

Tackling Complex Mixture by NMR

25th Jan., 2022

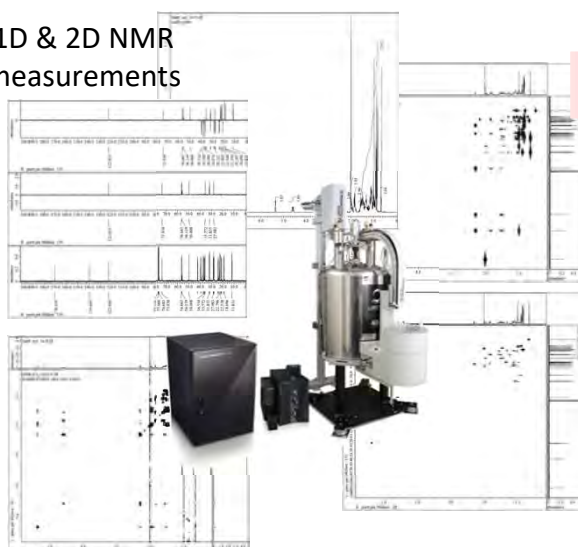
JEOL RESONANCE, Inc.

Takanori Komatsu, Ph.D.

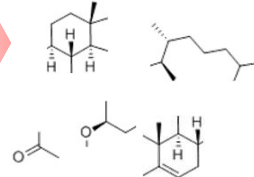
NMR in laboratories (pure world): Structural Elucidation

NMR is essential tool for *de-novo* structural elucidation of organic molecules in the laboratories. 😊

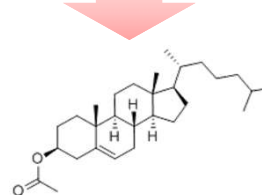
1D & 2D NMR
measurements



Partial structures

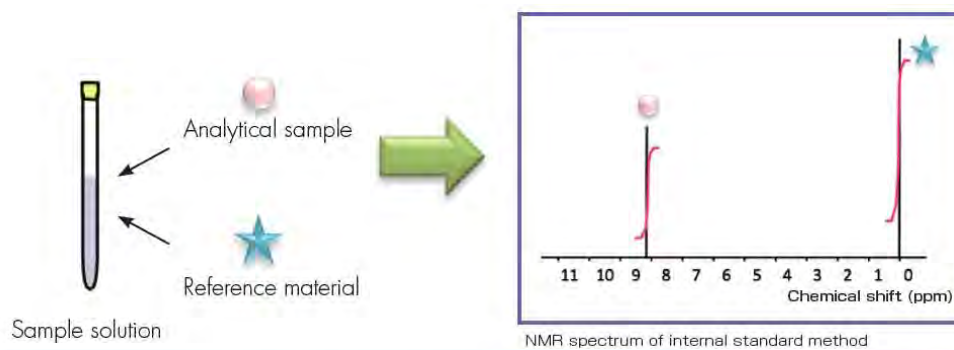


Structural elucidation by NMR



NMR in laboratories (pure world): Quantitative NMR

Quantitative nature of NMR is getting much attention for assay for pure chemicals. 😊



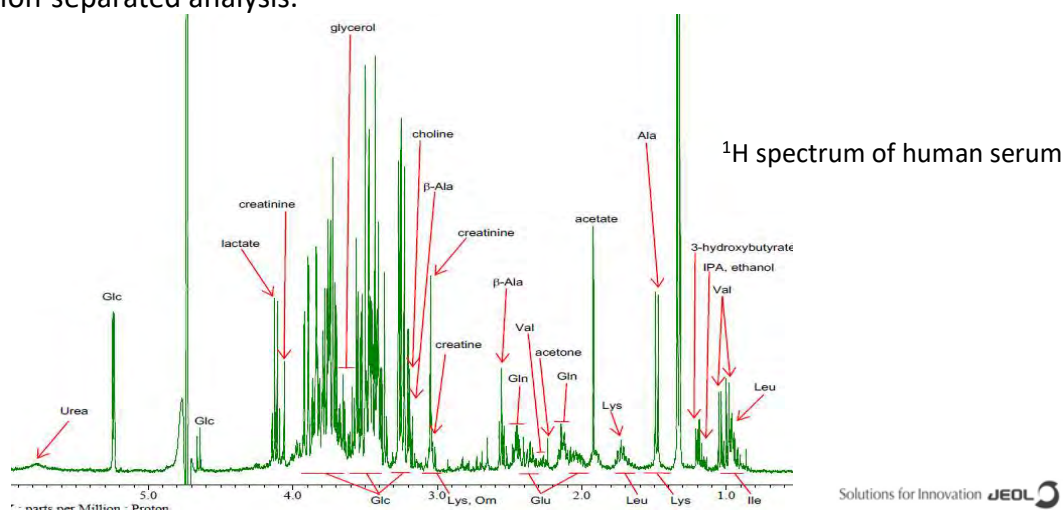
$$\text{Formula (1)} \quad P_{\text{sample}} = \frac{I_{\text{sample}}}{I_{\text{std}}} \times \frac{H_{\text{std}}}{H_{\text{sample}}} \times \frac{m_{\text{std}}}{m_{\text{sample}}} \times \frac{M_{\text{sample}}}{M_{\text{std}}} \times P_{\text{std}}$$

I = signal intensity (integral value), H = number of protons (number of hydrogen atoms in the functional group),
 m = mass (weight), M = molecular weight, P = purity (%)

Solutions for Innovation JEOL

Toward laboratory (pure) to real world (mixture)

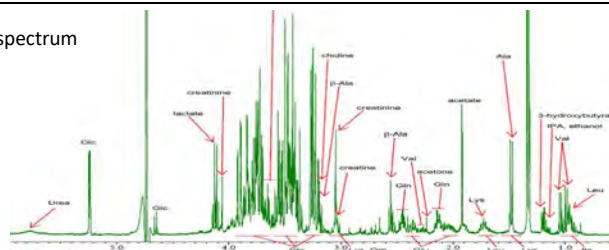
Real world: Very complex mixture -> Severe spectral overlapping 😞 Umm...
 NMR measurements are highly reproducible with highly quantitative nature, but it is typically non-separated analysis.



Solutions for Innovation JEOL

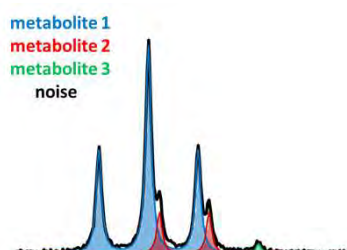
Strategies for Complex Mixture Analysis by solution NMR

ex. Serum ^1H -NMR spectrum



Targeted Analysis

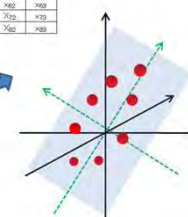
- "Reductionism"
- Spectral library is required.



Non-Targeted Analysis

- "Holistic approach"
- Statistical analyses are typically used.

	py	sp	sp	sp
r1	0	x11	x12	x13
r2	0	x21	x22	x23
r3	0	x31	x32	x33
r4	0	x41	x42	x43
r5	1	x51	x52	x53
r6	1	x61	x62	x63
r7	1	x71	x72	x73
r8	1	x81	x82	x83



Solutions for Innovation JEOL

Outline

Introduction (done)

Workflow of NMR Non-targeted analysis

Chemometrics tool in the Delta NMR software

Unsupervised Multivariate Analyses in the Chemometrics tool

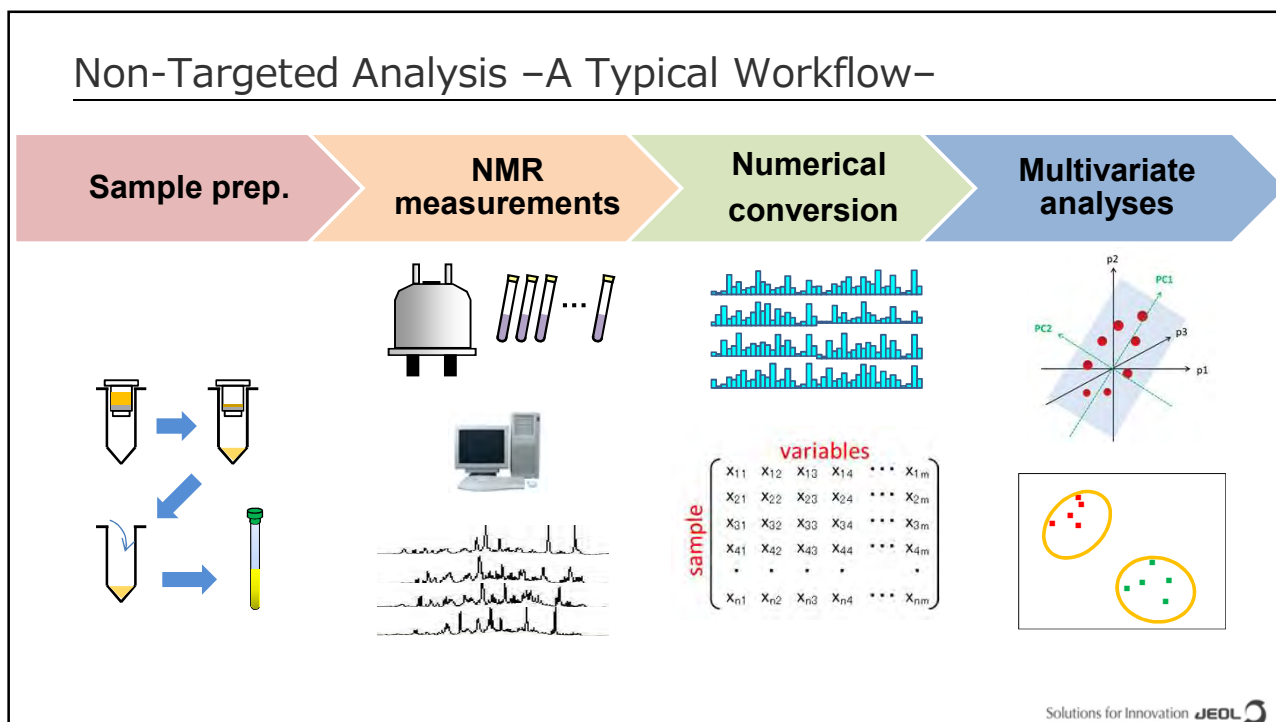
Classical and Robust Principle Component Analysis

Hierarchal Cluster Analysis

Application Example

Solutions for Innovation JEOL

Non-Targeted Analysis –A Typical Workflow–



Sample Preparation

Unfortunately there are no "golden recipe" for NMR sample preparation with all materials.

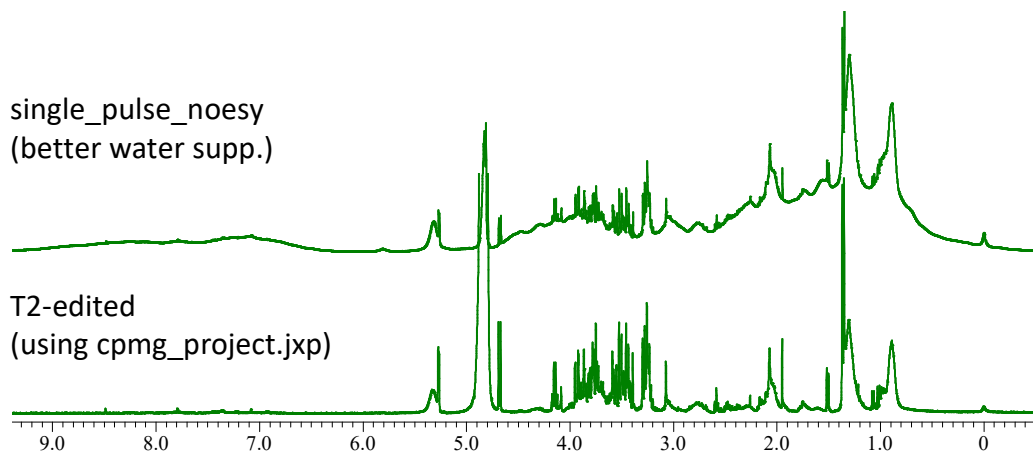
Generally simple sample preparation procedures are okay for NMR measurements (ex. add few deuterated solvent for 2H lock or even though without deuterated solvent), but that doesn't mean good spectral quality.

Liquid-liquid extraction (ex. variants of Bligh-Dyer method) is sometimes employed to remove non-polar (or polar) and neutral compounds.

Ultrafilterization is sometimes employed to remove macromolecules (ex. biomacromolecules in serum).

Previous studies using similar material are helpful for choosing procedures of sample preparation.

“Edited” or “As it is”?



Any 1D experiment is applicable.

Solvent suppression is often employed when solvent peaks hamper further analyses.
Spectral-editing is sometimes used for emphasizing what you want to see.

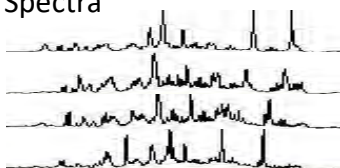
Solutions for Innovation JEOL

Numerical conversion (Spectra to Matrix)

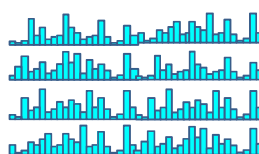
Bucket Integration

(Chemometrics tool supports this)

Spectra



Spectra are divided into even width windows (ex. 0.04 ppm)

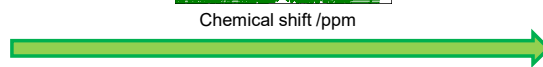
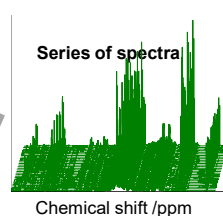
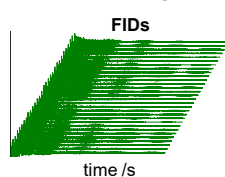


matrix

$$\begin{matrix}
 & \text{variables} & \\
 \text{sample} & \begin{pmatrix}
 X_{11} & X_{12} & X_{13} & X_{14} & \dots & X_{1m} \\
 X_{21} & X_{22} & X_{23} & X_{24} & \dots & X_{2m} \\
 X_{31} & X_{32} & X_{33} & X_{34} & \dots & X_{3m} \\
 X_{41} & X_{42} & X_{43} & X_{44} & \dots & X_{4m} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 X_{n1} & X_{n2} & X_{n3} & X_{n4} & \dots & X_{nm}
 \end{pmatrix}
 \end{matrix}$$

CRAFT

(Please be patient, it takes long time for converting...)



Sample	Variable 1	Variable 2	Variable 3	Variable 4	Variable 5	Variable 6	Variable 7	Variable 8	Variable 9
1	0.123	0.456	0.789	0.234	0.567	0.890	0.345	0.678	0.901
2	0.234	0.567	0.890	0.345	0.678	0.901	0.456	0.789	0.123
3	0.345	0.678	0.901	0.456	0.789	0.123	0.567	0.890	0.234
4	0.456	0.789	0.123	0.567	0.890	0.234	0.678	0.901	0.345
5	0.567	0.890	0.234	0.678	0.901	0.345	0.789	0.123	0.456
6	0.678	0.901	0.345	0.789	0.123	0.456	0.890	0.234	0.567
7	0.789	0.123	0.456	0.890	0.234	0.567	0.901	0.345	0.678
8	0.890	0.234	0.567	0.901	0.345	0.678	0.123	0.456	0.789
9	0.901	0.345	0.678	0.123	0.456	0.789	0.234	0.567	0.890
10	0.123	0.456	0.789	0.234	0.567	0.890	0.345	0.678	0.901

Solutions for Innovation JEOL

Multivariate Analysis (Extract “something interesting” from “illegible matrix”)

Unsupervised Analysis

Uses only explanatory variables (here NMR bins) and doesn't use objective variables.

PCA; Principal Component Analysis #The chemometric tool supports PCA.

HCA; Hierarchical Cluster Analysis #The chemometric tool supports HCA.

Supervised Analysis

Uses both explanatory variables and objective variables.

PLS-DA; Partial Least Squared-Discrimination Analysis

PLS-R; Partial Least Squared-Regression #The chemometric tool doesn't support these supervised analyses.

Solutions for Innovation JEOL 

Outline

Introduction (done)

Workflow of NMR Non-targeted analysis

Chemometrics tool in the Delta NMR software

Unsupervised Multivariate Analyses in the Chemometrics tool

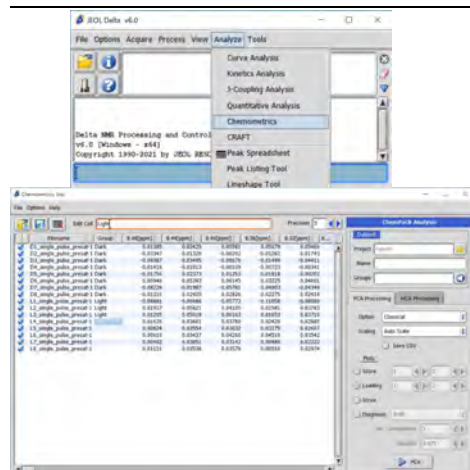
Classical and Robust Principle Component Analysis

Hierarchical Cluster Analysis

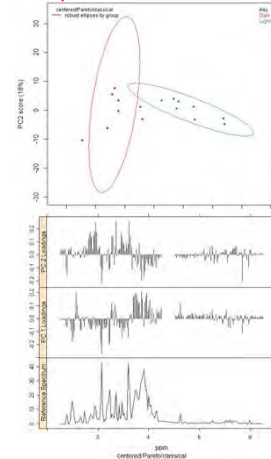
Application Example

Solutions for Innovation JEOL 

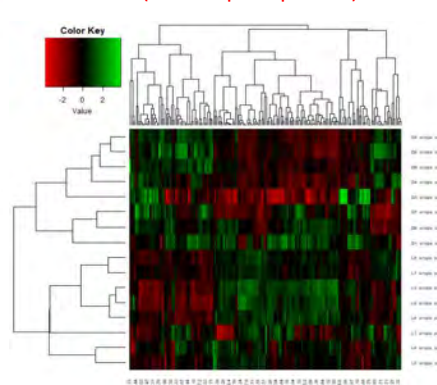
Chemometrics tool



PCA (both classical and robust)



HCA
(heatmap is optional)

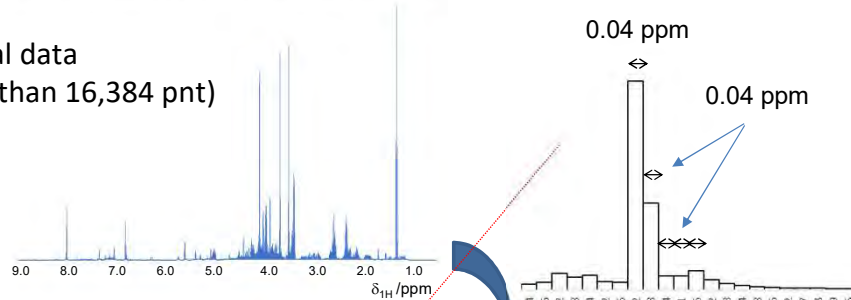


- Delta NMR software has “Chemometrics tool” for non-target explanatory analyses.
- The “Chemometrics tool” supports bucket integration (convert multiple spectra into matrix), PCA*, HCA*.
- Robust version PCA is available as well as classical PCA*.
- * R (statistical language environment) and the ChemoSpec package (Developed by Dr. Bryan Hanson) are required

Solutions for Innovation JEOL

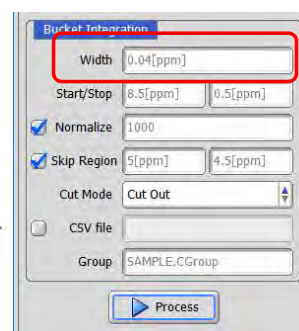
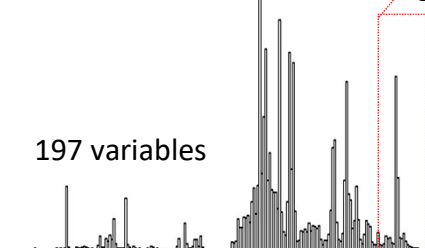
Bucket Integration in the Chemometrics tool (1/2)

Original data
(more than 16,384 pnt)



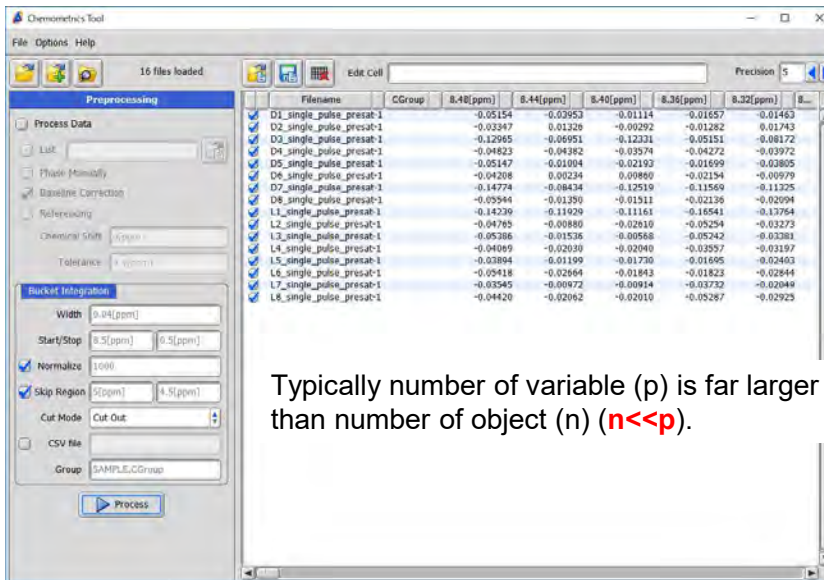
Bucket integration

197 variables



Solutions for Innovation JEOL

Bucket Integration in the Chemometrics tool (2/2)



The screenshot shows the Chemometrics Tool interface. On the left, the 'Preprocessing' panel is active, with the 'Bucket Integration' section expanded. The settings are: Width: 0.04[ppm], Start/Stop: 8.5[ppm] | 6.5[ppm], Normalize: 1000, Skip Region: 5[ppm] | 4.5[ppm], Cut Mode: Cut Out, and Group: SAMPLE.CGroup. The 'Process' button is visible at the bottom of the panel. The main window displays a table with 16 rows of data, each representing a file (e.g., D1_single_pulse_presat-1) and its corresponding values across five columns: 8.48[ppm], 8.44[ppm], 8.40[ppm], 8.36[ppm], and 8.32[ppm].

Filename	CGroup	8.48[ppm]	8.44[ppm]	8.40[ppm]	8.36[ppm]	8.32[ppm]
D1_single_pulse_presat-1		-0.05154	-0.03953	-0.01114	-0.01657	-0.01463
D2_single_pulse_presat-1		-0.02947	0.01329	-0.00292	-0.01282	0.01743
D3_single_pulse_presat-1		-0.12945	-0.04951	-0.12371	-0.05151	-0.00172
D4_single_pulse_presat-1		-0.04823	-0.04382	-0.03574	-0.04272	-0.03972
D5_single_pulse_presat-1		-0.05147	-0.01004	-0.02193	-0.01699	-0.03805
D6_single_pulse_presat-1		-0.04208	0.00234	0.09860	-0.02154	-0.00979
D7_single_pulse_presat-1		-0.14774	-0.00434	-0.12519	-0.11569	-0.11225
D8_single_pulse_presat-1		-0.05944	-0.01350	-0.01511	-0.02136	-0.02094
L1_single_pulse_presat-1		-0.14239	-0.11929	-0.11161	-0.16541	-0.13764
L2_single_pulse_presat-1		-0.04765	-0.00880	-0.02610	-0.05254	-0.03273
L3_single_pulse_presat-1		-0.05206	-0.01535	-0.00568	-0.05242	-0.03281
L4_single_pulse_presat-1		-0.04069	-0.02030	-0.02040	-0.03557	-0.03197
L5_single_pulse_presat-1		-0.03894	-0.01199	-0.01730	-0.01665	-0.02403
L6_single_pulse_presat-1		-0.05418	-0.02664	-0.01843	-0.01823	-0.02844
L7_single_pulse_presat-1		-0.03545	-0.00972	-0.00914	-0.01722	-0.02040
L8_single_pulse_presat-1		-0.04420	-0.02062	-0.02010	-0.05287	-0.02923

Typically number of variable (p) is far larger than number of object (n) ($n \ll p$).

Solutions for Innovation JEOL

Outline

Introduction (done)

Workflow of NMR Non-targeted analysis

Chemometrics tool in the Delta NMR software

Unsupervised Multivariate Analyses in the Chemometrics tool

Classical and Robust Principle Component Analysis

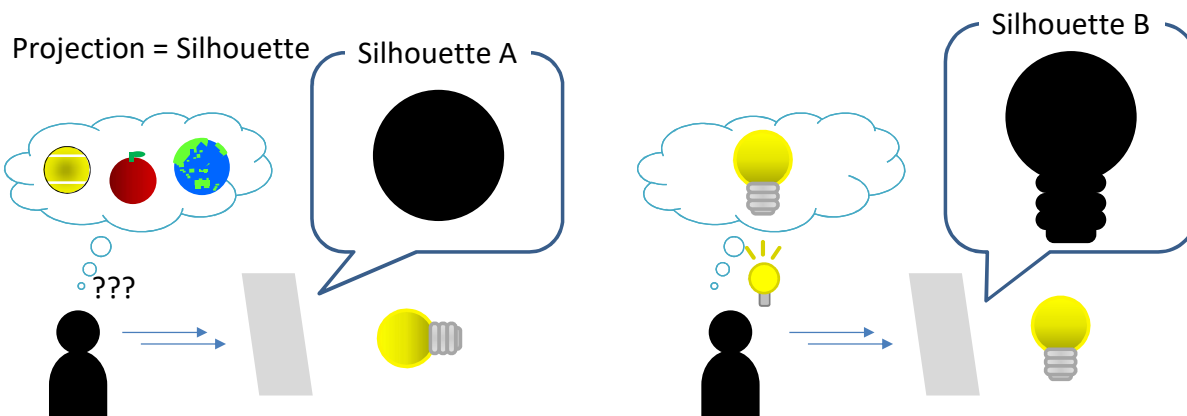
Hierarchical Cluster Analysis

Application Example

Solutions for Innovation JEOL

What is PCA?

PCA projects multivariate data set into the new orthogonal coordinate system.



Here silhouettes are 2D projections of (3D) shape of a light bulb.

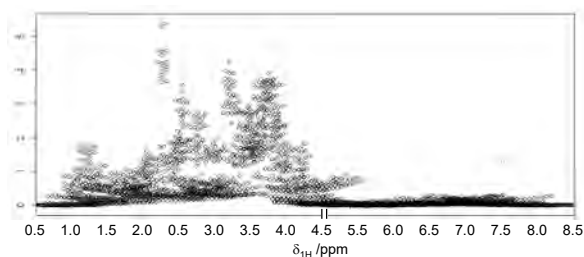
Analogous to this, PCA projects high dimensional data into lower dimensional space.

How does PCA select “projecting angle”? → Maximizing variance of a projection.

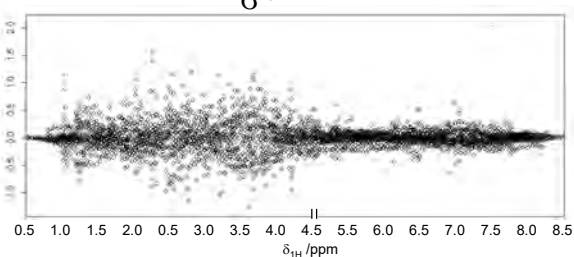
Solutions for Innovation JEOL

Data preprocessing (Scaling each variable)

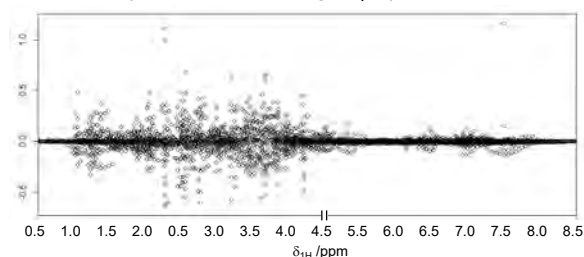
Original data (x_i)



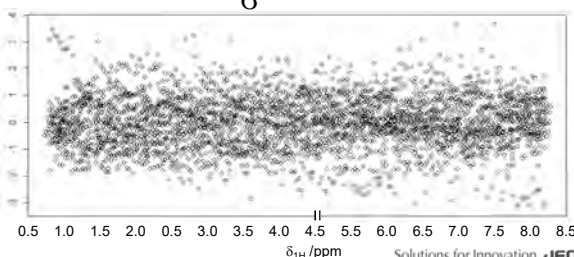
Pareto scale $\frac{x_i - \bar{x}}{\sigma^{1/2}}$



No scale (mean centering, $x_i - \bar{x}$)



Auto scale $\frac{x_i - \bar{x}}{\sigma}$

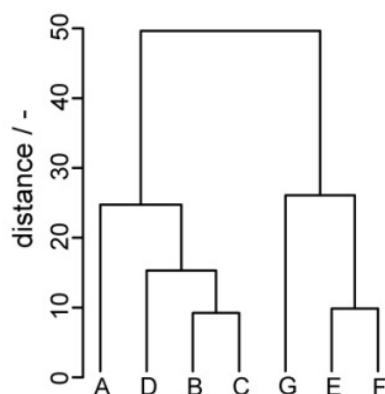


Solutions for Innovation JEOL

What is HCA (Hierarchical Cluster Analysis) ?

HCA, a clustering method, builds a hierarchical cluster which shows similarity based on particular distance between samples (or variables).

HCA connects the closet distance pair (observation or cluster) until whole data become one dendrogram.



Solutions for Innovation JEOL

Outline

Introduction (done)

Workflow of NMR Non-targeted analysis

Chemometrics tool in the Delta NMR software

Unsupervised Multivariate Analyses in the Chemometrics tool

Classical and Robust Principle Component Analysis

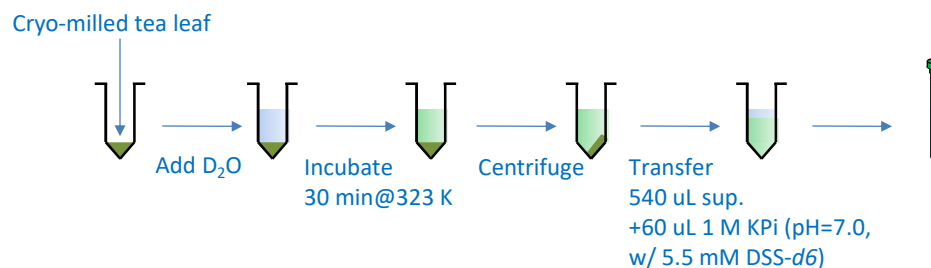
Hierarchical Cluster Analysis

Application Example

Solutions for Innovation JEOL

Exploratory Analysis of Japanese Green Tea

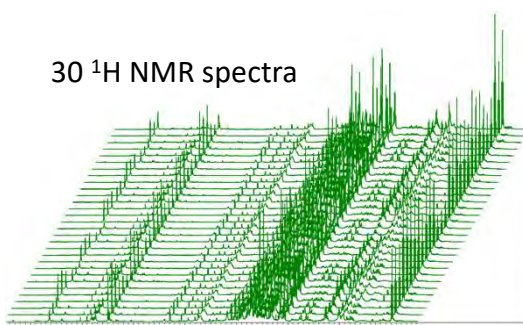
- 10 commercially available Japanese green tea leaves were purchased in a tea store.
- 5 teas are *gyokuro* tea, specialty tea (Tea trees are covered for several weeks before harvest in the production).
- Other 5 teas are *sencha* tea, normal tea.



- Three replicates were prepared for each tea leaf; therefore totally 30 NMR samples were prepared.
- ¹H-NMR spectra (with presaturation) were recorded with JNM-ECZ500R system equipped with ROYALPROBE™ and ASC30.

Solutions for Innovation JEOL

Convert ¹H NMR spectra into matrix by bucket integration

30 ¹H NMR spectra


bucket integration

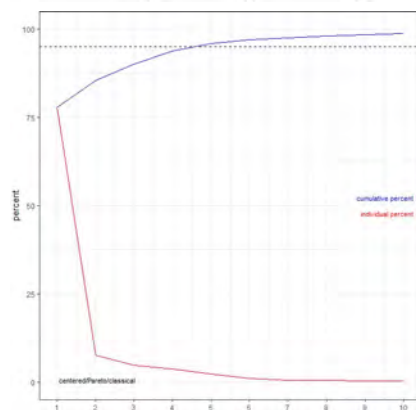
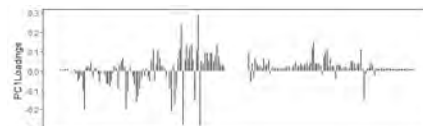
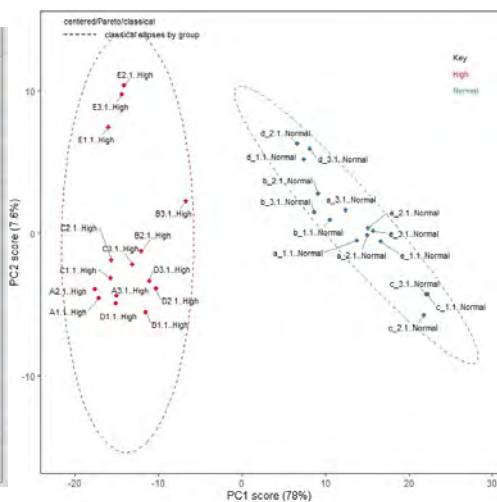
The screenshot shows the ChemMatrix Tool interface with a list of 30 NMR samples. The table below represents the data shown in the interface, where each row corresponds to a sample and its chemical shift values in ppm.

Filename	CGroup	0.40[ppm]	0.80[ppm]	1.60[ppm]	3.00[ppm]	4.00[ppm]	5.00[ppm]
A1-1	High	-0.14487	-0.10563	-0.14169	-0.14418	-0.12576	-0.14636
A2-1	High	-0.14268	-0.13822	-0.14168	-0.14064	-0.13625	-0.14281
A3-1	High	-0.11817	-0.11678	-0.10242	-0.13369	-0.12563	-0.12811
a_1-1	Normal	-0.07054	-0.05540	-0.07311	-0.06644	-0.05213	-0.05448
a_2-1	Normal	-0.07151	-0.04879	-0.05370	-0.05982	-0.07025	-0.05811
a_3-1	Normal	-0.05876	-0.03192	-0.03702	-0.07223	-0.05679	-0.04806
B1-1	High	-0.08956	-0.09488	-0.10488	-0.09881	-0.08029	-0.10751
B2-1	High	-0.10690	-0.11145	-0.09388	-0.12972	-0.09644	-0.09397
B3-1	High	-0.12764	-0.12238	-0.12752	-0.12917	-0.13128	-0.13200
b_1-1	Normal	-0.09547	-0.08001	-0.10562	-0.11081	-0.08729	-0.10301
b_2-1	Normal	-0.13064	-0.09038	-0.09250	-0.13164	-0.09895	-0.08834
b_3-1	Normal	-0.07053	-0.05268	-0.05132	-0.03965	-0.05952	-0.04521
C1-1	High	-0.15156	-0.14153	-0.13808	-0.15979	-0.13871	-0.15214
C2-1	High	-0.17758	-0.14553	-0.16990	-0.16405	-0.14520	-0.15471
C3-1	High	-0.15033	-0.12739	-0.14360	-0.13054	-0.12015	-0.13084
c_1-1	Normal	-0.08053	-0.08015	-0.06540	-0.07867	-0.07920	-0.07255
c_2-1	Normal	-0.09167	-0.08169	-0.07013	-0.07271	-0.06613	-0.06721
c_3-1	Normal	-0.11183	-0.09855	-0.08666	-0.10308	-0.10475	-0.10671
D1-1	High	-0.07895	-0.07697	-0.09140	-0.10734	-0.06855	-0.09421
D2-1	High	-0.12886	-0.12701	-0.13376	-0.13756	-0.10996	-0.15001
D3-1	High	-0.09189	-0.08461	-0.08063	-0.06983	-0.06598	-0.06805
d_1-1	Normal	-0.08195	-0.07657	-0.10424	-0.10319	-0.10536	-0.09721
d_2-1	Normal	-0.14997	-0.15586	-0.14887	-0.16625	-0.14654	-0.15524
d_3-1	Normal	-0.12058	-0.12305	-0.14597	-0.12962	-0.15515	-0.15156
E1-1	High	-0.19180	-0.17359	-0.18259	-0.20628	-0.17313	-0.18714
E2-1	High	-0.26500	-0.24697	-0.24170	-0.22348	-0.22187	-0.25191
E3-1	High	-0.19238	-0.17514	-0.18807	-0.18689	-0.18636	-0.19551
e_1-1	Normal	-0.05338	-0.02498	-0.01893	-0.05299	-0.03630	-0.05454
e_2-1	Normal	-0.04362	-0.04178	-0.04177	-0.04646	-0.01974	-0.02223
e_3-1	Normal	-0.09054	-0.05230	-0.07514	-0.09536	-0.08775	-0.07601

30*197 matrix

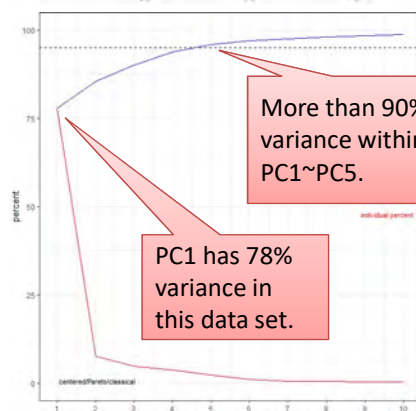
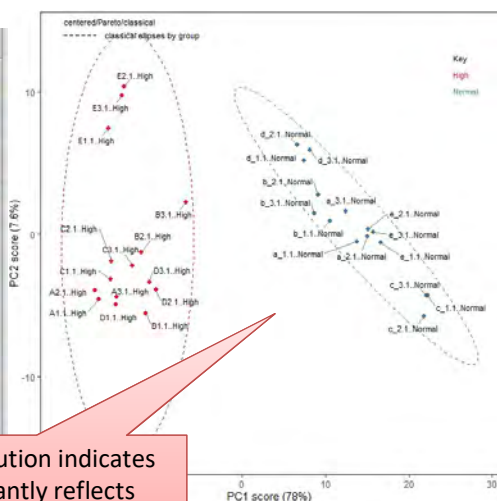
Solutions for Innovation JEOL

[Result] PCA



Classical PCA was performed with “Pareto scaling”. (Normalized with total sum)

[Result] PCA

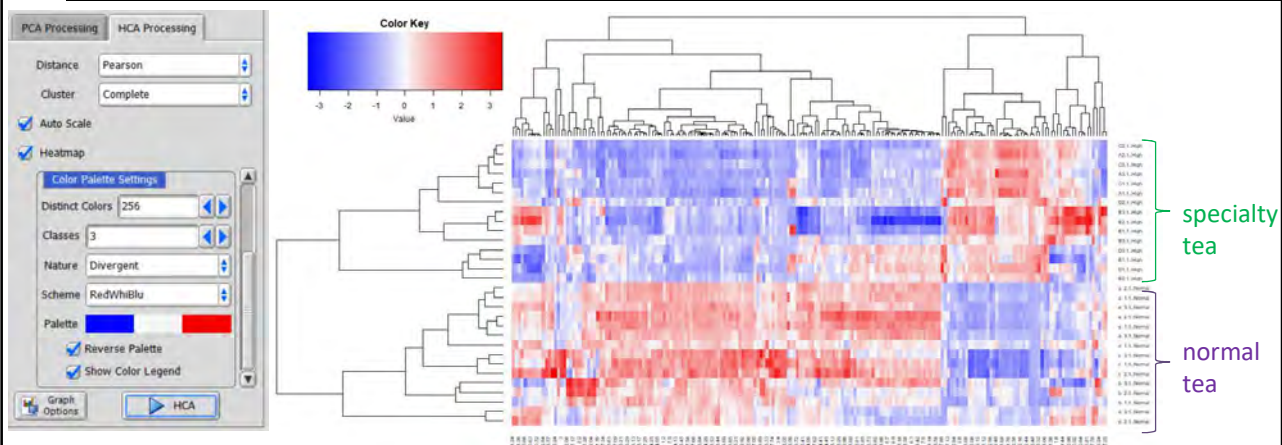


Sample distribution indicates PC1 predominantly reflects sample group differences (specialty or normal)

PC1 has 78% variance in this data set.

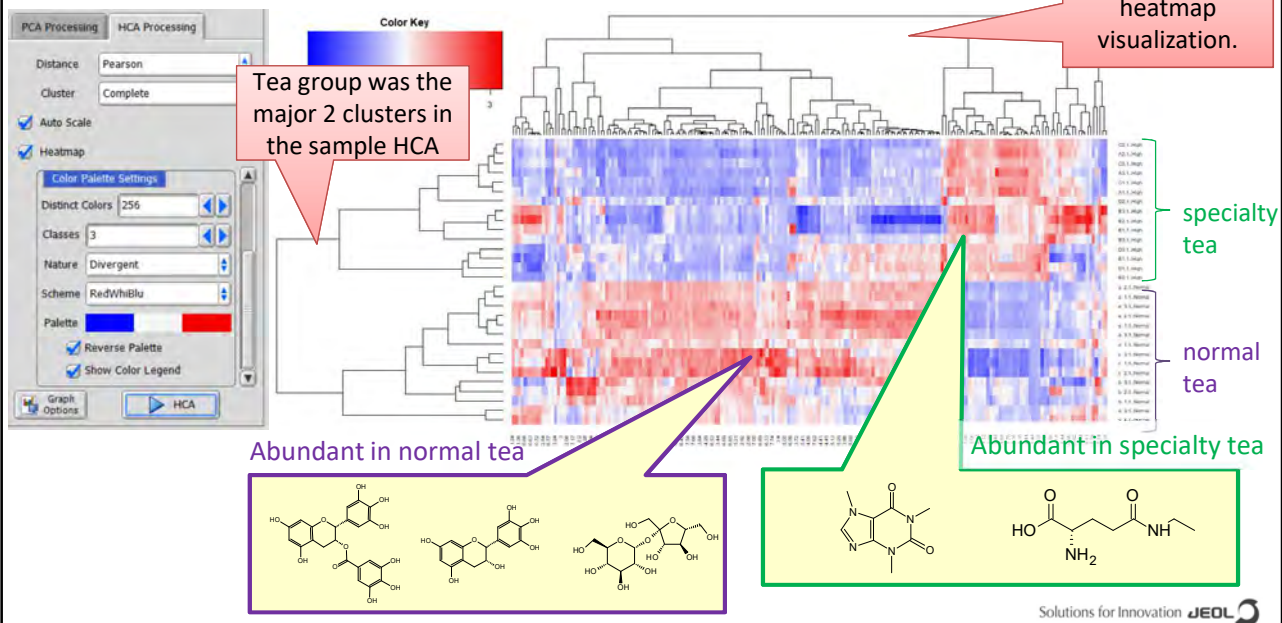
More than 90% variance within PC1~PC5.

[Result] HCA



HCA was performed using "Person dissimilarity" and "Complete clustering method".
(Variables were scaled by variance (auto scale))

[Result] HCA



Summary

NMR is robust analytical platform for complex mixture analysis.

Multivariate analyses are typically employed for non-targeted analysis strategies with NMR data. The Chemometrics tool in Delta NMR software Ver. 6.0 (or later) supports PCA and HCA, unsupervised analyses.

Multivariate analyses are based on "simple" mathematics. Not mysterious.

The chemometrics tool is "Gateway to Chemometrics". Not only PCA and HCA but various data analysis are available with other both commercial and free software.

Acknowledgement

Dr. Bryan Hanson

-Prof. Emeritus of Chemistry & Biochemistry at DePauw University

Developing and maintaining the "ChemoSpec" package

