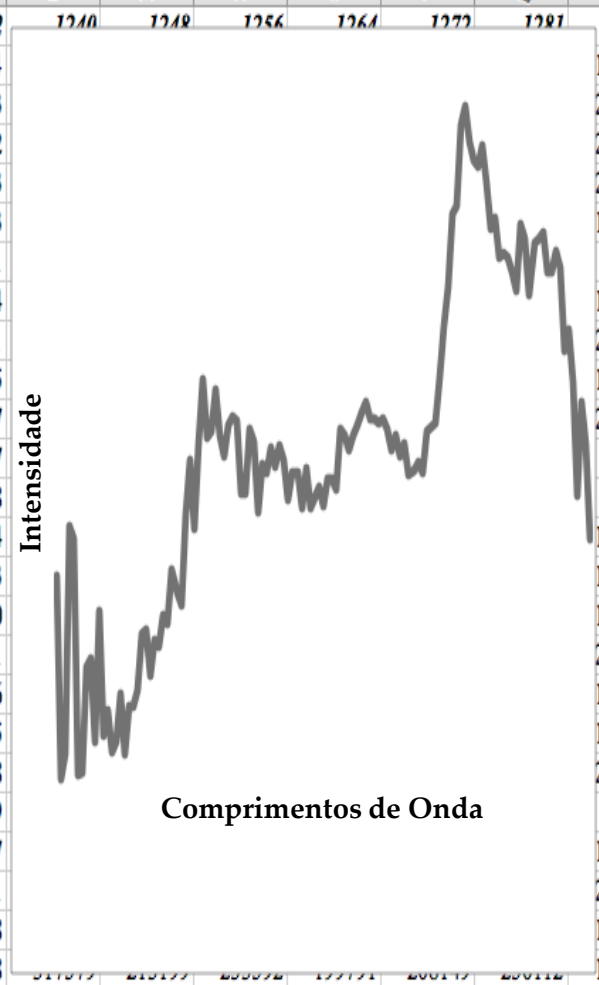
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	<b>Deslocamento Químico</b>	<b>A.1</b>	<b>A.2</b>	<b>A.3</b>	<b>B.1</b>	<b>B.2</b>	<b>B.3</b>	<b>C.1</b>	<b>C.2</b>	<b>C.3</b>	<b>D.1</b>	<b>D.2</b>	<b>D.3</b>	<b>E.1</b>	<b>E.2</b>	<b>E.3</b>							
2	9,0008	70503	6603	-20719	-9714	-39401	-14184	36280	-30333	-25688	-36269	-6377	-4819	-67051	-13699	34							
3	8,9995	68838	687	-16184	-17514	-28572	-11150	59508	-13414	-15669	-33726	-10094	-12891	-69446	-21692	27							
4	8,9981	66184	-3856	-16742	-1217	-32013	-2625	63455	-6498	-25644	-21311	-13856	-1503	-63699	-17761	-33							
5	8,9968	77365	10571	-5880	-17921	-10984	-2593	80240	-4628	-21000	-27937	-5050	940	-61526	-14402	149							
6	8,9955	78976	258	553	6645	-15876	-10592	83828	-5623	-9514	-25929	-8388	-10709	-65197	-18001	30							
7	8,9941	108703	-8324	2671	12442	-14530	-17837	84961	-662	-19010	-29716	1838	1237	-47751	-20918	57							
8	8,9928	116031	25290	9259	23980	-505	12398	94442	1920	-25003	-31297	4041	-1498	-50932	-19179	156							
9	8,9914	116802	14294	18747	10737	-7143	2636	92603	17177	-16438	-27420	634	985	-50309	-14194	200							
10	8,9901	109654	3179	4086	9070	-12244	24742	87447	22440	-17566	-26411	-660	-7441	-54181	-15852	146							
11	8,9888	94804	4497	5431	3785	6282	18542	85359	22588	-20819	-31814	4189	-314	-52245	-18042	123							
12	8,9874	91499	-12803	-19653	-6895	5108	23018	86284	-4093	-15786	-40176	7780	-6875	-61617	-15312	92							
13	8,9861	92033	-3007	-17455	-6828	-70	5995	92736	2264	-12779	-35442	1856	-3573	-55627	-26470	207							
14	8,9847	84308	-13024	-19378	-17265	-27258	-7821	70808	731	-12199	-34950	-10565	-11203	-62039	-17946	140							
15	8,9834	98633	-7151	-10044	-10777	-22482	-4257	78532	-10861	-17362	-26685	3001	-8060	-53730	-24508	100							
16	8,9821	92379	-6659	-24080	-10299	-34528	-14232	67373	-6404	-20631	-30000	-12197	-9308	-52680	-28562	103							
17	8,9807	92496	-20328	-11429	-10501	-12105	-33167	69944	-14995	-14949	-30933	-10356	-3815	-53923	-23546	110							
18	8,9794	70918	-3800	7242	-12003	-18113	-26919	62773	-23330	-10664	-32039	-13768	-10494	-58305	-21872	69							
19	8,9781	76147	-9353	-4519	-11036	-30088	-18188	85278	15519	-20698	-35861	-17533	-13007	-60188	-14043	74							
20	8,9767	76435	-179	-208	-16229	-38623	-12075	61194	-6146	-20807	-29981	-10393	-9044	-64751	-17431	76							
21	8,9754	72219	-3504	5567	-3639	-36617	-2205	71774	-3463	-21739	-27616	-11397	-8473	-55199	-24326	-2							
22	8,974	99448	-7410	2451	2577	-17548	270	78945	-8080	-23386	-31084	-22659	-9675	-61930	-24394	57							
23	8,9727	82646	-4761	-17932	-6852	-40498	-13635	74918	-6138	-22510	-38566	-5030	-6547	-54598	-23221	99							
24	8,9714	70352	-20956	-25411	-8988	-38903	-9643	72910	-16949	-10608	-37304	-7959	-6792	-49718	-28939	97							
25	8,97	70338	12794	-23244	10882	-24641	12036	66119	695	-12849	-28905	-10934	-13194	-59783	-20445	47							
26	8,9687	82632	2187	-12958	2225	-29988	-26596	78357	-5030	-16243	-33163	-20000	-7644	-57670	-12987	166							



# Disposição dos dados para Quimiometria amostras e variáveis

*Espectro de IV*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	<b>Comprimentos de Onda</b>	<b>1158</b>	<b>1166</b>	<b>1175</b>	<b>1183</b>	<b>1191</b>	<b>1199</b>	<b>1207</b>	<b>1215</b>	<b>1224</b>	<b>1232</b>	1240	1248	1256	1264	1272	1281
2	A.1	230582	91892	108996	265134	256272	95063	96957	169000	174272	118074						
3	A.2	459103	230994	253517	522469	496823	231740	235063	348726	363146	278573						
4	A.3	358639	160697	182988	412187	392183	162558	165343	267056	281640	205762						
5	B.1	392045	177959	198733	450815	422711	179037	181457	287482	304278	221473						
6	B.2	356684	151269	174743	413999	392204	158629	160802	262918	279625	199643						
7	B.3	229484	68911	85711	267043	255989	69349	70097	153705	164630	101231						
8	C.1	354921	149523	169971	411807	386961	152620	154121	257411	271739	192574						
9	C.2	395513	172267	198377	454027	436246	182088	182521	295341	311664	224811						
10	C.3	299202	111210	132842	346227	332091	114961	116043	211934	224395	150885						
11	D.1	370063	155155	179502	426097	406350	160757	160665	268407	285309	201697						
12	D.2	250559	80533	95898	286122	267570	74137	74454	158823	170285	103087						
13	D.3	249709	77317	95075	282366	266993	74387	76957	159214	170596	105208						
14	E.1	369784	162364	181397	421564	402990	161979	164412	268762	285090	204124						
15	E.2	359352	152456	174253	407014	382057	152602	152121	255067	269296	191513						
16	E.3	290836	107773	126741	332918	317642	111776	111737	204237	216571	143540						
17	F.1	466241	229870	251763	537518	511306	234645	238385	356701	376202	280521						
18	G.1	296339	121506	140187	342550	330428	126660	128927	219030	230404	158486						
19	G.2	314384	132957	154432	358724	345416	135765	137481	229465	238915	166785						
20	G.3	391699	188267	208893	448508	430510	193424	194527	296993	310612	228968						
21	H.1	110944	6850	5574	137667	131654	6711	5240	56820	62801	10709						
22	H.2	353629	169380	183867	394092	376206	164889	169736	260398	271770	203087						
23	H.3	390988	195721	215246	431680	415991	192146	196679	293146	305691	233111						
24	I.1	314368	146937	164453	352839	341587	146393	148891	235350	244490	180718						
25	I.2	342552	167750	182097	383033	372159	166871	169035	259020	270402	202148						



# Importância de pesquisas relacionadas à Quimiometria

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	<i>Amostras de Caju</i>	<i>Aroma de Caju</i>	<i>Aroma Doce</i>	<i>Aroma Sulfuroso</i>	<i>Sabor Caju</i>	<i>Gosto doce</i>	<i>Gosto ácido</i>	<i>Maciez</i>	<i>Suculência</i>	<i>Fibrosidade</i>	<i>Adstringência</i>	<i>Aspezeza na Boca</i>	<i>Irritação na Garganta</i>
2	A6.1	6,0	6,5	2,4	6,9	6,8	2,5	4,4	4,6	4,4	2,6	4,6	4,0
3	A6.2	7,4	5,1	1,2	1,5	3,0	5,9	6,8	6,5	2,8	2,5	0,8	0,7
4	A6.3	7,8	0,9	0,6	8,1	6,2	6,0	4,5	6,9	7,1	0,5	0,7	0,9
5	B9.1	3,8	3,2	2,3	4,3	3,2	2,2	4,2	4,2	4,1	3,7	4,9	4,9
6	B9.2	5,1	3,1	3,1	3,4	2,5	5,0	2,2	2,2	2,6	2,2	1,6	4,6
7	B9.3	7,7	6,2	0,1	7,2	0,8	8,7	5,4	6,5	7,8	2,1	2,0	1,9
8	C76.1	5,6	4,4	2,0	5,6	6,5	1,1	5,9	5,9	2,3	2,2	2,1	2,0
9	C76.2	6,5	2,5	2,2	2,3	3,2	1,3	8,7	8,7	2,1	1,0	0,7	0,2
10	C76.3	8,6	2,6	2,8	8,5	4,7	6,3	6,7	6,5	7,0	2,2	1,4	0,7
11	C10.1	7,2	6,8	2,2	5,6	6,1	3,4	4,5	4,4	4,6	5,2	5,9	5,9
12	C10.2	5,5	6,8	2,2	3,7	5,3	4,4	4,5	4,5	3,5	4,2	1,4	3,4
13	C10.3	7,2	0,4	0,1	6,8	5,1	2,9	5,6	5,3	6,8	1,6	2,9	1,8
14	B89.1	5,6	5,4	1,6	6,5	5,5	2,3	5,5	5,5	3,5	5,0	5,3	6,5
15	B89.2	5,0	1,2	5,2	5,0	1,9	5,0	1,3	0,7	2,7	1,7	1,8	1,6
16	B89.3	7,2	2,5	0,1	6,5	5,7	6,8	6,4	5,7	7,2	1,4	1,4	0,7
17	B65.1	5,8	5,0	2,3	6,0	6,3	2,4	4,9	4,7	3,4	2,1	3,3	3,2
18	B65.2	1,7	1,8	0,0	1,7	1,8	4,3	2,0	2,0	4,4	3,0	3,0	1,6
19	B65.3	7,3	2,4	0,1	8,4	8,3	0,2	7,7	8,3	5,4	0,1	0,2	0,1
20	B26.1	7,2	7,1	1,3	7,0	6,5	7,2	7,5	7,5	7,5	7,7	7,6	6,3
21	B26.2	7,0	6,7	6,9	9,0	7,9	1,0	8,1	8,4	5,7	1,5	1,5	1,5
22	B26.3	8,0	7,4	1,5	6,8	5,0	6,6	5,1	6,6	7,1	2,9	3,7	1,6
23	B74.1	8,5	8,6	0,8	5,5	1,6	6,3	4,5	2,7	7,5	7,5	7,5	8,2
24	B74.2	7,1	7,0	2,4	5,1	3,7	8,8	3,6	3,9	3,5	8,6	6,5	7,2
25	B74.3	6,0	4,1	1,9	6,5	1,8	8,1	1,8	7,0	7,1	2,4	3,4	6,7
26	B75.1	8,4	8,5	2,7	8,1	7,5	5,1	7,5	7,5	5,5	6,8	4,4	4,1
27	B75.2	4,8	3,3	5,5	7,8	7,1	6,4	7,9	8,3	2,4	2,0	1,2	2,2
28	B75.3	3,9	7,8	6,5	7,6	6,6	3,2	3,3	7,4	6,7	2,1	1,6	1,2
29	E50.1	7,9	8,1	1,9	3,9	4,0	4,5	7,7	7,7	6,6	7,5	7,4	7,5
30	E50.2	5,6	3,3	6,5	4,6	1,0	6,6	3,1	6,5	6,8	6,5	5,9	6,1
31	E50.3	7,6	6,6	5,6	6,5	2,5	7,2	2,0	2,1	7,3	4,0	2,6	1,4



- grande quantidade de informações
- elevada dimensionalidade de dados
- não permite uma simples inspeção visual

estudar  
matrizes complexas:

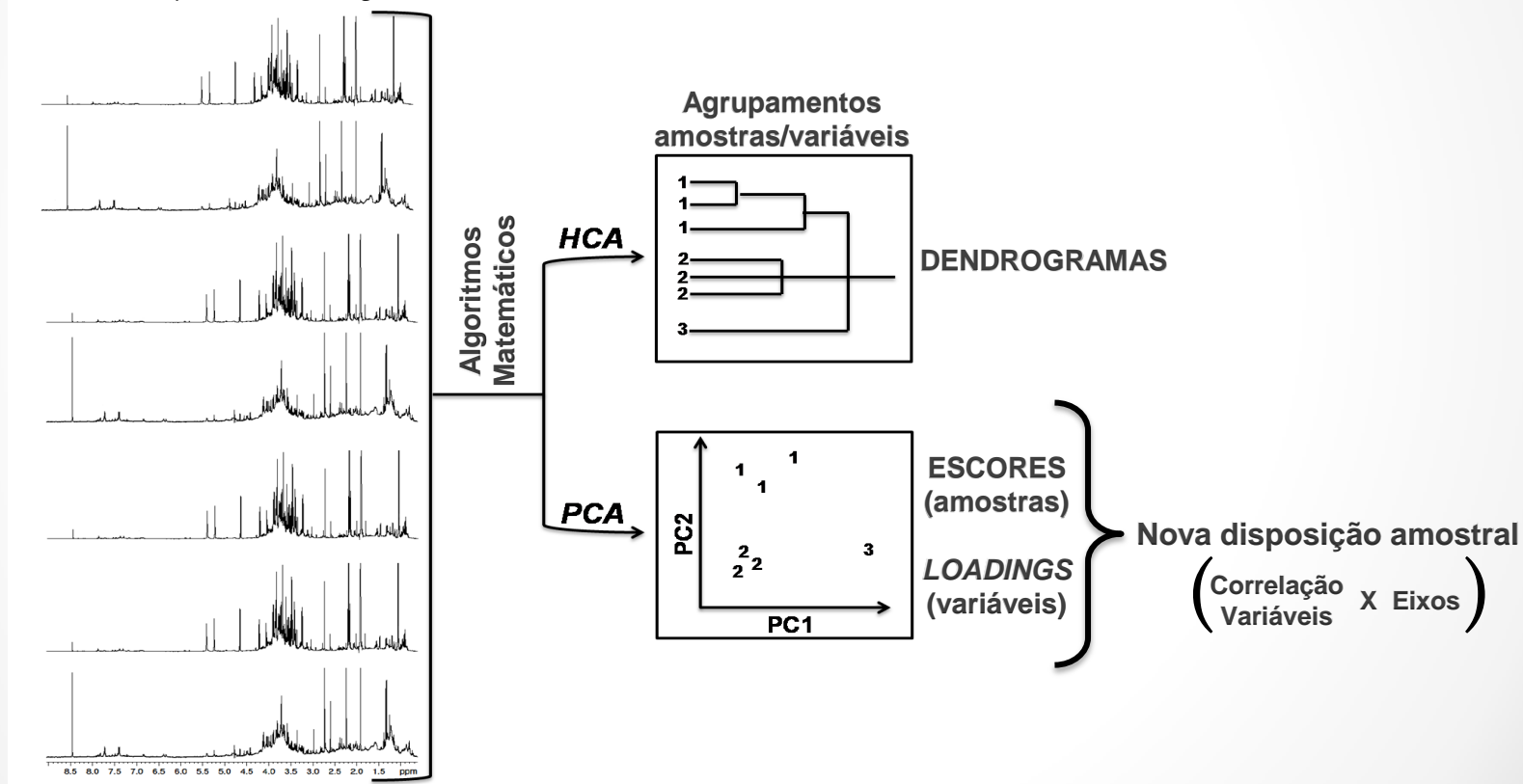
**Sensorial**  
**RMN**  
**HPLC-MS**  
**HPLC-UV**  
**GC-MS**  
**IV...**

*poucas variáveis fornecerão correlações existentes entre as amostras e as variáveis responsáveis por estas correlações.*

# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de alimentos

## MODELAGENS QUIMIOMÉTRICAS

Sensorial, Espectros, Cromatogramas...





# Exemplo da importância de pesquisas relacionadas à Quimiometria: ALIMENTOS



melhoria na qualidade



aumento na produção



melhoramento genético

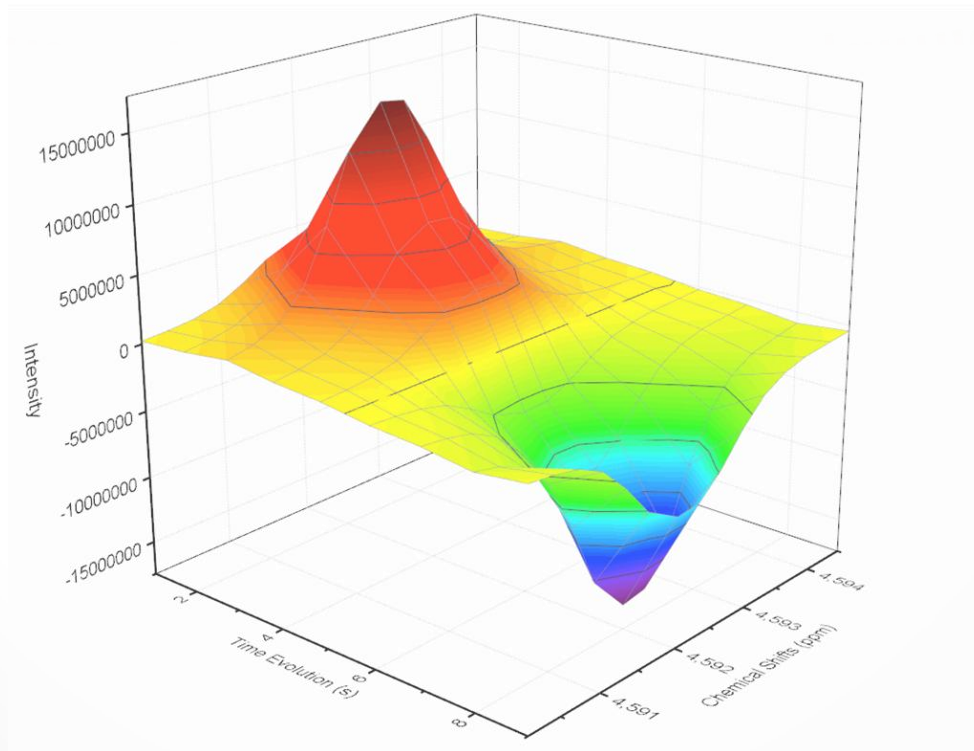
redução de custos



acesso à baixa renda



# Quais os cuidados anterior, durante e posterior as análises quimiométricas ?



# Quimiometria aplicada em análises de alimentos

## SUCESSO NO ESTUDO

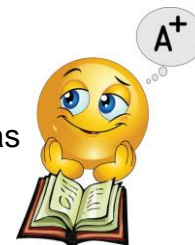


- Questão Biológica → o que será avaliados → ex.: ponto de partida para um estudo metabólico
- Planejamento experimental → estabelecer critérios para análise; fatores que podem influenciar; quais se correlacionam...
- Coleta e preparo apropriados das amostras
- Aquisição apropriada dos dados
- Processamentos apropriados destes dados
- Construção e extração dos dados para análises



## QUALIDADE NOS RESULTADOS (4 passos)

- Desenhar o processo → montar a análise com todas as etapas a serem concluídas
- Quais etapas são críticas → afetam mais os resultados
- Estabelecer os processos e equipamentos críticos → quais partes da análise são mais instáveis
- Conhecer a tolerância dos equipamentos e medições. Limites!



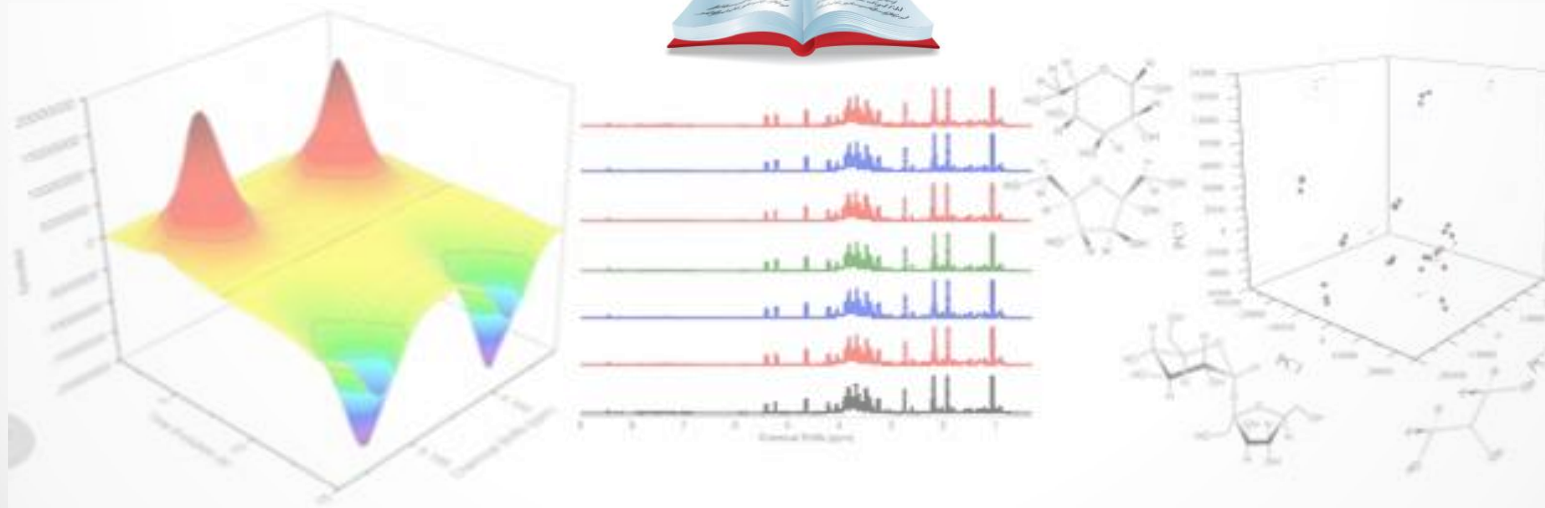
# Quimiometria aplicada em análises de alimentos



**Conhecimento → Saber o que está fazendo; não apenas aplicar!**

**Experiência em aplicações quimiométricas → Praticar bastante!**

**Noções de matemática/estatística**





# Coleta e preparo amostral



Análises dados (Sensorial, RMN, HPLC-UV, GC-MS...) por métodos quimiométricos:

- Profissional capacitado deverá ser consultado;
- Resultados representativos e reprodutíveis;
- Material deve representar o sistema a ser avaliado – Replicatas!

obs.:  $80 > 75$  ? Modelo Analítico  $\rightarrow \exists / 80 \pm 6 = 75 \pm 5 \rightarrow \exists$  erros!

- replicata de aquisição  $\rightarrow$  estabilidade do equipamento;
- replicata de preparo amostral, no mínimo!
- replicata de tratamentos; replicata biológica...  $\rightarrow$  mais indicado!

*Protocolos para coleta das amostras e identificação de variáveis importantes:*

## **Acondicionamento do Material:**

- Temperatura para conservação
- Embalagens resistentes
- Contaminantes das embalagens (ftalatos)
- Códigos para identificação
- Origem da amostra
- Data da Coleta

Anal. Chem. 2009, 81, 2884–2894

## **Coleta das Amostras $\rightarrow$ Represente o sistema:**

Regionalidade  
Sazonalidade  
Amadurecimento  
Tamanho e forma do fruto  
Horário da coleta

## **<sup>1</sup>H NMR, GC–EI-TOFMS, and Data Set Correlation for Fruit Metabolomics: Application to Spatial Metabolite Analysis in Melon**

**Benoit Biais,<sup>\*,†,§</sup> J. William Allwood,<sup>\*,||</sup> Catherine Deborde,<sup>†,§</sup> Yun Xu,<sup>||</sup> Mickael Maucourt,<sup>‡,§</sup> Bertrand Beauvoit,<sup>‡</sup> Warwick B. Dunn,<sup>‡</sup> Daniel Jacob,<sup>\*,V</sup> Royston Goodacre,<sup>||,‡</sup> Dominique Rolin,<sup>†,‡</sup> and Annick Moing<sup>†,§</sup>**

# Coleta e preparo amostral



Preparo das amostras:

- mínimo de tratamento e manipulação o material – de acordo com objetivo do estudo
- mais representativo possível do meio onde foi coletado

- analista deverá diminuir fontes de erros → modelo quimiométrico reproduzível/representativo
- as amostras devem ser preparadas, no mínimo, em triplicada.

## Preparo das Amostras:



*Quantidade?* → ↓tempo de análise, ↓erros, ↑resolução

*Solvente?* → completa solubilidade; volatilidade; não sobreposição de sinais

*Processamento amostral?* → Ultrassom; Centrifugação; Filtração; Liofilização; Correção de pH

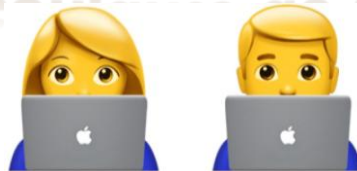
*Padrão interno ou externo?* → alinhar e avaliar a qualidade dos dados



## ETAPAS CONCLUÍDAS:

- *Planejamento experimental: bastante leitura e consulta com especialistas da área.*
- *Coleta, armazenamento e preparo amostral.*
- *Experimentos analíticos mais apropriados.*

**Como construir/trabalhar/extrair as informações de um conjunto de dados ?**

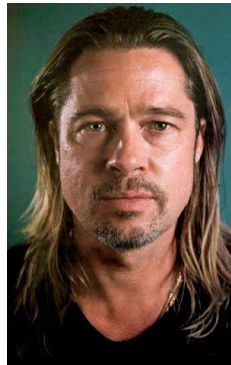


***Etapa trabalhosa, porém determina a qualidade dos resultados de uma modelagem quimiométrica!***

# Moldar os dados para construção de uma matriz numérica “uso de maquiagem”



pré-tratamento  
→



pré-tratamento  
→



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

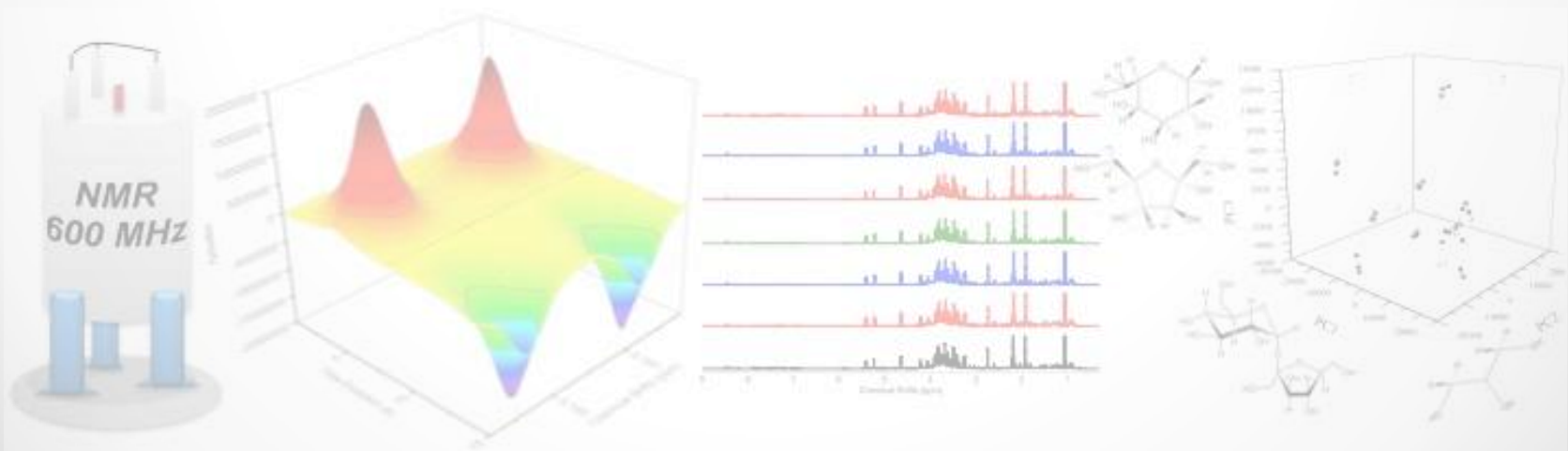
## Pré-tratamento dos dados “embelezamento”



- Dados vindos de análises sensoriais, espectros ou cromatogramas → Tendências na composição
- Pré-processamentos → redução ou remoção de informações irrelevantes (randômicas ou sistemáticas):

*Acondicionar a matriz numérica para uma análise mais eficiente*

- Remover informações irrelevantes que mascaram as de interesse
- Remover fontes de erros
- Remover informações importantes quando mal aplicados!



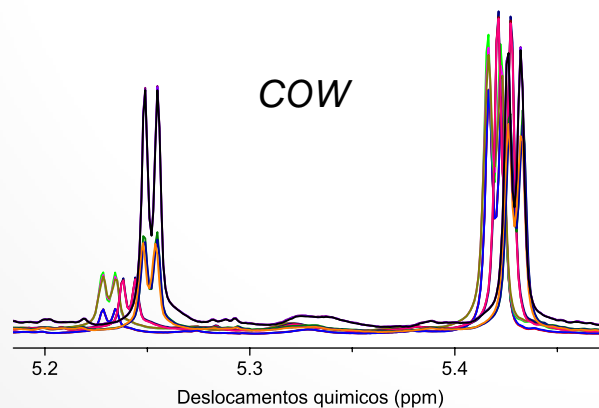
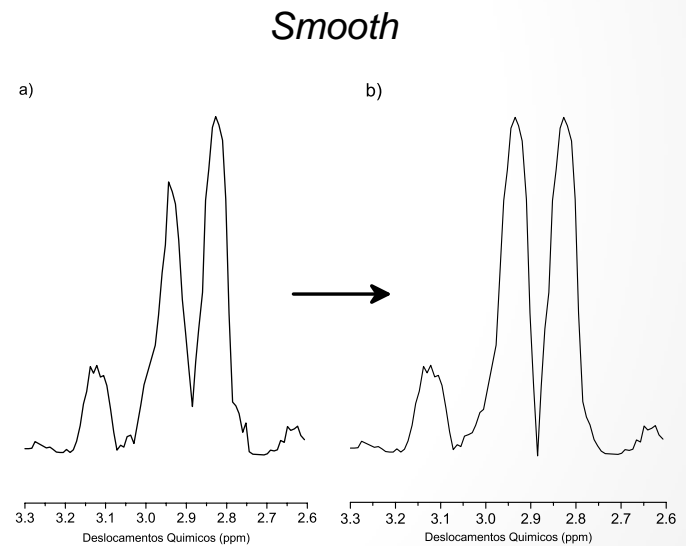
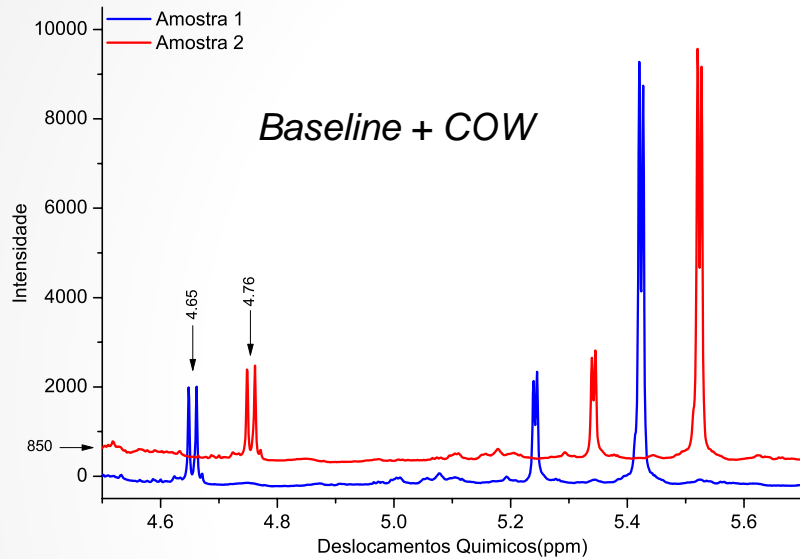
# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

## ***Principais pré-processamento sobre as variáveis:***

- Primeira derivada → melhorar a definição dos sinais sobrepostos
- Segunda derivada
- Alinhamento → COW (*Correlation Optimized Warping*)
- *Bucket Table*
- Logaritmo na base 10
- *Smooth* → filtros polinomiais: Savitzky-Golay; Gaussiano; Média...  
*remover informações aleatórias*
- Correção por espalhamento multiplicativo (MSC)
- *Standard Normal Variate* (SNV)
- Normalização → dividir cada variável por uma constante  
*remover informações sistemáticas*
- Correção da linha de base
- Ponderação → dar peso a umas amostras de interesse sobre outras  
*remover informações sistemáticas*
- Dentre muito outros...

# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

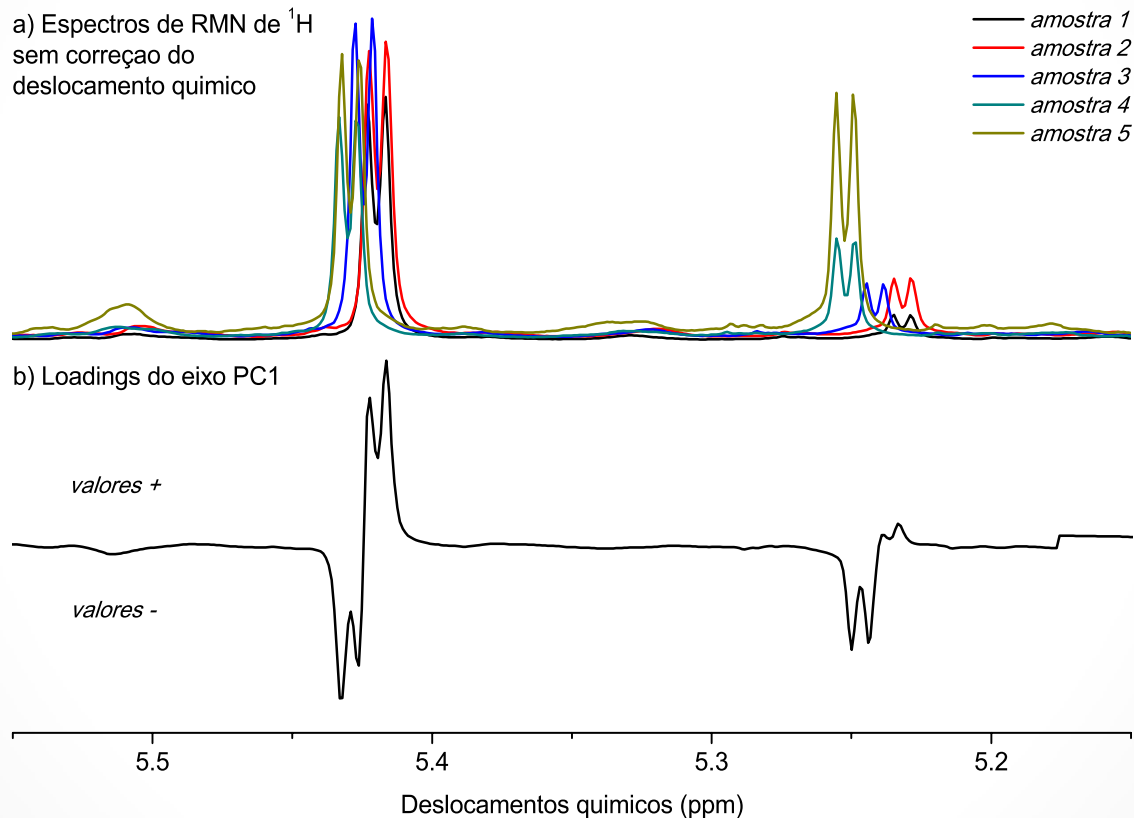
## Principais pré-processamento sobre as variáveis



# Principal problema do desalinhamento de sinais correspondentes

a) Espectros de RMN de  $^1\text{H}$  desalinhados (descalibrados);

b) Gráfico de *loadings* do eixo PC1 ilustrando o efeito negativo deste problema para quimiometria.

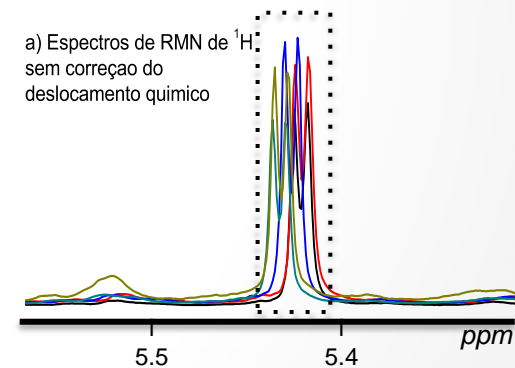
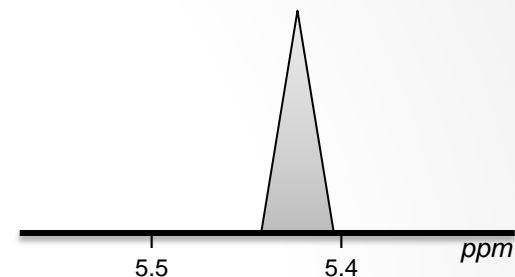
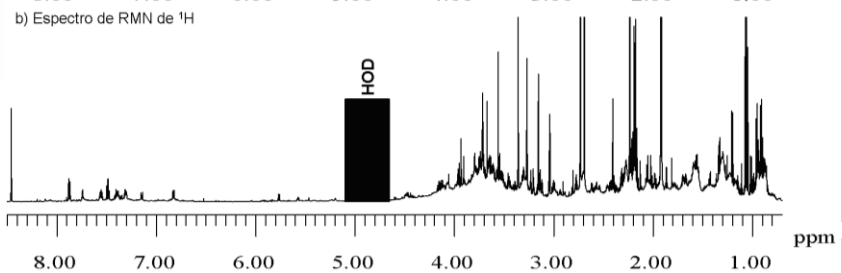
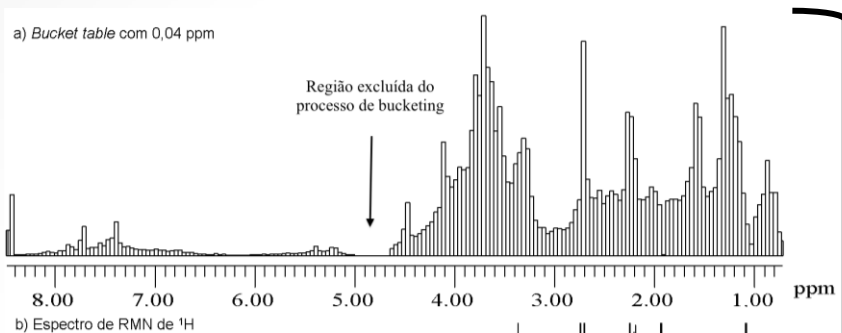




# Processo por *Bucketing*

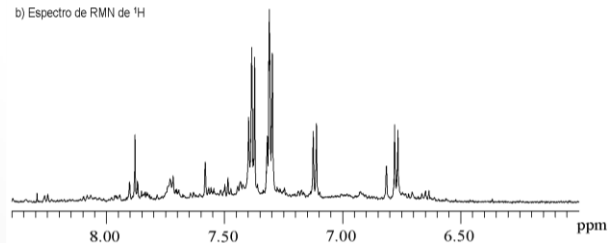
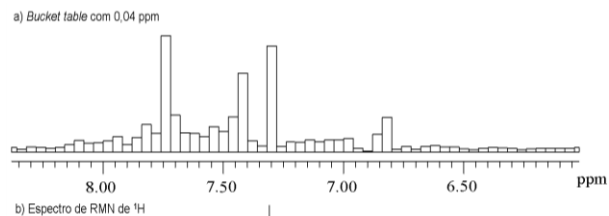
a) Resultado da aplicação do processo de *bucketing* (0,04 ppm) – Amix<sup>®</sup>

b) Espectro de RMN <sup>1</sup>H



Espectro Completo →

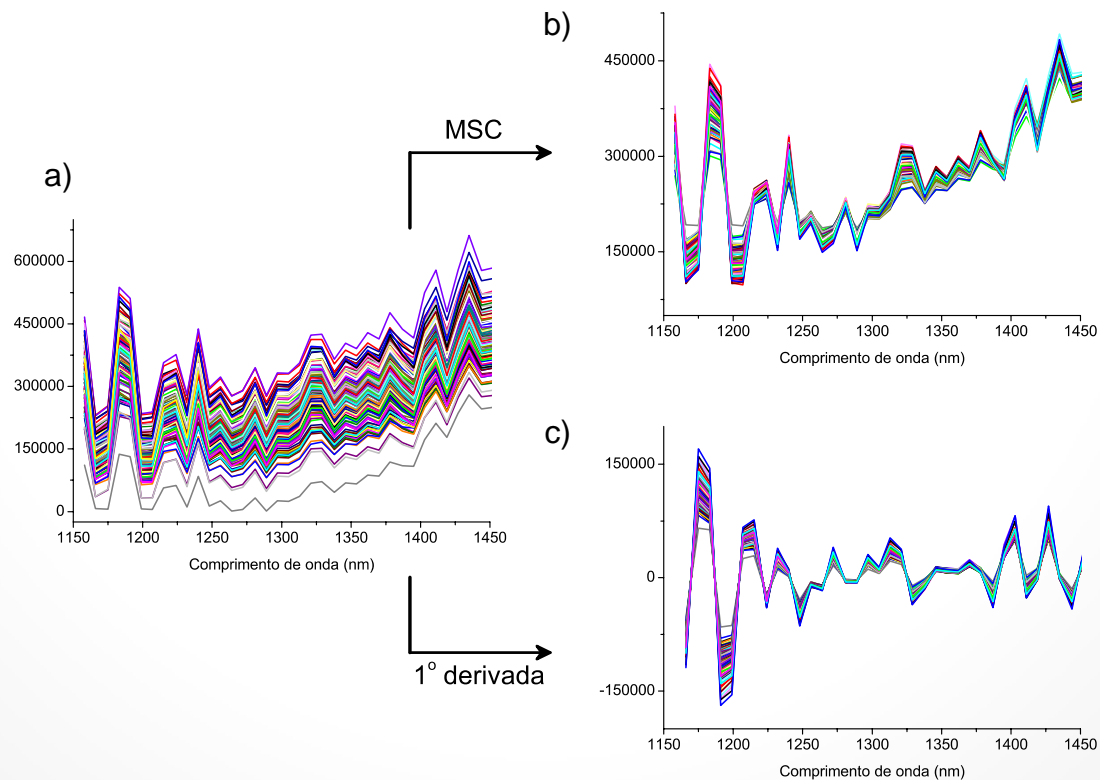
Região de <sup>1</sup>H Aromáticos →



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

## Principais pré-processamento sobre as variáveis

Espectros de IV antes (a) e após (b) o pré-processamento de correção por espalhamento multiplicativo (MSC)



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

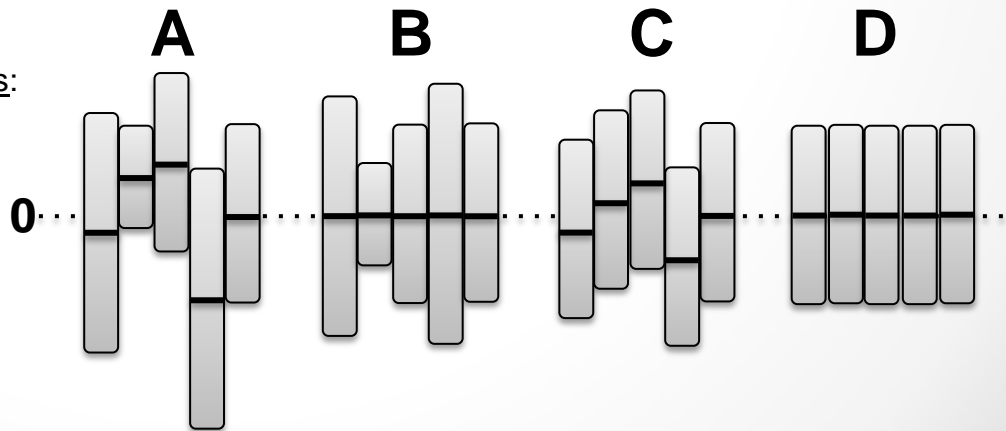
## Principais pré-processamento sobre as amostras

- Centralização em torno da média  
subtrai a média dos valores por cada valor → média zero
- Escalonamento da variância - padronização  
dividi a média de cada elemento de cada variável pelo desvio padrão, igualando a variância de cada variável
- Autoescalamento  
centralização dos dados na média + escalonamento  
média zero e desvio padrão 1

*obs.: Escalonamento da variância e autoescalamento darão o mesmo peso às variáveis*

Ilustração dos processamentos dos dados:

- (A) sem tratamento
- (B) após a centralização em torno da média
- (C) após escalonamento
- (D) autoescalamento





## ETAPAS CONCLUÍDAS:

- *Planejamento experimental: bastante leitura e consulta com especialistas da área.*
- *Coleta, armazenamento e preparo amostral mais apropriados.*
- *Experimentos analíticos mais indicados.*
- *Pré-tratamentos (maquiagem) dos dados para que informações de interesse (qualidade) sejam destacadas, e imperfeições, erros e informações não relevantes sejam minimizadas ou eliminadas.*

**Dados tratados e  
pré-processados!**



**Análises exploratórias  
HCA e PCA 🧐**

**Análises classificatórias  
PLS, PLS-DA, SIMCA e KNN**



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

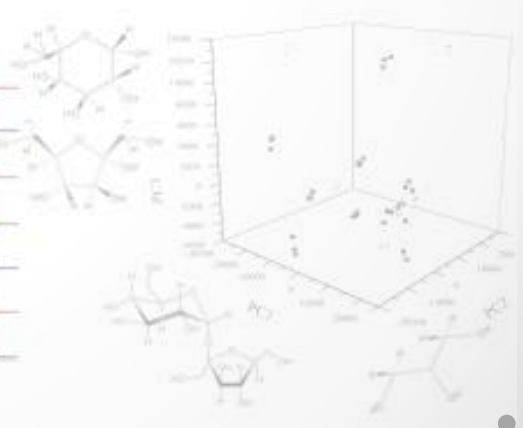
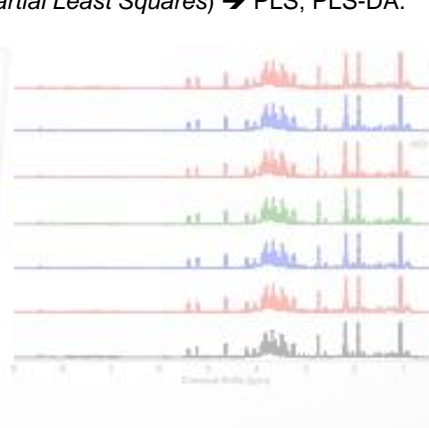
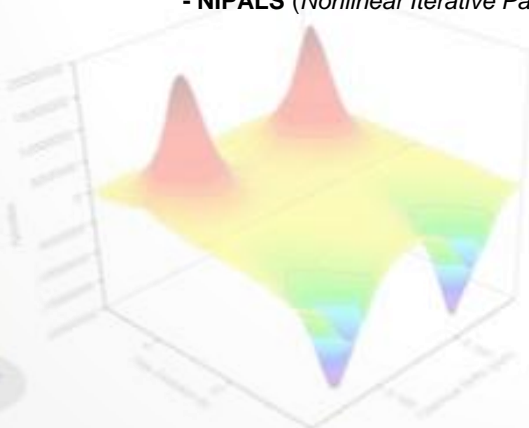
## HCA, PCA, PLS, PLS-DA, SIMCA, KNN



- **Similaridade entre as amostras – agrupamentos**
- **Quais são as variáveis responsáveis**
- **Quais destas variáveis estão correlacionadas (a favor ou contra) ou independentes**
- **HCA e PCA → exploratórias, não supervisionadas**
- **PLS-DA, PLS, SIMCA e KNN → calibração e regressão, supervisionadas**

### Algoritmos mais utilizados:

- **SVD (Singular Value Decomposition) → projeção dos dados em um novo espaço vetorial.**
- **NIPALS (Nonlinear Iterative Partial Least Squares) → PLS; PLS-DA.**

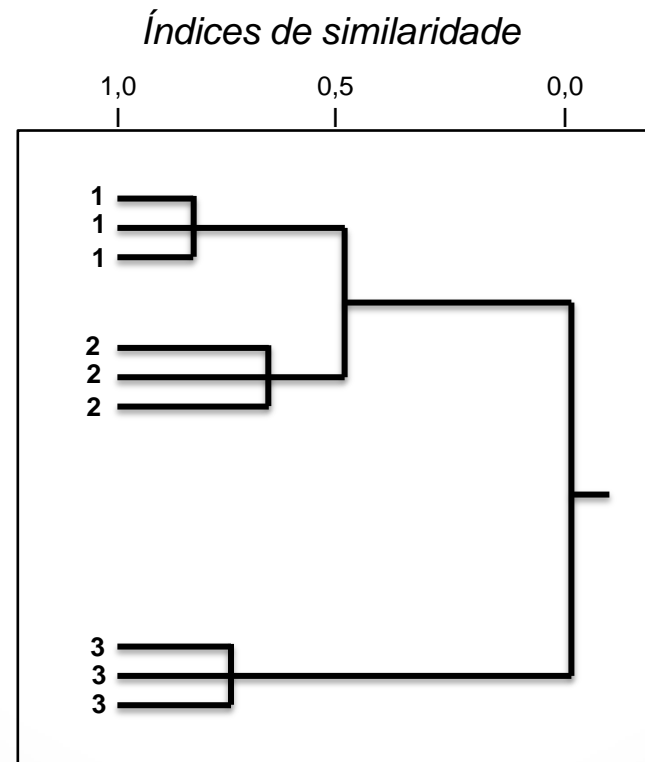


# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

## Análise de Agrupamentos por HCA



### Dendrogramas



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

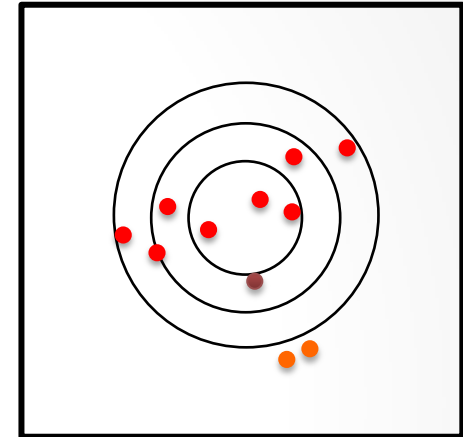
Medidas para estabelecer o conceito de distância entre duas amostras:

- *coeficiente de correlação de Pearson*
- **distância Euclidiana**
- *distância de Manhattan*
- *distância de Mahalanobis* → *considera a correlação entre variáveis*

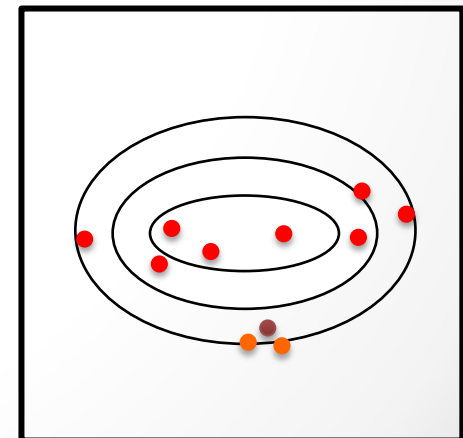
Tipos de Agrupamento:

- *K-means*
- *K-medians*
- *Hierarchical Single-linkage*
- **Hierarchical Complete-linkage**
- *Hierarchical Average-linkage*
- *Hierarchical Median-linkage*
- **Ward's method**

*Distância Euclidiana*



*Distância de Mahalanobis*





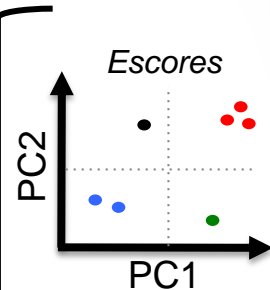
# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

## Análise de Componentes Principais – PCA = Autovalores + Autovetores

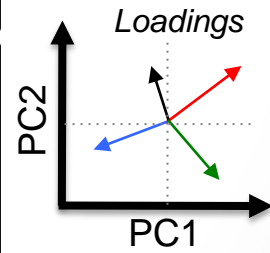
Long Name	A(X)	Q(Y)	R(Y)	S(Y)	T(Y)	U(Y)	V(Y)	W(Y)	X(Y)	Y(Y)	Z(Y)	AA(Y)	AB(Y)	AC(Y)	AD(Y)	AE(Y)
Amostras	0.5 ppm	1.0 ppm	1.5 ppm	2.0 ppm	2.5 ppm	3.0 ppm	3.5 ppm	4.0 ppm	4.5 ppm	5.0 ppm	5.5 ppm	6.0 ppm	6.5 ppm	7.0 ppm	7.5 ppm	
1 Amostra 1	-6.12960	-3.93400	-3.74400	-16.0422	-8.56630	-7.90010	-11.4652	-11.9444	-10.9445	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-11.4652	
2 Amostra 2	-5.60530	-4.88580	-2.85890	-16.6824	-6.89770	-8.29170	-8.78110	-8.75380	-13.2263	-6.00780	-11.8920	-10.2839	-32.6477	-7.23870	-8.78110	
3 Amostra 3	-4.35330	-3.18580	-2.57540	-15.3622	-8.07880	-8.92000	-6.13200	-14.0918	-15.3337	-9.03270	-14.0169	-13.1059	-23.5460	-11.0579	-6.13200	
4 Amostra 4	-4.00100	-4.82000	-4.60380	-11.6526	-8.19190	-9.88750	-17.3969	-11.7166	-13.4233	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-17.3969	
5 Amostra 5	-5.41880	-7.11660	-6.36720	-10.7203	-12.6022	-8.01010	-10.8263	-10.5608	-14.8458	-8.03990	-8.80460	-12.8046	-25.8944	-7.10760	-10.8263	
6 Amostra 6	-3.78450	-7.77060	-5.74360	-12.8627	-7.24940	-10.0148	-10.3575	-16.4900	-18.1446	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-10.3575	
7 Amostra 7	-4.32290	-5.20820	-5.12730	-12.0741	-3.94300	-11.5984	-9.62000	-16.2788	-20.5803	-10.0780	-11.8920	-10.2839	-32.6477	-7.23870	-9.63900	
8 Amostra 8	-5.46260	-7.09780	-1.25950	-13.7586	-11.6288	-5.57110	-11.0355	-11.8333	-9.3273	-9.12770	-14.0169	-13.1059	-23.5460	-11.0579	-12.8154	
9 Amostra 9	-4.73020	-6.87950	-2.54220	-10.9181	-11.4177	-10.5614	-11.2495	-9.28660	0.7388	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-13.2492	
10 Amostra 10	-3.76680	-8.45230	-5.43130	-15.9566	-11.5827	-11.1877	-11.9411	-9.16610	-15.5852	-10.0398	-8.80460	-12.8046	-25.8944	-7.10760	-12.9457	
11 Amostra 11	-3.77910	-6.18320	-6.25970	-17.3355	-14.1177	-11.4752	-4.35330	-15.6672	-16.9064	-8.48370	-12.6168	-16.5185	-25.6425	-12.9705	-4.35330	
12 Amostra 12	-5.60250	-7.63830	-2.63660	-15.9857	-9.06870	-7.65940	-11.4723	-15.8113	-11.0772	-14.4538	-17.6184	-11.5801	-21.4435	-21.0691	-11.4723	
13 Amostra 13	-6.63250	-7.20500	-5.19560	-16.3599	-8.38810	-8.44650	-5.97750	-14.8479	-17.1128	-6.06940	-12.6701	-12.7060	-16.1058	-14.1553	-5.97750	
14 Amostra 14	-7.01780	-6.84440	-5.55070	-16.0415	-14.5741	-13.0748	-6.65650	-11.5459	-11.4694	-7.92760	-4.40530	-13.3188	-17.2564	-28.2828	-6.65650	
15 Amostra 15	-6.53900	-5.97200	-3.14880	-20.4323	-13.2241	-17.7318	-8.71800	-13.8167	-13.9193	-11.7536	-8.54060	-8.15650	-11.6307	-21.3215	-8.71800	
16 Amostra 16	-4.75900	-5.45510	-5.19170	-15.1259	-7.81480	-13.8332	-16.0618	-11.9624	-12.2261	-12.4957	-11.0392	-11.9333	-15.8174	-15.9668	-16.0618	
17 Amostra 17	-3.95600	-5.07640	-4.16330	-16.2087	-10.2615	-10.0024	-13.6478	-15.0621	-14.3768	-10.5041	-10.4618	-8.31990	-28.5540	-19.7976	-13.6478	
18 Amostra 18	-2.55840	-5.77230	-6.15910	-13.9531	-13.6938	-9.01310	-8.67210	-15.8102	-14.7981	-8.71770	-11.9790	-13.4979	-23.8291	-22.8878	-8.67210	
19 Amostra 19	-4.83510	-7.22520	-2.81990	-15.1478	-9.16250	-8.04410	-9.37930	-15.1017	-15.0568	-11.1856	-10.7040	-11.8671	-13.7654	-17.6918	-9.37930	

**Matriz X**

**Algoritmos**



→ AMOSTRAS autovalores



→ VARIÁVEIS autovetores

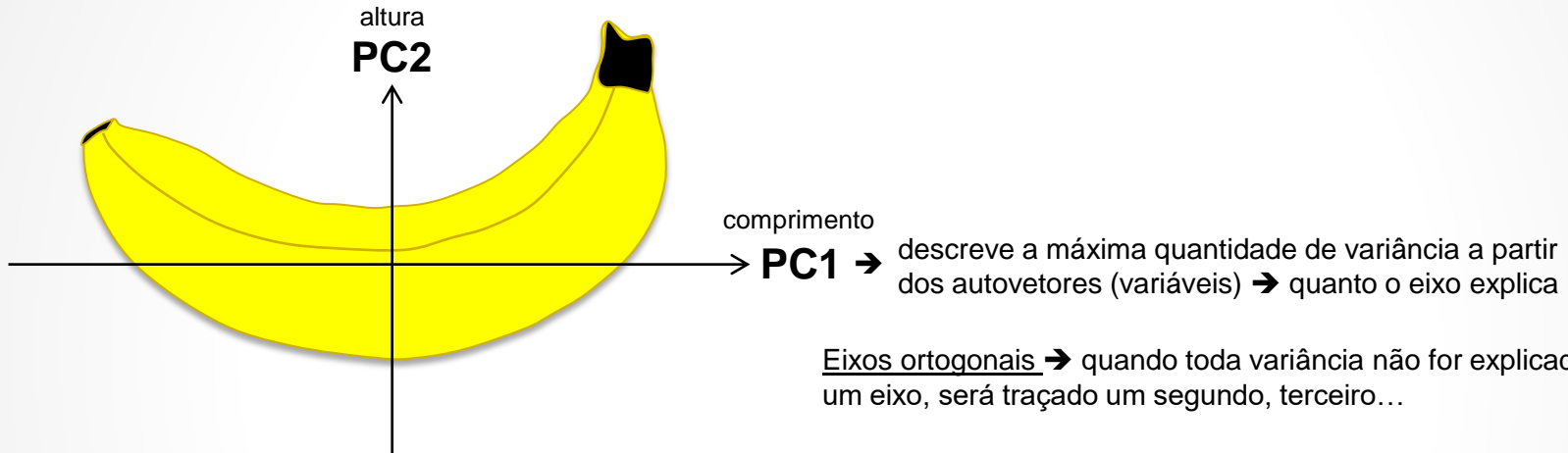
**Matriz de Erros**

→ Variância não explicada  
→ Resíduo da amostra

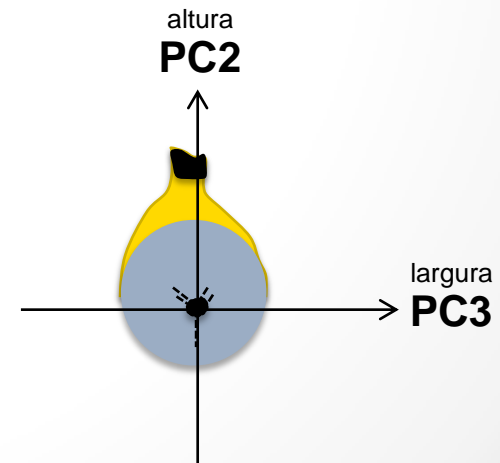
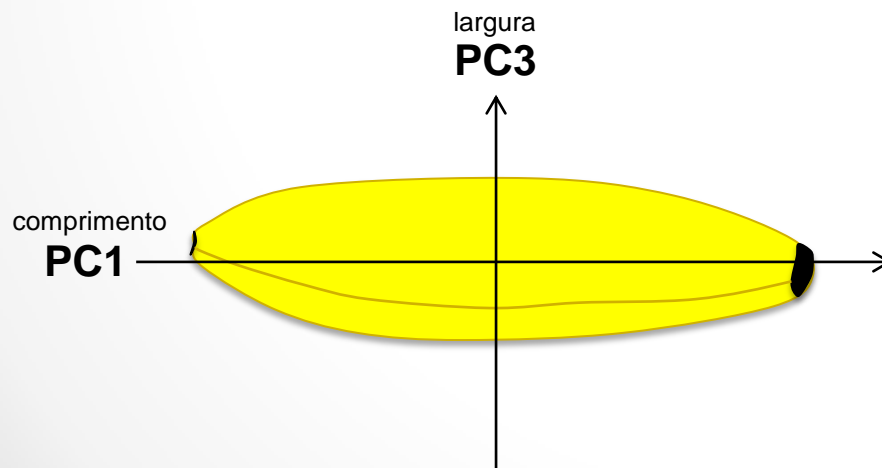
# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes alimentícias

## Análise de Componentes Principais – PCA

*Princípio do funcionamento*



Eixos ortogonais → quando toda variância não for explicada em um eixo, será traçado um segundo, terceiro...



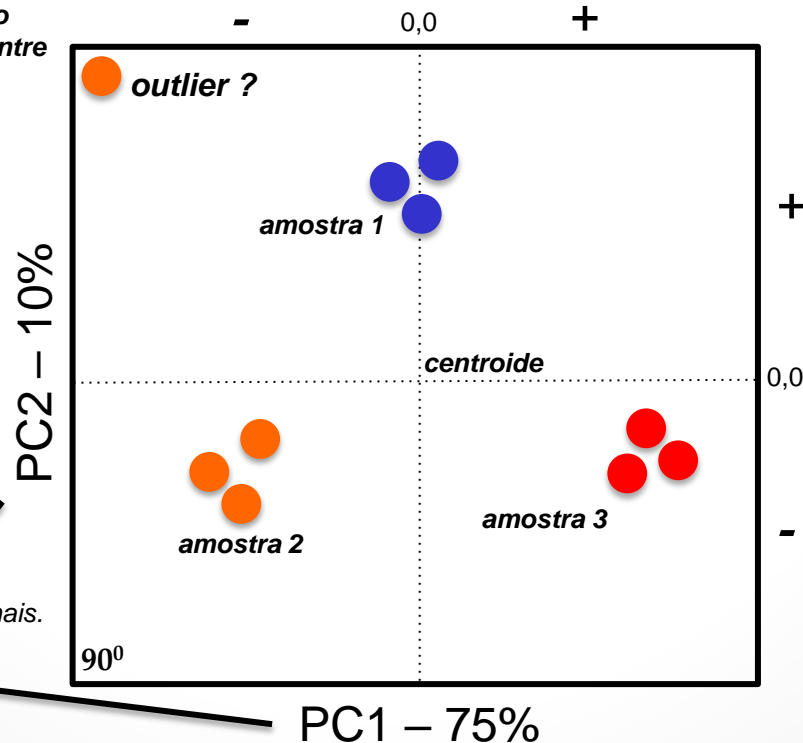
$$PC1 + PC2 + PC3 = 100\%$$

# Análises Estatísticas Multivariadas dos Dados e suas Aplicações nas Análises de Matrizes Complexas

## PCA → GRÁFICO DE ESCORES - AMOSTRAS

- variância dos escores → proporcional ao autovalor para cada eixo
- autovalor → valores que representam a contribuição de cada componente na explicação da variância total dos dados.
- olhar projeções dos dados nos eixos
- residuais das amostras → distância do ponto ao eixo → erros da modelagem

Covariância → grau da correlação linear de cada variância → zero entre eixos

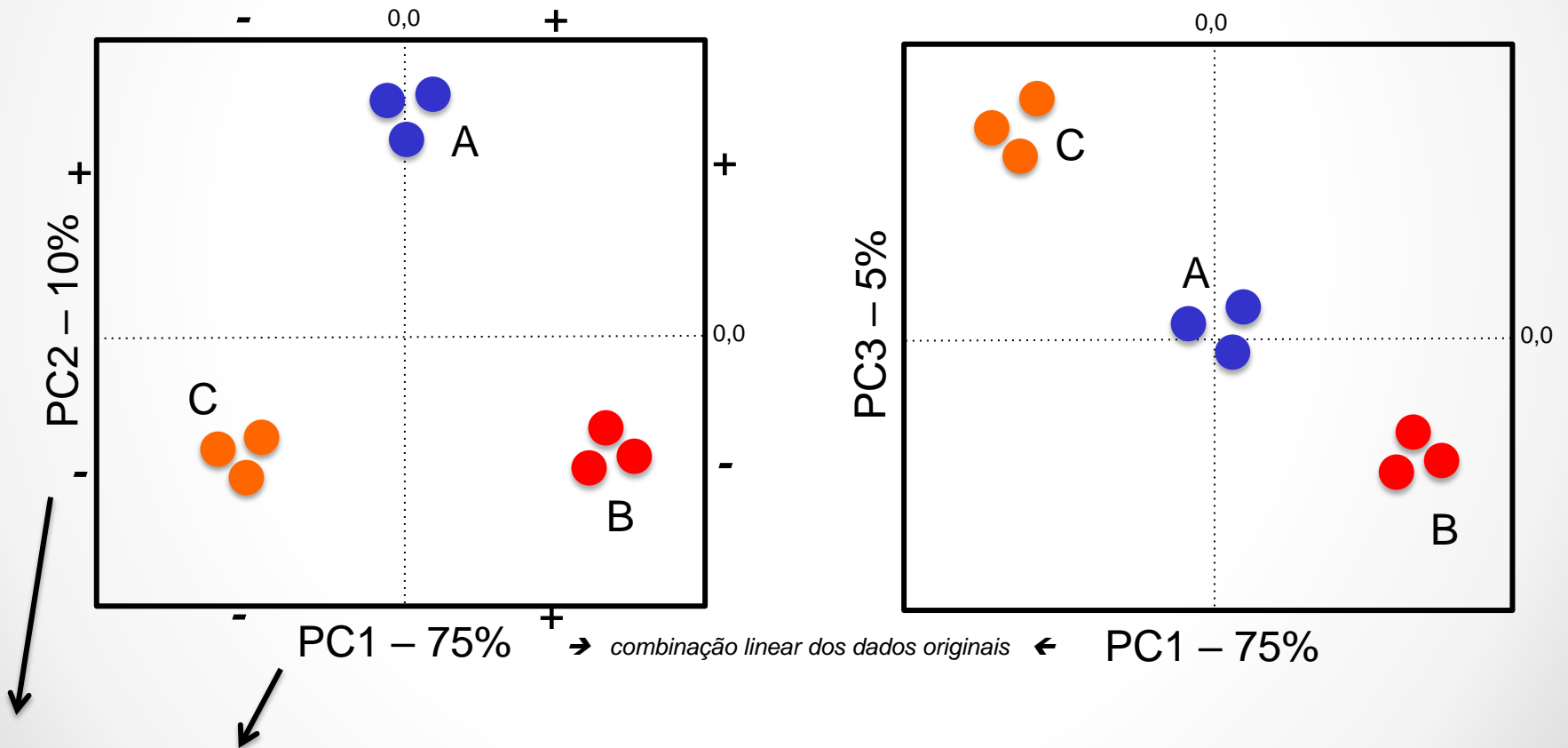


RESÍDUO TOTAL  
VARIÂNCIA TOTAL  
↓  
Quão bem foi a adaptação do modelo aos dados

Combinção linear das variáveis originais.  
Melhor ajustados de acordo com a nuvem de pontos.

# Análises Estatísticas Multivariadas dos Dados e suas Aplicações nas Análises de Matrizes Complexas

## PCA → GRÁFICO DE ESCORES - AMOSTRAS

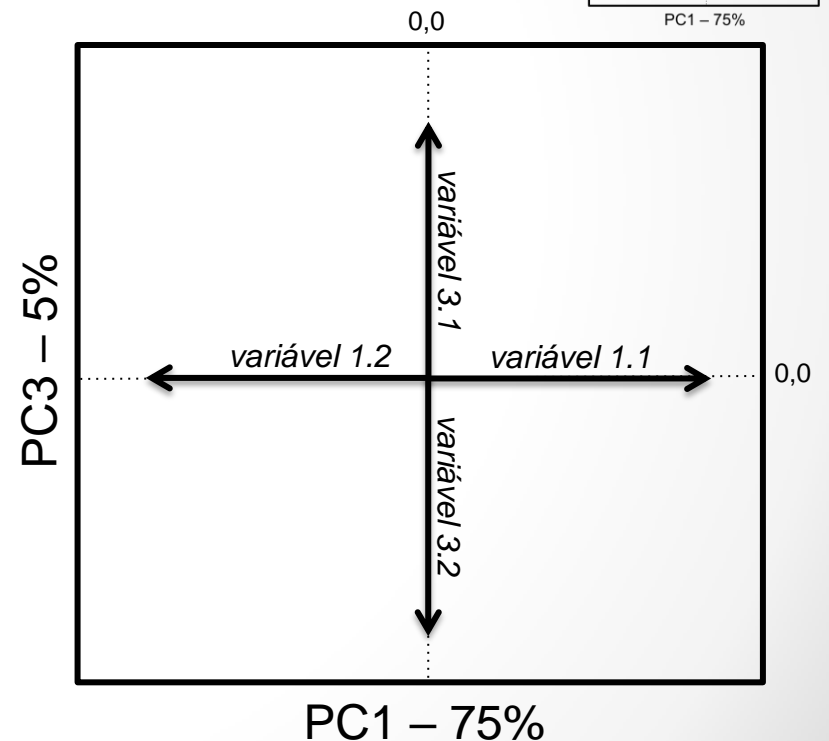
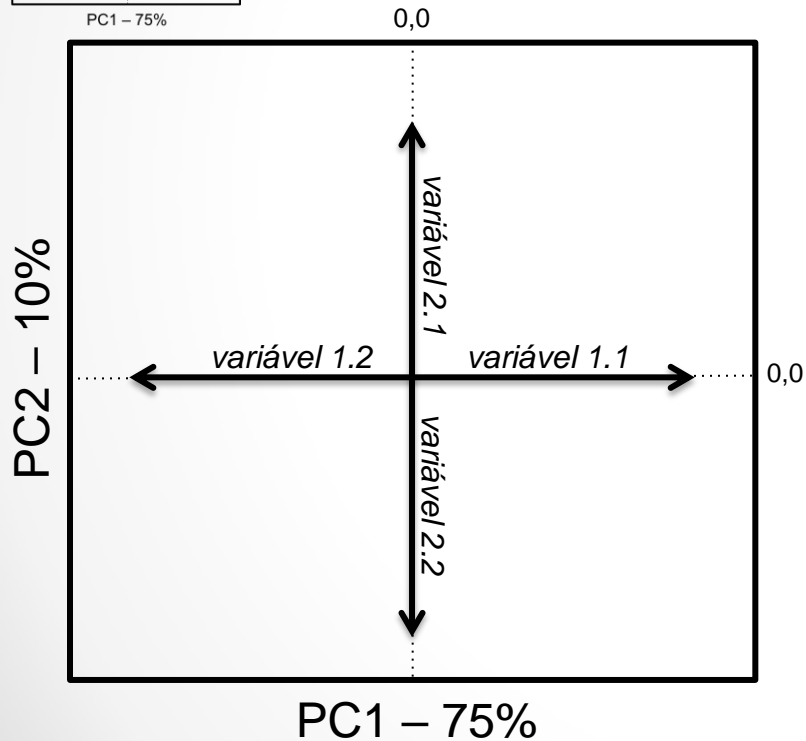
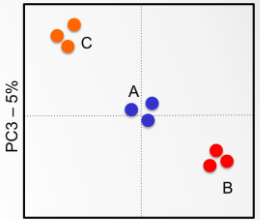
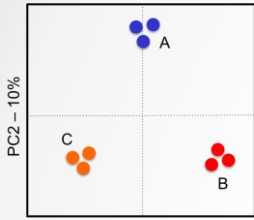


Melhor ajustados de acordo com a nuvem de pontos

# Análises Estatísticas Multivariadas dos Dados e suas Aplicações nas Análises de Matrizes Complexas

## PCA → GRÁFICO DE PESOS (LOADINGS) – VARIÁVEIS

AUTOVETORES → VALORES QUE REPRESENTAM O PESO DE CADA VARIÁVEL SOBRE A COMPONENTE



Correlação → Covariância entre as variáveis normalizadas → tirar efeito de altas concentrações  
Variáveis com variância 1

- residuais das variáveis → distância do ponto ao eixo



**Exemplo real:**  
**Análise quimiométrica exploratória**  
**(não supervisionada)**



**Teorias básicas:**

- *Quanto de variância acumulada é o suficiente?*
- *Remover de outlier?*
- *O que e como informar os resultados?*



## Exemplos reais: Análises quimiométricas exploratórias (não supervisionadas)



- Amostragem; **falar dos artigos de processamentos**
- Alterações climáticas;
- Condições de cultivo e Sazonalidade;
- Regionalidade;
- Fatores externos: doenças, predadores;
- Bancos de germoplasma;
- Técnicas corroborativas;
- *Heat map.*



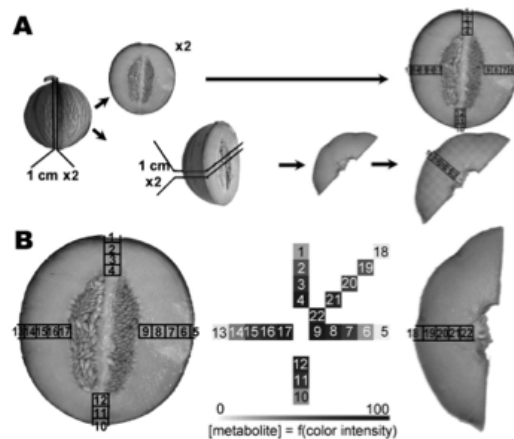
# Estudo quimiométrico relacionado a amostragem

*Anal. Chem.* 2009, 81, 2884–2894

## $^1\text{H}$ NMR, GC–EI-TOFMS, and Data Set Correlation for Fruit Metabolomics: Application to Spatial Metabolite Analysis in Melon

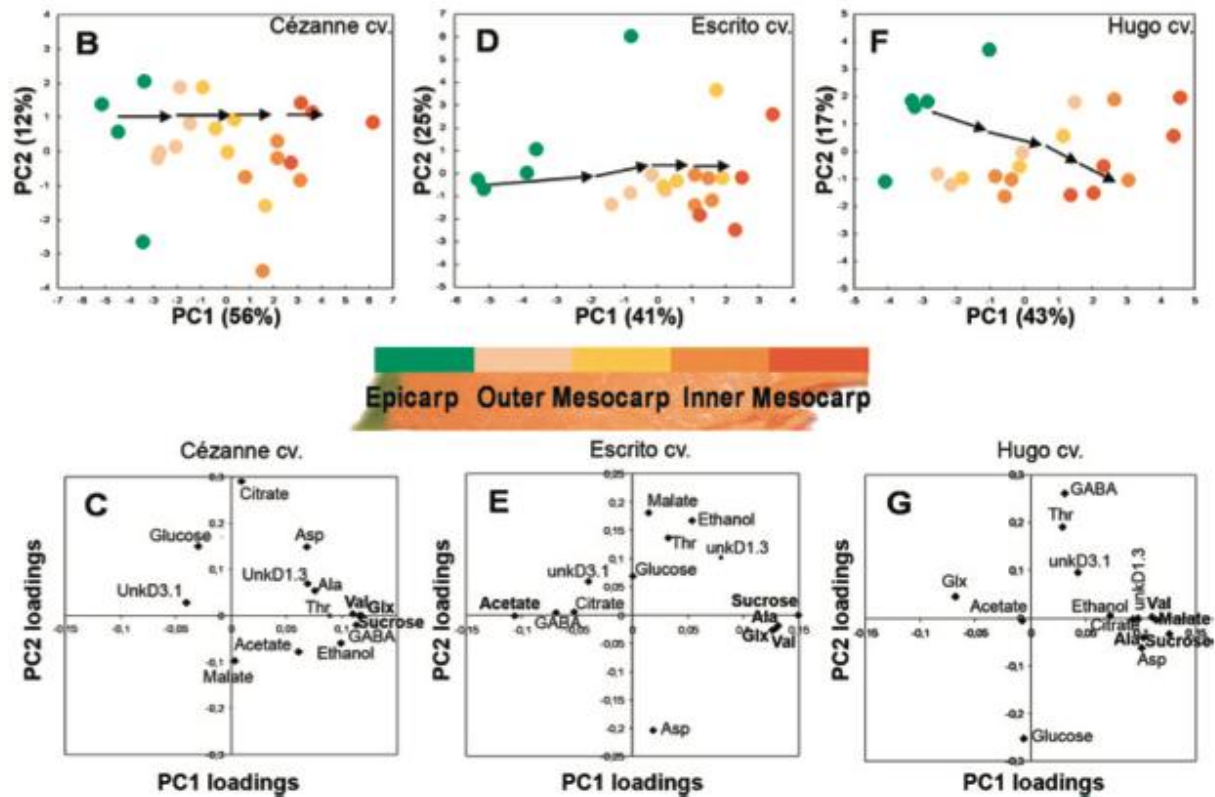
**Benoit Biais**,<sup>\*,†,§</sup> **J. William Allwood**,<sup>\*,||</sup> **Catherine Deborde**,<sup>†,§</sup> **Yun Xu**,<sup>||</sup> **Mickael Maucourt**,<sup>‡,§</sup> **Bertrand Beauvoit**,<sup>‡</sup> **Warwick B. Dunn**,<sup>⊥</sup> **Daniel Jacob**,<sup>#,∇</sup> **Royston Goodacre**,<sup>||,⊥</sup> **Dominique Rolin**,<sup>†,‡</sup> and **Annick Moing**<sup>†,§</sup>

INRA, UMR619 Fruit Biology, BP81, F-33140 Villenave d'Ornon, France, Université de Bordeaux, UMR619 Fruit Biology, BP81, F-33140 Villenave d'Ornon, France, Plateforme Métabolome-Fluxome–Génomique Fonctionnelle Bordeaux, IFR 103 BVI, BP81, F-33140 Villenave d'Ornon, France, School of Chemistry, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K., Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K., Centre de Bioinformatique de Bordeaux–Génomique Fonctionnelle Bordeaux, Université de Bordeaux, 146 rue Léo Saignat, F-33076 Bordeaux, France, and INRA, UMA1251, IFR103 BVI, BP 81, F-33140 Villenave d'Ornon, France



**Figure 1.** Melon sample collection and data representation. (A) Preparation of melon slices and localization of collected flesh pieces. (B) Schematic representation of concentration gradients depending on the location in the slice.

# Estudo quimiométrico relacionado a amostragem



**Figure 2.** <sup>1</sup>H NMR spatial metabolite profiling of melon slices. (A) Representative parts of <sup>1</sup>H NMR spectra of each cultivar of melon juice in 70% methanol-*d*<sub>4</sub>. Resonances are annotated according to the chemical shifts in Supporting Information Table S1. Glx corresponds to glutamine plus glutamic acid resonances. (B–G) Principal component analysis of 14 metabolites (absolute concentration) obtained from the <sup>1</sup>H NMR spectra of 22 melon sections for each of the three cultivars (Cézanne, Escrito, and Hugo). The metabolites were identified and quantified using <sup>1</sup>H NMR spectra of melon juices. Mean-centered data scaled to unit variance were used for PCA. (B, D, F) Scores plots. (C, E, G) Loading plots. Metabolites with PC1 loading values greater than 0.1 are indicated with a bold font. Acids are noted with their conjugate base name for readability purpose.

# Estudo quimiométrico relacionado a alterações climáticas



AQUECIMENTO GLOBAL



Análise da resposta metabólica  
(*non-targeted profiling*)

{ *Phyllanthus amarus*  
*Phyllanthus niruri* }

Escassez hídrica

[www.nature.com/scientificreports](http://www.nature.com/scientificreports)

## SCIENTIFIC REPORTS

OPEN

### Physiological changes for drought resistance in different species of *Phyllanthus*

Received: 11 April 2018  
Accepted: 24 September 2018  
Published online: 11 October 2018

Elenilson G. Alves Filho<sup>1</sup>, Luiza N. Braga<sup>2</sup>, Lorena Mara A. Silva<sup>1</sup>, Fábio R. Miranda<sup>1</sup>, Ebenézer O. Silva<sup>1</sup>, Kirley M. Canuto<sup>1</sup>, Maria Raquel Miranda<sup>3</sup>, Edy S. de Brito<sup>1</sup> & Guilherme J. Zocolo<sup>1</sup>

The *Phyllanthus* genus is widely distributed in tropical and subtropical areas of the world and present several pharmacological applications. Drought is a restrictive factor for crop development and production, and is becoming a severe problem in many regions of the world. The species *Phyllanthus amarus* and *Phyllanthus niruri* were subjected to drought stress for varying periods of time (0, 3, 5, 7, and 10 days), and afterwards, leaves were collected and evaluated for physiological and biochemical responses, such as oxidative stress markers and drought-associated defense mechanisms. Results show that *P. amarus* has an endogenously higher level of variables of the oxidative/antioxidant metabolism, and *P. niruri* presents the most significant changes in those variables when compared to control and stressed plants. For both *Phyllanthus* species, drought stress induces higher levels of organic acids such as malic, succinic, and citric acids, and amino acids such as proline, GABA, alanine, and valine. Moreover, *P. niruri* plants respond with greater glucose and corilagin contents. Therefore, considering the evaluated metabolic changes, *P. amarus* is better adapted to drought-stress, while *P. niruri* presents an acclimation strategy that increases the corilagin levels induced by short-term drought stress.

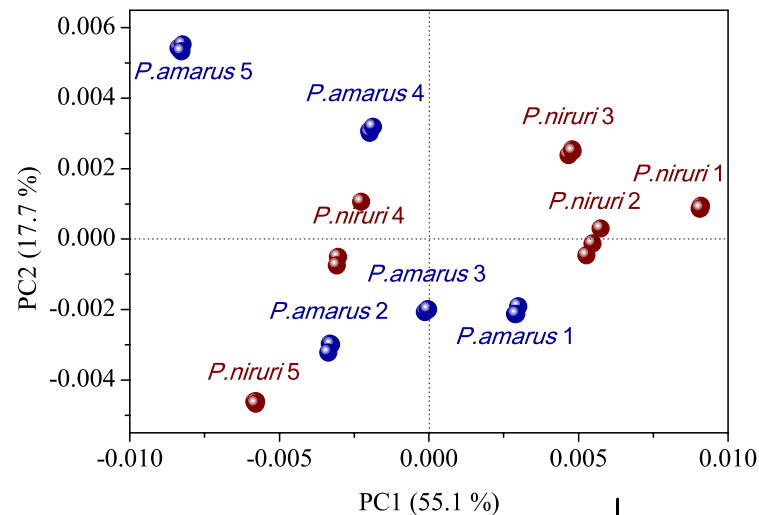
# Estudo quimiométrico relacionado a alterações climáticas

## PCA

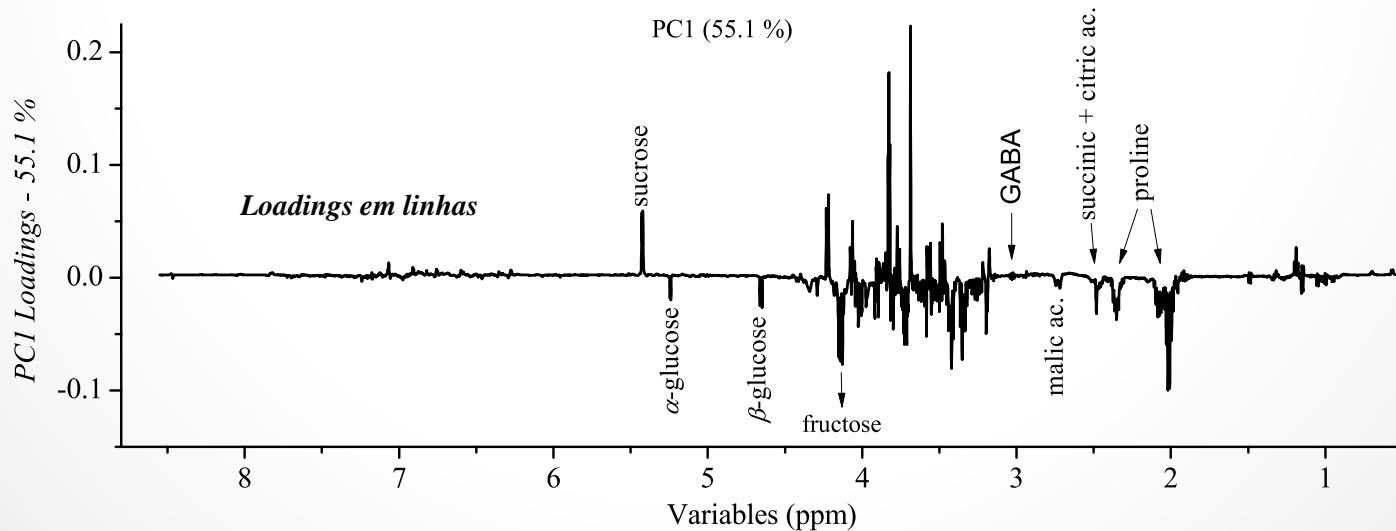
*P. amarus* → azul

*P. niruri* → vermelho

a) Scores



b) Loadings





# Estudo quimiométrico relacionado a condições de cultivo/sazonalidade



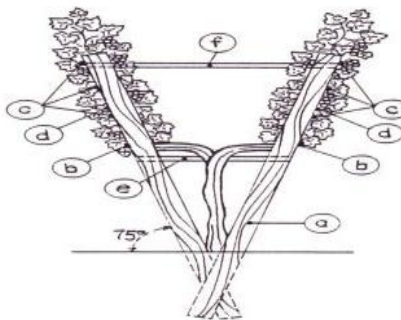
Análise da resposta metabólica  
(*non-targeted profiling*)

{ vinho tinto – *Syrah*  
vinho branco – *Chenin Blanc*

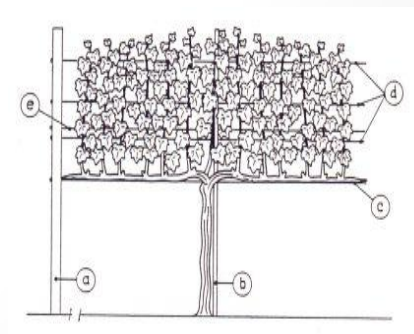
## Variáveis:

- Sazonalidade: janeiro/agosto
- Sistema de condução: lira/espaldeira
- Porta-enxerto: 313/572/766/Paulsen 1103/SO4
- Duplicata de colheita e vinificação

*Lira*



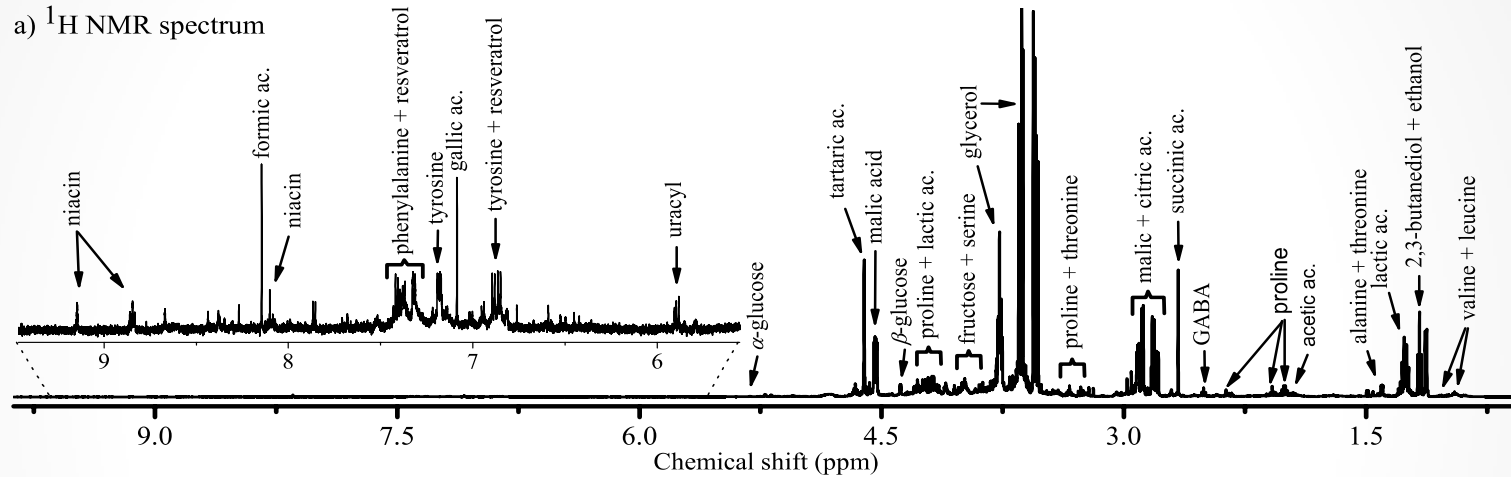
*Espaldeira*



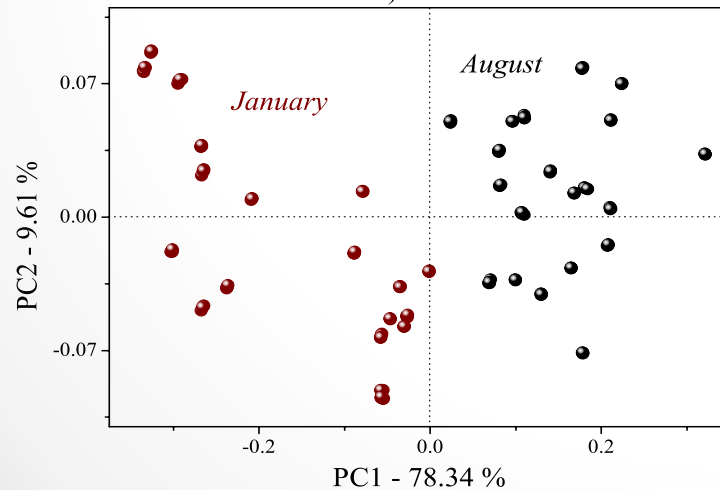
# Estudo quimiométrico relacionado a condições de cultivo/sazonalidade

- a) Espectro de RMN de  $^1\text{H}$  de vinhos da uva Chenin Blanc;
- b) Scores
- c) Loadings

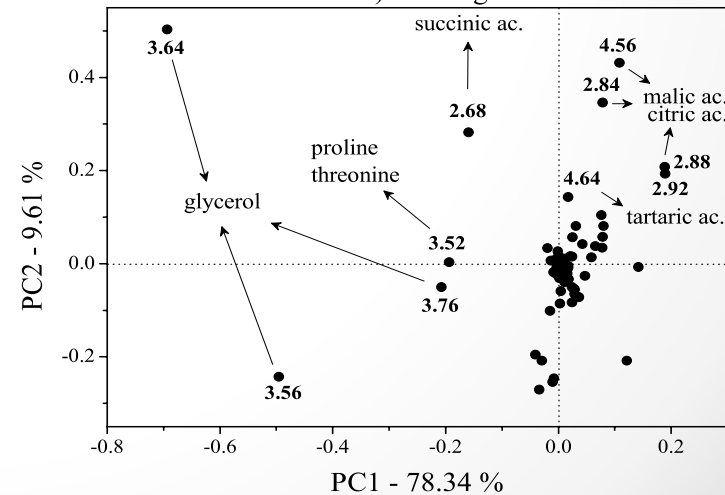
a)  $^1\text{H}$  NMR spectrum



b) Scores

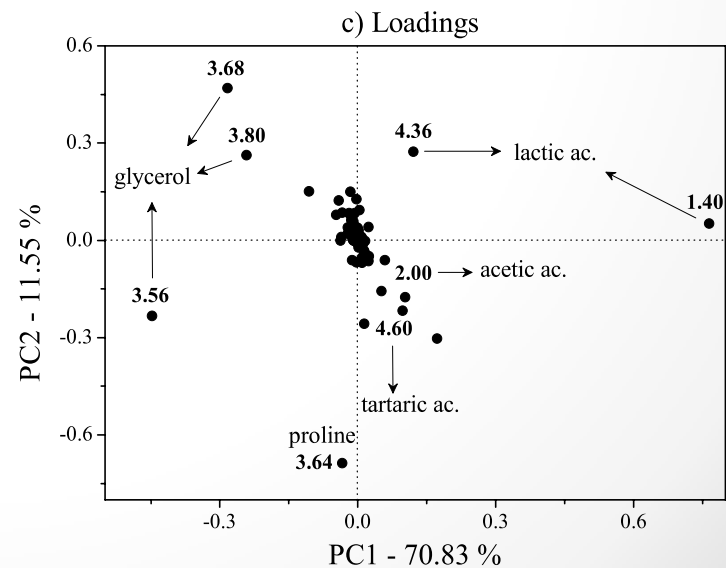
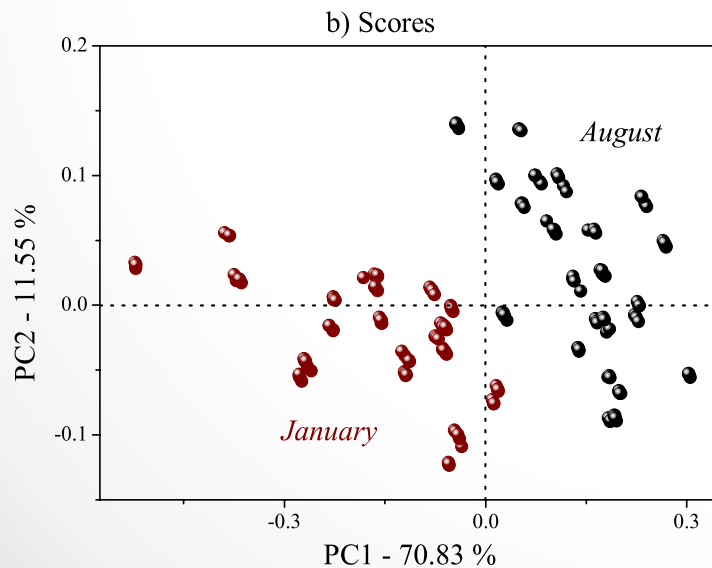
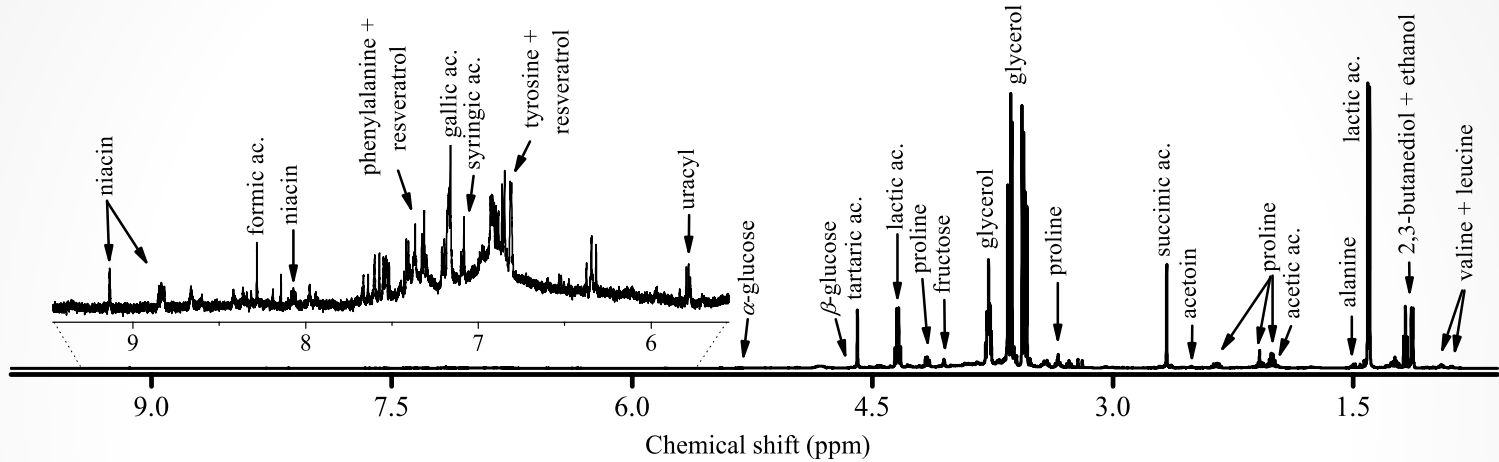


c) Loadings



# Estudo quimiométrico relacionado a condições de cultivo/sazonalidade

- a) Espectro de RMN de  $^1\text{H}$  de vinhos da uva Syrah;
- b) Scores
- c) Loadings



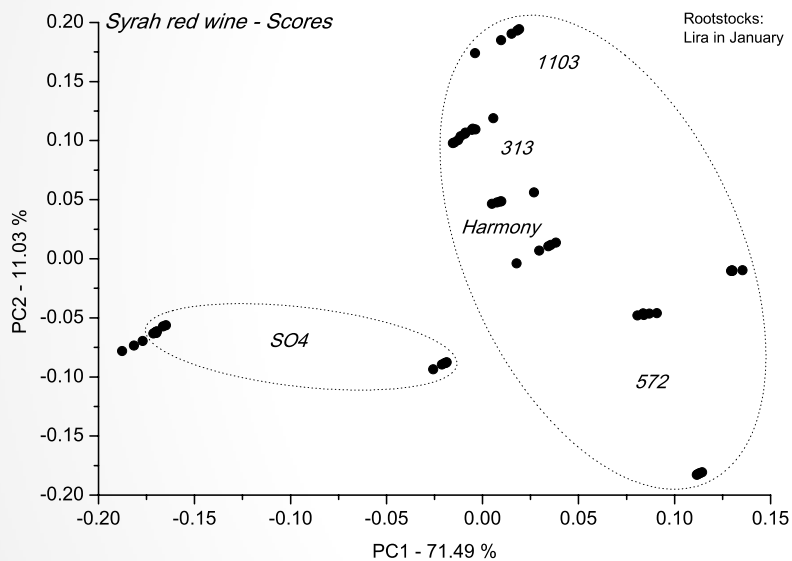


# Estudo quimiométrico relacionado a condições de cultivo/sazonalidade

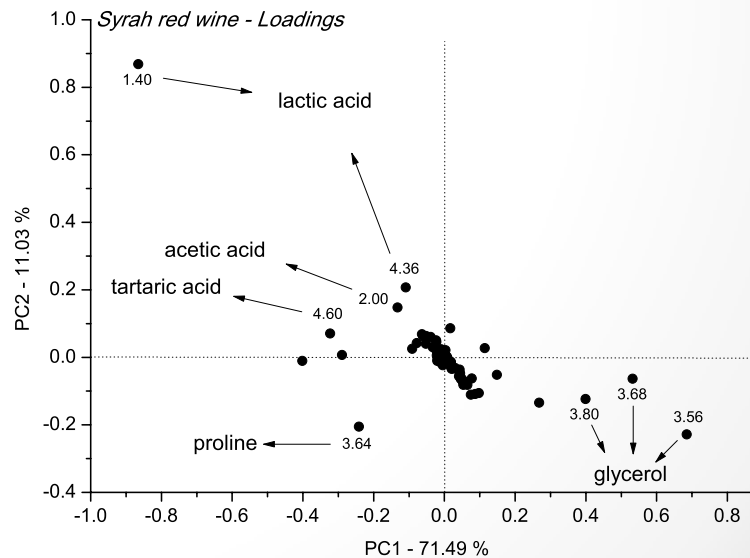
## PCA detalhada dos vinhos:

- uva Syrah
- Janeiro
- Lira

a) Scores - amostras



b) Loadings - variáveis (compostos)



# Estudo quimiométrico relacionado a regionalidade



Magnetic Resonance in Food Science 2016 Proceedings

doi: 10.1255/mrfs.5

## <sup>1</sup>H quantitative nuclear magnetic resonance and principal component analysis as tool for discrimination of guarana seeds from different geographic regions of Brazil

Givaldo Souza da Silva<sup>a</sup>, Lorena Mara Alexandre e Silva<sup>a\*</sup>, Elenilson de Godoy Alves Filho<sup>b</sup>, Kirley Marques Canuto<sup>b</sup>, Edy Sousa de Brito<sup>b</sup> and Ralido Mota de Jesus<sup>b</sup>

<sup>a</sup>Santa Cruz State University, Ilhéus-BA, Brazil

<sup>b</sup>Embrapa Tropical Agroindustry, Fortaleza-CE, Brazil. Corresponding Author: [lorena.mara@embrapa.br](mailto:lorena.mara@embrapa.br)

Guarana (*Paullinia cupana*) is an Amazonian native fruit, whose seeds are rich in caffeine and phenolic compounds. Due to the economic importance of guarana, it is necessary to improve the knowledge of its chemical characteristics, which may vary depending on several factors, such as soil characteristics, crop variety, and environmental conditions. In this context, this study describes the application of quantitative nuclear magnetic resonance associated with principal components analysis to evaluate the composition of guarana seeds from the two largest Brazilian producer regions: Amazonas and Bahia states. The principal component analysis discriminated the samples according to the source, revealing that the seeds from Bahia contain higher caffeine and phenolic contents (catechin, epi-catechin), but the seeds from Amazonas present higher content of fatty acids. Additionally, caffeine was quantified using quantitative nuclear magnetic resonance and liquid chromatography coupled to photodiode array detector, which corroborate the chemometrics. The results show a slightly larger content of caffeine for seeds from Bahia. Therefore, *q*NMR combined with chemometrics was important to detect the metabolic variability of guarana seeds from different regions.

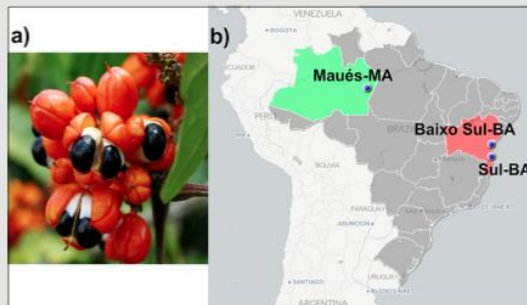


Figure 1. a) *Paullinia cupana* (guarana) fruit. Source: Embrapa archive; b) geographic regions of guarana sampling.



# Estudo quimiométrico relacionado a regionalidade

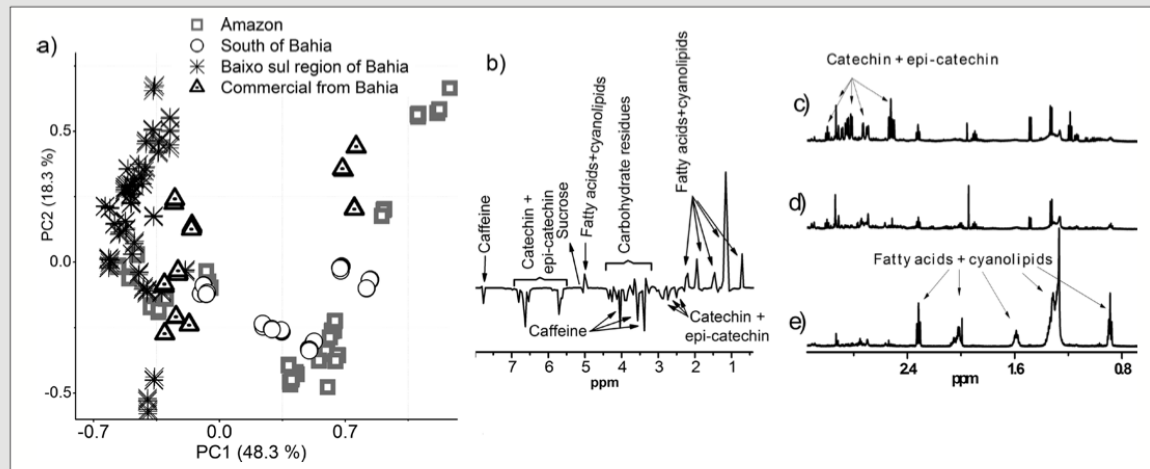


Figure 3. a) PC1xPC2 scores plot of guarana seeds extract from different geographic origin; b) PC1 loadings graph. Comparison among the  $^1\text{H}$  NMR spectra from: c) seeds of Baixo sul region of Bahia state; d) seeds of south of Bahia state; e) seeds of the Amazonas state.

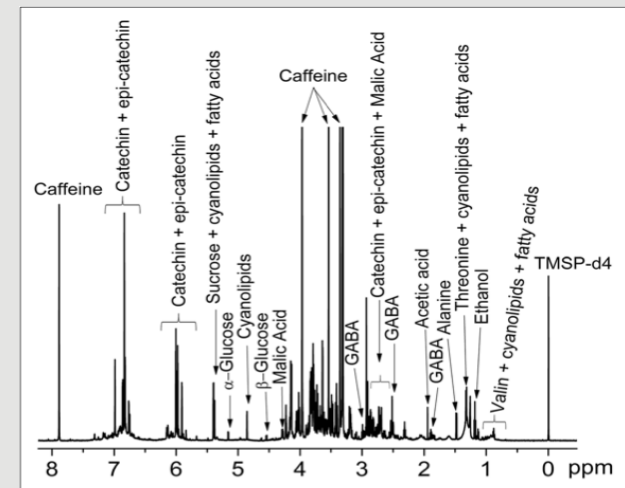
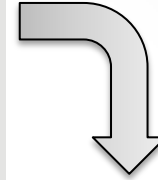


Figure 2.  $^1\text{H}$  NMR spectrum of guarana seeds showing signals labelled with the compounds identified. Legend: GABA – gamma aminobutyric acid.

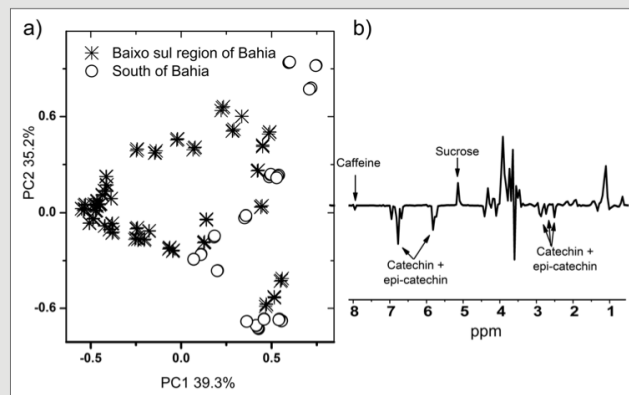
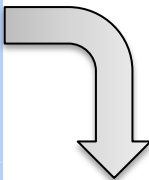
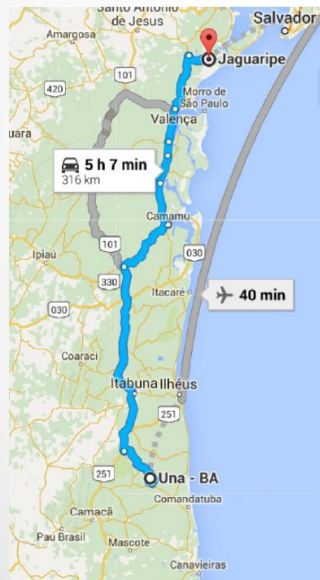


Figure 4. a) PC1 x PC2 of guarana seeds extract from the Bahia state; and b) loadings graph plotted in line from PC1 axis (top) and  $^1\text{H}$  NMR spectra (bottom) to illustrate the NMR regions.

# Estudo quimiométrico relacionado a banco de germoplasma

Banco de Germoplasma do Centro de Ciências Agrárias da  
UFC

Food Research International 91 (2017) 140–147



Contents lists available at ScienceDirect

Food Research International

journal homepage: [www.elsevier.com/locate/foodres](http://www.elsevier.com/locate/foodres)



## Genotype evaluation of cowpea seeds (*Vigna unguiculata*) using $^1\text{H}$ qNMR combined with exploratory tools and solid-state NMR



Elenilson G. Alves Filho <sup>a,b,\*</sup>, Lorena M.A. Silva <sup>a</sup>, Elizita M. Teofilo <sup>c</sup>, Flemming H. Larsen <sup>d</sup>, Edy S. de Brito <sup>a</sup>

<sup>a</sup> EMBRAPA Agroindústria Tropical, Fortaleza, CE, Brazil

<sup>b</sup> LABIOTEC, Dept. Food Technology, Federal University of Ceará, Brazil

<sup>c</sup> Center of Agricultural Science, Federal University of Ceará, Fortaleza, CE, Brazil

<sup>d</sup> Department of Food Science, University of Copenhagen, Denmark

### ARTICLE INFO

#### Article history:

Received 4 October 2016

Received in revised form 2 December 2016

Accepted 10 December 2016

Available online 18 December 2016

#### Keywords:

Cowpea seeds genotype

Oligosaccharides

NMR

Chemometrics

CP-MAS

SP/MAS

### ABSTRACT

The ultimate aim of this study was to apply a non-targeted chemometric analysis (principal component analysis and hierarchical clustering analysis using the heat map approach) of NMR data to investigate the variability of organic compounds in nine genotype cowpea seeds, without any complex pre-treatment. In general, both exploratory tools show that Tvu 233, CE-S84, and Setentão genotypes presented higher amount mainly of raffinose and Tvu 382 presented the highest content of choline and least content of raffinose. The evaluation of the aromatic region showed the Setentão genotype with highest content of niacin/vitamin B3 whereas Tvu 382 with lowest amount. To investigate rigid and mobile components in the seeds cotyledon,  $^{13}\text{C}$  CP and SP/MAS solid-state NMR experiments were performed. The cotyledon of the cowpea comprised a rigid part consisting of starch as well as a soft portion made of starch, fatty acids, and protein. The variable contact time experiment suggests the presence of lipid-amylose complexes.

© 2016 Elsevier Ltd. All rights reserved.

Data in Brief 11 (2017) 136–146



Contents lists available at ScienceDirect

Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

## $^1\text{H}$ NMR spectra dataset and solid-state NMR data of cowpea (*Vigna unguiculata*)



Elenilson G. Alves Filho <sup>a,b,\*</sup>, Lorena M.A. Silva <sup>a</sup>, Elizita M. Teofilo <sup>c</sup>, Flemming H. Larsen <sup>d</sup>, Edy S. de Brito <sup>a</sup>

<sup>a</sup> EMBRAPA Agroindústria Tropical, Fortaleza, CE, Brazil

<sup>b</sup> LABIOTEC, Dept. Food Technology, Federal University of Ceará, Brazil

<sup>c</sup> Center of Agricultural Science, Federal University of Ceará, Fortaleza, CE, Brazil

<sup>d</sup> Department of Food Science, University of Copenhagen, Denmark

### ARTICLE INFO

#### Article history:

Received 22 December 2016

Received in revised form

4 January 2017

Accepted 27 January 2017

Available online 3 February 2017

#### Keywords:

Cowpea seeds

$^1\text{H}$  qNMR

Chemometrics

CP-MAS

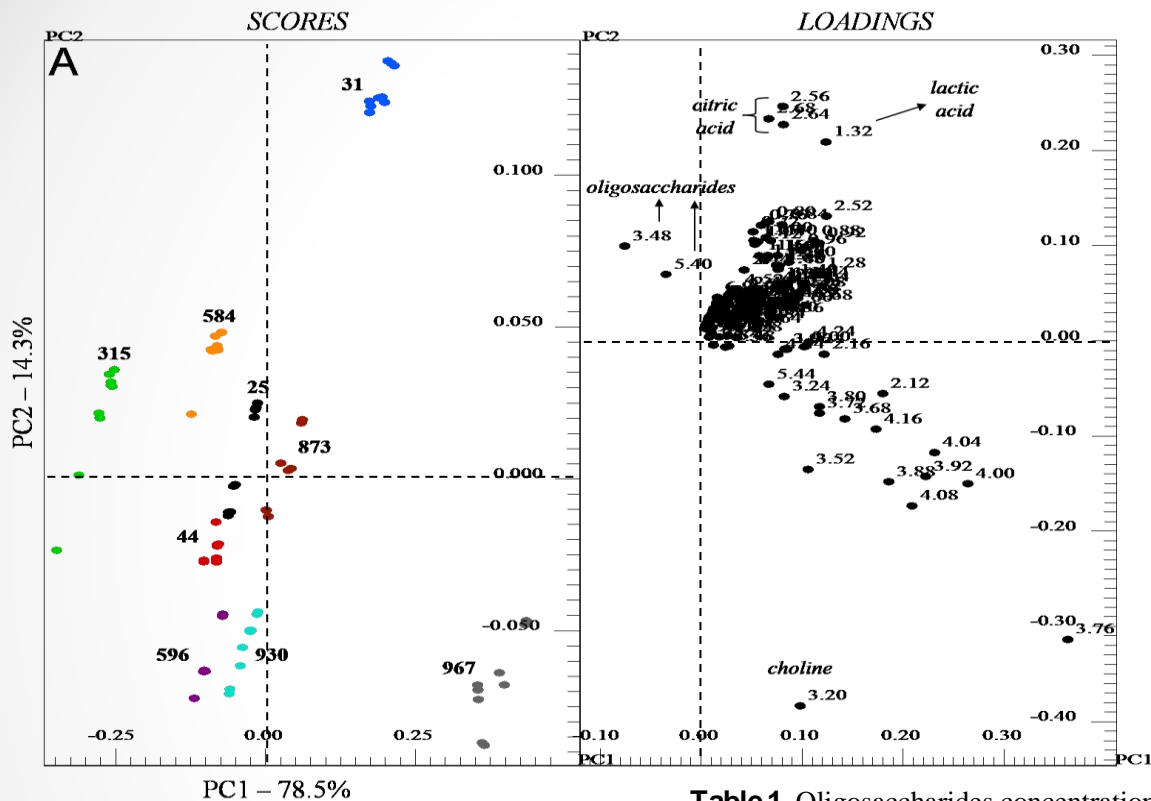
SP/MAS

### ABSTRACT

In this article the NMR data from chemical shifts, coupling constants, and structures of all the characterized compounds were provided, beyond a complementary PCA evaluation for the corresponding manuscript (E.G. Alves Filho, L.M.A. Silva, E.M. Teofilo, F.H. Larsen, E. S. de Brito, 2017) [3]. In addition, a complementary assessment from solid-state NMR data was provided. For further chemometric analysis, numerical matrices from the raw  $^1\text{H}$  NMR data were made available in Microsoft Excel workbook format (.xls).

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

# Estudo quimiométrico relacionado a banco de germoplasma



Quimiometria não supervisionada



Quantificação



Table 1. Oligosaccharides concentrations in mg.g<sup>-1</sup> of cowpea seeds.

Cultivars/ Compounds	CE-25	CE-31	CE-44	CE-315	CE-584	CE-596	CE-873	CE-930	CE-967	
Oligossacarídeos	STA	34.3 ±0.7	28.8 ±0.8	32.2 ±0.8	51.0 ±0.8	42.4 ±0.6	45.2 ±0.8	26.2 ±0.8	32.4 ±0.9	15.1 ±2.0
	RAF	31.7 ±0.5	35.6 ±0.5	34.5 ±0.5	34.1 ±0.7	34.6 ±0.5	43.1 ±0.5	35.2 ±0.5	35.9 ±0.7	33.7 ±0.7
Total	66.0 ±1.2	64.4 ±1.3	66.7 ±1.3	85.1 ±1.5	77.0 ±1.1	88.3 ±1.3	61.4 ±1.3	68.2 ±1.6	48.9 ±2.7	



# Estudo quimiométrico relacionado a banco de germoplasma

Food Chemistry 141 (2013) 2841–2847



Contents lists available at SciVerse ScienceDirect

Food Chemistry

journal homepage: [www.elsevier.com/locate/foodchem](http://www.elsevier.com/locate/foodchem)



Analytical Methods

## Distinction between a transgenic and a conventional common bean genotype by $^1\text{H}$ HR-MAS NMR



Rafael Choze<sup>a</sup>, Glaucia B. Alcantara<sup>a</sup>, Elenilson de G. Alves Filho<sup>a</sup>, Lorena Mara A. e Silva<sup>a</sup>, Josias C. Faria<sup>b</sup>, Luciano M. Lião<sup>a,\*</sup>

<sup>a</sup>Instituto de Química, Universidade Federal de Goiás, C.P. 131, 74001-970 Goiânia, GO, Brazil

<sup>b</sup>Embrapa Arroz e Feijão, C.P. 179, Santo Antônio de Goiás, 75375-000 GO, Brazil

### ARTICLE INFO

#### Article history:

Received 30 November 2010

Received in revised form 21 November 2012

Accepted 24 May 2013

Available online 4 June 2013

#### Keywords:

*Phaseolus vulgaris*

Common beans

Transgenic

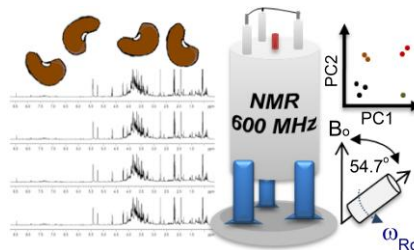
HR-MAS NMR

### ABSTRACT

In this work, it is proposed a methodology which allows to differentiate a conventional and a specific transgenic common beans, grown in greenhouse or under field conditions, based on modifications in chemical composition using  $^1\text{H}$  HR-MAS NMR. It is demonstrated that the influence of typical variables from field planting conditions had no significant influence on the ability to set apart transgenic from conventional. This methodology was corroborated by multivariate data analysis of the  $^1\text{H}$  NMR and IR spectra. This study also points out the simplicity of using the HR-MAS NMR technique for food analyses. The measurement is highly simplified because it does not require any pretreatment of the sample apart from the addition of a small amount of  $\text{D}_2\text{O}$  necessary to produce homogeneous dough and a field frequency lock. Moreover, due to the high concentration of the sample, measurement time in HR-MAS NMR is very short.

© 2013 Elsevier Ltd. All rights reserved.

**Embrapa**  
Arroz e Feijão

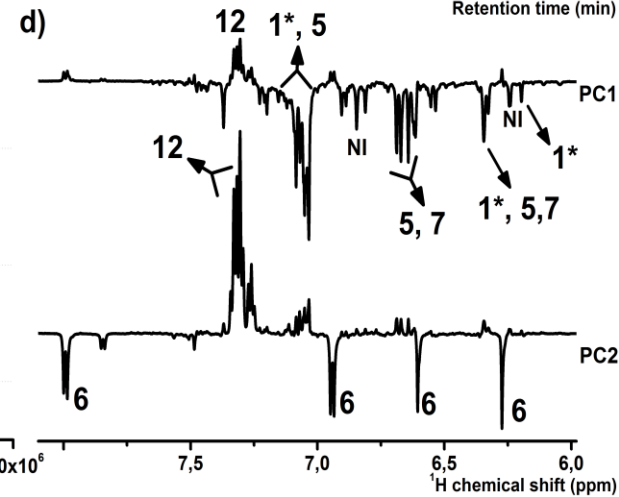
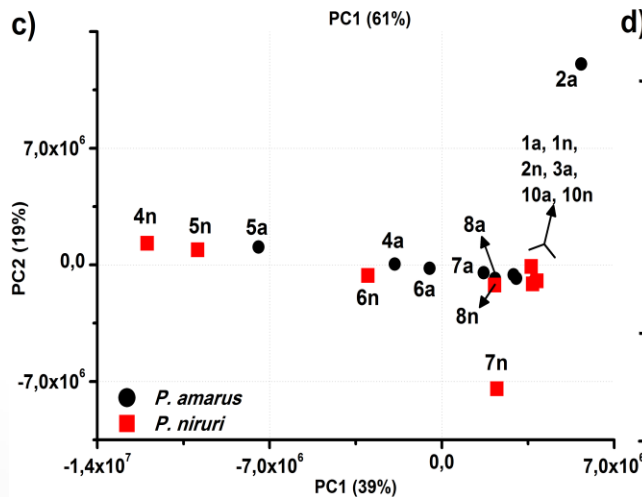
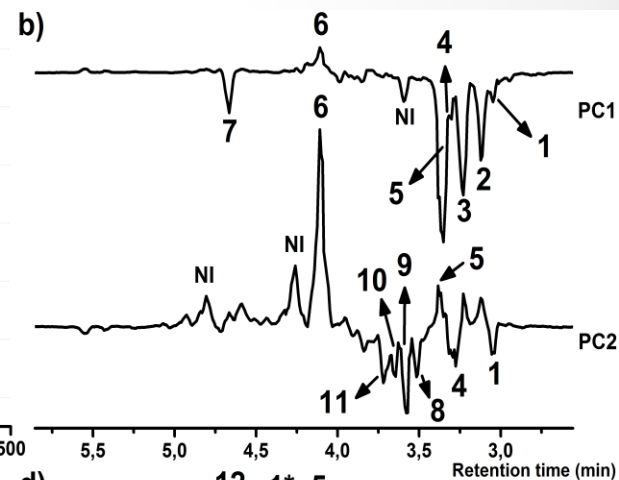
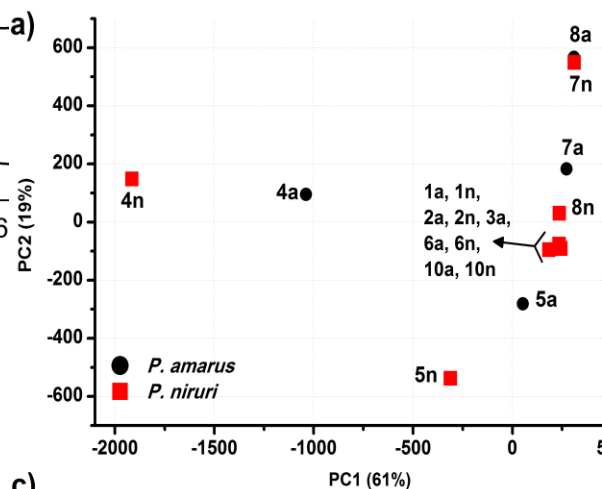
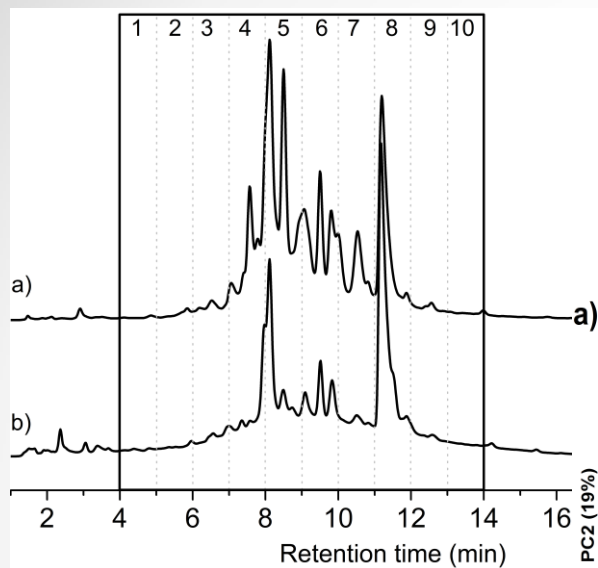


# Estudo quimiométrico RMN × HPLC-MS

10 extratos:

- *a* = *P. amarus*

- *n* = *P. niruri*





# Estudo quimiométrico RMN × GC-MS

Talanta 180 (2018) 329–336



Contents lists available at ScienceDirect

Talanta

journal homepage: [www.elsevier.com/locate/talanta](http://www.elsevier.com/locate/talanta)



## Chemometric analysis of NMR and GC datasets for chemotype characterization of essential oils from different species of *Ocimum*



João Vito B. Freitas<sup>a</sup>, Elenilson G. Alves Filho<sup>b</sup>, Lorena Mara A. Silva<sup>b</sup>, Guilherme J. Zocolo<sup>b</sup>, Edy S. de Brito<sup>b</sup>, Nilce V. Gramosa<sup>a,\*</sup>

<sup>a</sup> Departamento de Química Orgânica e Inorgânica, Universidade Federal do Ceará, Fortaleza, CE, Brazil

<sup>b</sup> Embrapa Agroindústria Tropical, Fortaleza, CE, Brazil

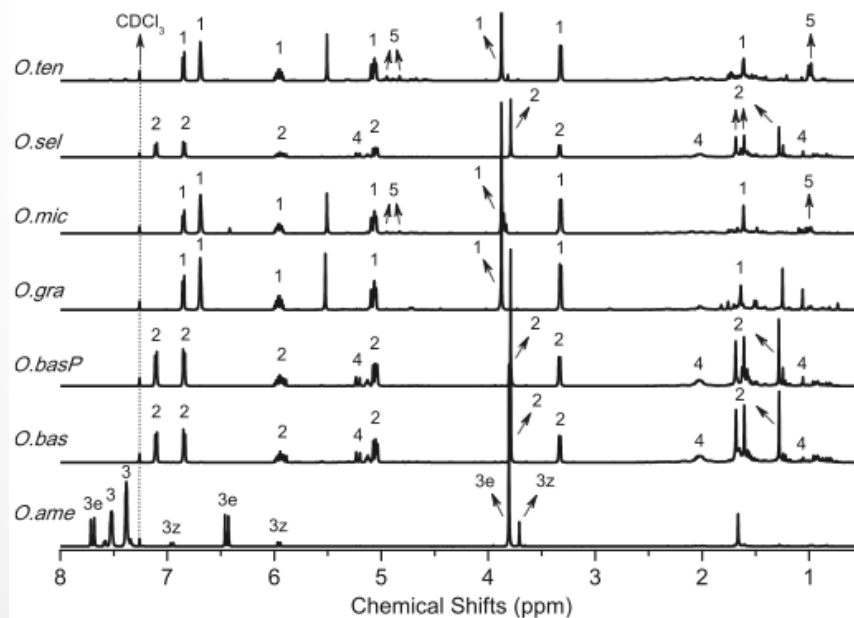
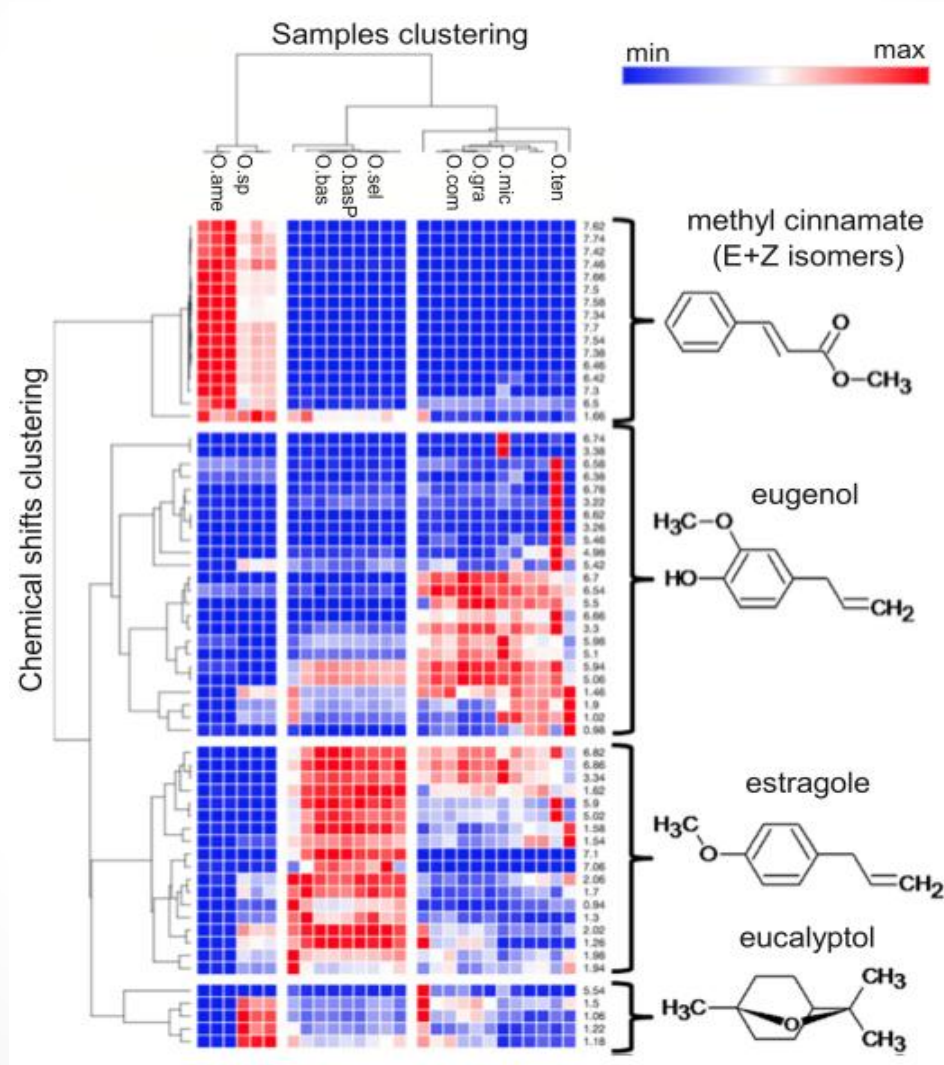


Fig. 1. Comparison of the  $^1\text{H}$  NMR spectra of different species from *O.ame* (*Ocimum americanum*), *O.bas* (*Ocimum basilicum*), *O.basP* (*Ocimum basilicum* var. *purpuracens*), *O.gra* (*Ocimum gratissimum*), *O.mic* (*Ocimum micranthum*), *O.sel* (*Ocimum selloi*), and *O.ten* (*Ocimum tenuiflorum*). Legend: 1 – eugenol, 2 – estrágole, 3e – *E*-methyl cinnamate, 3z – *Z*-methyl cinnamate, 4 – linalool, and 5 –  $\beta$ -caryophyllene.



# Estudo quimiométrico RMN × GC-MS

*heat map*



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes complexas

## ***Principais análises de calibração/regressão dos dados (supervisionadas, conhecimento prévio):***

*obs.: métodos que procuram criar um modelo que descreva os dados de forma a quantificar a relação entre dois grupos de variáveis → X e Y*

- Método de regressão por mínimos quadrados parciais → PLS/PLS-DA (*Partial Least Squares – Discriminant Analysis*)
- Método do vizinho mais próximo → KNN (*K-Nearest Neighbor*)
- Modelagem suave independente por analogia de classe → SIMCA (*Soft Independent Modeling of Class Analogy*)

### ***Regressão multivariada***

*Método que procuram modelar dados de forma a conhecer uma relação entre dois ou mais grupos de variáveis, para descrever esta relação ou para prever novos valores.*

# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes complexas

## Método de Regressão por Mínimos Quadrados Parciais – PLS

Explicar variações das variáveis Y através das variações nas variáveis X

### VARIÁVEIS INDEPENDENTES

Long Name	A(X)	Q(Y)	R(Y)	S(Y)	T(Y)	U(Y)	V(Y)	W(Y)	X(Y)	Y(Y)	Z(Y)	AA(Y)	AB(Y)	AC(Y)	AD(Y)	AE(Y)	B(Y)
Amostras	0.5 ppm	1.0 ppm	1.5 ppm	2.0 ppm	2.5 ppm	3.0 ppm	3.5 ppm	4.0 ppm	4.5 ppm	5.0 ppm	5.5 ppm	6.0 ppm	6.5 ppm	7.0 ppm	7.5 ppm	Função Resposta	
1 Amostra 1	-6.12960	-3.93400	-3.74400	-16.0422	-8.56630	-7.90010	-11.4652	-11.9444	-10.9445	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-11.4652	1.00000	
2 Amostra 2	-5.60530	-4.88580	-2.85890	-16.6824	-6.89770	-8.29170	-8.78110	-8.75380	-13.2263	-6.00780	-11.8920	-10.2839	-32.6477	-7.23870	-8.78110	1.00000	
3 Amostra 3	-4.35330	-3.18580	-2.57540	-15.3622	-8.07880	-8.92000	-6.13200	-14.0918	-15.3337	-9.03270	-14.0169	-13.1059	-23.5460	-11.0579	-6.13200	1.00000	
4 Amostra 4	-4.00100	-4.82000	-4.60380	-11.6526	-8.19190	-9.88750	-17.3969	-11.7166	-13.4233	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-17.3969	2.00000	
5 Amostra 5	-5.41880	-7.11660	-6.36720	-10.7203	-12.6022	-8.01010	-10.8263	-10.5608	-14.8458	-8.03990	-8.80460	-12.8046	-25.8944	-7.10760	-10.8263	2.00000	
6 Amostra 6	-3.78450	-7.77060	-5.74360	-12.8627	-7.24940	-10.0148	-10.3575	-16.4900	-18.1446	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-10.3575	2.00000	
7 Amostra 7	-4.32290	-5.20820	-5.12730	-12.0741	-3.94300	-1.59840	-9.62900	-16.2888	-20.5803	-0.00780	-11.8920	-10.2839	-32.6477	-7.23870	-9.63900	3.00000	
8 Amostra 8	-5.46260	-7.09780	-1.25950	-13.7586	-11.6288	-5.55540	-11.2555	-10.8263	-10.7388	-9.03270	-14.0169	-13.1059	-23.5460	-11.0579	-12.8154	3.00000	
9 Amostra 9	-4.73020	-6.67950	-2.54220	-10.9181	-9.91770	-10.3575	-11.2555	-9.88750	-10.7388	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-13.2492	3.00000	
10 Amostra 10	-3.76680	-8.45230	-5.43130	-15.9566	-7.95820	-5.18270	-11.9444	-8.03990	-15.5852	-0.03990	-8.80460	-12.8046	-25.8944	-7.10760	-12.9457	3.00000	
11 Amostra 11	-3.77910	-6.18320	-6.25970	-17.3355	-14.1177	-11.4752	-4.35330	-15.6672	-16.9064	-8.48370	-12.6168	-16.5185	-25.6425	-12.9705	-4.35330	4.00000	
12 Amostra 12	-5.60250	-7.63830	-2.63660	-15.9857	-9.06870	-7.65940	-11.4723	-15.8113	-11.0772	-14.4538	-17.6184	-11.5801	-21.4435	-21.0691	-11.4723	4.00000	
13 Amostra 13	-6.63250	-7.20500	-5.19560	-16.3599	-8.38810	-8.44650	-5.97750	-14.8479	-17.1128	-6.06940	-12.6701	-12.7060	-16.1058	-14.1553	-5.97750	5.00000	
14 Amostra 14	-7.01780	-6.84440	-5.55070	-16.0415	-14.5741	-13.0748	-6.65650	-11.5459	-11.4694	-7.92760	-4.40530	-13.3188	-17.2564	-28.2828	-6.65650	5.00000	
15 Amostra 15	-6.53900	-5.97200	-3.14880	-20.4323	-13.2241	-17.7318	-8.71800	-13.8167	-13.9193	-11.7536	-8.54060	-8.15650	-11.6307	-21.3215	-8.71800	5.00000	
16 Amostra 16	-4.75900	-5.45510	-5.19170	-15.1259	-7.81480	-13.8332	-16.0618	-11.9624	-12.2261	-12.4957	-11.0392	-11.9333	-15.8174	-15.9668	-16.0618	6.00000	
17 Amostra 17	-3.95600	-5.07640	-4.16330	-16.2087	-10.2615	-10.0024	-13.6478	-15.0621	-14.3768	-10.5041	-10.4618	-8.31990	-28.5540	-19.7976	-13.6478	6.00000	
18 Amostra 18	-2.55840	-5.77230	-6.15910	-13.9531	-13.6938	-9.01310	-8.67210	-15.8102	-14.7981	-8.71770	-11.9790	-13.4979	-23.8291	-22.8878	-8.67210	6.00000	
19 Amostra 19	-4.83510	-7.22520	-2.81990	-15.1478	-9.16250	-8.04410	-9.37930	-15.1017	-15.0568	-11.1856	-10.7040	-11.8671	-13.7654	-17.6918	-9.37930	7.00000	

Matriz X

Matriz Y

### VARIÁVEIS DEPENDENTES

Informações intrínsecas das amostras

- Matrizes “X” e “Y” serão decompostas em escores e *loadings*
- Cada PC será modificada para buscar máxima covariância entre “X” e “Y”
- Componentes Principais (PC) → Variáveis Latentes (LV): perdem ortogonalidade

Buscar a melhor relação linear entre os valores de escores e *loadings* de “X” com os escores de “Y”  
(cálculo dos coeficientes de regressão)

# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes complexas

## Análise Discriminante com Método de Regressão por Mínimos Quadrados Parciais – PLS-DA

### VARIÁVEIS INDEPENDENTES

Long Name	A(X)	Q(Y)	R(Y)	S(Y)	T(Y)	U(Y)	V(Y)	W(Y)	X(Y)	Y(Y)	Z(Y)	AA(Y)	AB(Y)	AC(Y)	AD(Y)	AE(Y)	B(Y)
Amostras	0.5 ppm	1.0 ppm	1.5 ppm	2.0 ppm	2.5 ppm	3.0 ppm	3.5 ppm	4.0 ppm	4.5 ppm	5.0 ppm	5.5 ppm	6.0 ppm	6.5 ppm	7.0 ppm	7.5 ppm		Função Resposta
1	-6.12960	-3.93400	-3.74400	-16.0422	-8.56630	-7.90010	-11.4652	-11.9444	-10.9445	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-11.4652		1.00000
2	-5.60530	-4.88580	-2.85890	-16.6824	-6.89770	-8.29170	-8.78110	-8.75380	-13.2263	-6.00780	-11.8920	-10.2839	-32.6477	-7.23870	-8.78110		1.00000
3	-4.35330	-3.18580	-2.57540	-15.3622	-8.07880	-8.92000	-6.13200	-14.0918	-15.3337	-9.03270	-14.0169	-13.1059	-23.5460	-11.0579	-6.13200		1.00000
4	-4.00100	-4.82000	-4.60380	-11.6526	-8.19190	-9.88750	-17.3969	-11.7166	-13.4233	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-17.3969		2.00000
5	-5.41880	-7.11660	-6.36720	-10.7203	-12.6022	-8.01010	-10.8263	-10.5608	-14.8458	-8.03990	-8.80460	-12.8046	-25.8944	-7.10760	-10.8263		3.00000
6	-3.78450	-7.77060	-5.74360	-12.8627	-7.24940	-10.0148	-10.3575	-16.4900	-18.1446	-7.66220	-4.03630	-8.98090	-14.5308	-14.3044	-10.3575		2.00000
7	-4.32290	-5.20820	-5.12730	-12.0741	-9.39430	-11.5984	-9.62900	-16.2788	-20.5803	-10.0780	-11.8920	-10.2839	-32.6477	-7.23870	-9.62900		3.00000
8	-5.46260	-7.09780	-1.25950	-13.7586	-11.6288	-5.92550	-10.2498	-9.2860	-10.7388	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-11.8921		3.00000
9	-4.73020	-6.67950	-2.54220	-10.9181	-11.9177	-10.5011	-11.2498	-9.2860	-10.7388	-11.8921	-4.93450	-10.0461	-23.9340	-16.4655	-13.2492		3.00000
10	-3.76680	-8.45230	-5.43130	-15.9566	-11.8258	-18.2794	-9.06660	-15.5852	-10.3938	-8.80460	-12.8046	-25.8944	-7.10760	-12.9457	-10.3938		4.00000
11	-3.77910	-6.18320	-6.25970	-17.3355	-14.1177	-11.4752	-4.35330	-15.6672	-16.9064	-8.48370	-12.6168	-16.5185	-25.6425	-12.9705	-4.35330		4.00000
12	-5.60250	-7.63830	-2.63660	-15.9857	-9.06870	-7.65940	-11.4723	-15.8113	-11.0772	-14.4538	-17.6184	-11.5801	-21.4435	-21.0691	-11.4723		5.00000
13	-6.63250	-7.20500	-5.19560	-16.3599	-8.38810	-8.44650	-5.97750	-14.8479	-17.1128	-6.06940	-12.6701	-12.7060	-16.1058	-14.1553	-5.97750		5.00000
14	-7.01780	-6.84440	-5.55070	-16.0415	-14.5741	-13.0748	-6.65650	-11.5459	-11.4694	-7.92760	-4.40530	-13.3188	-17.2564	-28.2828	-6.65650		5.00000
15	-6.53900	-5.97200	-3.14880	-20.4323	-13.2241	-17.7318	-8.71800	-13.8167	-13.9193	-11.7536	-8.54060	-8.15650	-11.6307	-21.3215	-8.71800		5.00000
16	-4.75900	-5.45510	-5.19170	-15.1259	-7.81480	-13.8332	-16.0618	-11.9624	-12.2261	-12.4957	-11.0392	-11.9333	-15.8174	-15.9668	-16.0618		6.00000
17	-3.95600	-5.07640	-4.16330	-16.2087	-10.2615	-10.0024	-13.6478	-15.0621	-14.3768	-10.5041	-10.4618	-8.31990	-28.5540	-19.7976	-13.6478		6.00000
18	-2.55840	-5.77230	-6.15910	-13.9531	-13.8938	-9.01310	-8.67210	-15.8102	-14.7981	-8.71770	-11.9790	-13.4979	-23.8291	-22.8878	-8.67210		6.00000
19	-4.83510	-7.22520	-2.81990	-15.1478	-9.16250	-8.04410	-9.37930	-15.1017	-15.0568	-11.1856	-10.7040	-11.8671	-13.7654	-17.6918	-9.37930		7.00000

Matriz X

Função Resposta

### VARIÁVEIS DEPENDENTES grupos de amostras

Geralmente, melhores agrupamentos → PLS-DA × PCA

PCA → método não supervisionado → cada uma das PCs buscará as direções com maior variância da matriz “X” de dados

PLS-DA → método supervisionado → fornecerá dados identificando as amostras, otimizando a separação entre os grupos através da rotação dos eixos (LVs) para obter o máximo de separação entre as classes; ou característica previamente adquirida, para uma PLS.

# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes complexas

## *Análises de calibração/regressão*

Erro de Classificação → Validação Cruzada (completa; randômica...)

**LEAVE-ONE-OUT**

Número de LVs → Parâmetros estatísticos:

- erro padrão na calibração (SEC – *Standard Error of Calibration*)
- erro padrão na validação (SEV – *Standard Error of Validation*)
- raiz quadrada do erro médio quadrático de calibração (RMSEC – *Root Mean Square Error of Calibration*)
- raiz quadrada do erro médio quadrático de validação cruzada (RMSECV – *Root Mean Square Error of Cross Validation*)

Saber se o modelo foi bem ajustado e conhecer o número de LVs → Proximidade dos valores de RMSECV e RMSEC

Sensibilidade do modelo → Divida o número de amostras previstas como pertencentes à classe pelo número de amostras que são realmente da classe

Especificidade do modelo → Dividir o número de amostras previstas como não pertencentes à classe pelo número real de amostras que não pertencem à classe

**Valores próximos a um (1):**  
acertos próximos de 100% na previsão das amostras de validação



# Análises estatísticas multivariadas dos dados e suas aplicações nas análises de matrizes complexas

- Avaliar a capacidade do modelo de calibração desenvolvido (robustez):  
*Raiz quadrada do erro médio quadrático de calibração (RMSEC), na validação cruzada (RMSECV) e na predição (RMSEP) devem ser analisados*
- RMSEC (*Root Mean Square Error of Calibration*) é uma medida do erro na modelagem e é calculado a partir das amostras de calibração
- RMSECV (*Root Mean Square Error of Cross-Validation*) é o erro padrão na validação cruzada que é obtido a partir das amostras da validação cruzada
- RMSEP (*Root Mean Square Error of Prediction*) é o erro padrão que é calculado a partir das amostras do conjunto de predição
- Um bom ajuste das amostras utilizadas para a calibração → Valores baixos de RMSEC, RMSECV e RMSEP
- Valores de RMSECV e RMSEP maiores que RMSEC → modelo sobre-ajustado, ou seja, a regressão encontrada considerou dados não correlacionados, como ruídos, erros sistemáticos...
- Um bom modelo de predição será o que minimiza os valores de RMSEP na avaliação do ajuste do modelo de calibração
- Curva de calibração × Curva de validação cruzada → quanto mais próximas, mais robusto será o modelo
- *Outliers* podem ser determinados por análise visual do gráfico “Q residuals × Hotelling’s T<sup>2</sup>”
- O valor da razão RMSEC/RMSECV acima de 0.75 → modelo bem ajustado