BEER

Current quality assessment:

appearance: color, foam, clarity
taste: sweetness, sourness, saltiness, bitterness
flavor, aroma

Factors affecting chemical composition:

water quality, malt, hop, yeasts
recipe and timing of the brewing process

Motivation:

the relationship of the current quality properties with chemical composition is not fully understood

JOURNAL OF
AGRICULTURAL AND
FOOD CHEMISTRY

Composition of Beer by ^{1}H NMR Spectroscopy: Effects of Brewing Site and Date of Production

CLÁUDIA ALMEIDA,[†] IOLA F. DUARTE,[†] ANTÓNIO BARROS,[‡] JOÃO RODRIGUES,[†]
MANFRED SPRAUL,[§] AND ANA M. GIL*,[†]

CICECO and QOPNAA, Department of Chemistry, Campus Universitário de Santiago,
University of Aveiro, 3810-193 Aveiro, Portugal, and Bruker Biospin GmbH, Silberstreifen,
D76287 Rheinstetten, Germany

0-3 ppm
alcohols (propanol, isobutanol, isopentanol)
organic acids (citric, malic, pyruvic, acetic, succinic)
amino acids (alanine, g-aminobutyric, proline)

3-6 ppm
fermentable sugars (glucose, malrose)
dextrins (glucose oligomers)

6-10 ppm
aromatic amino acids (tyrosine, phenylalanine, tryptophan)
nucleosides (cytidine, uridine, adenosine/inosine)
aromatic alcohols (2-phenylethanol, tyrosol, tryptophol)
polyphenolic compounds

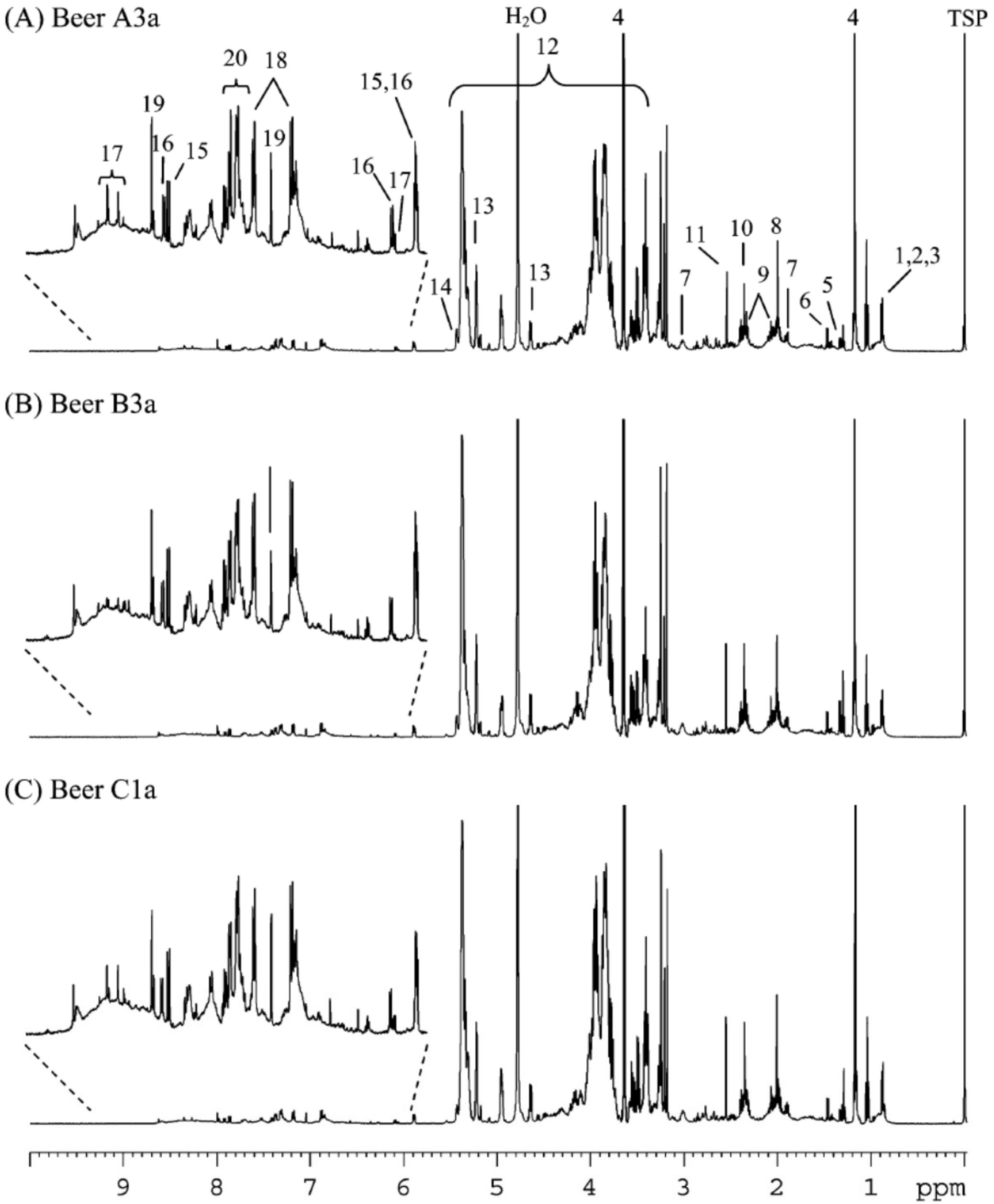


Figure 1. 500 MHz ^{1}H NMR spectra of (A) beer A3a, (B) beer B3a, and (C) beer C1a (named according to Table 1): 1, propanol; 2, isobutanol; 3, isopentanol; 4, ethanol; 5, lactate; 6, alanine; 7, γ-butyric acid (GABA); 8, acetate; 9, proline; 10, pyruvate; 11, succinate; 12, dextrins; 13, glucose; 14, maltose; 15, uridine; 16, cytidine; 17, adenosine/inosine; 18, tyrosine and/or tyrosol; 19, histidine; 20, 2-phenylethanol.

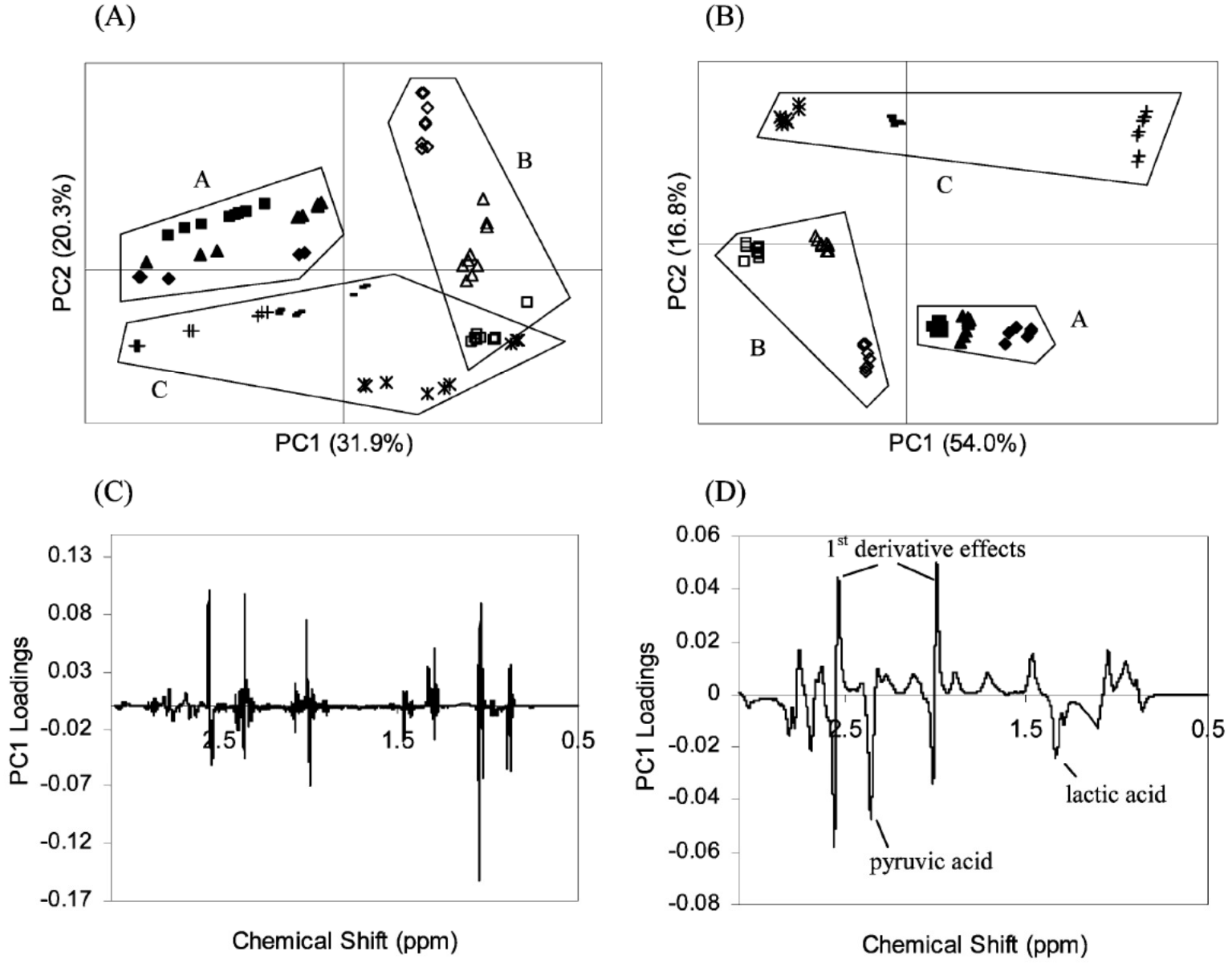


Figure 2. PCA of aliphatic NMR spectral regions (0.5–3.1 ppm): (A) scores scatter plot of PC1 vs PC2 for spectra processed with LB 0.3 Hz; (B) scores scatter plot of PC1 vs PC2 for spectra processed with LB 10 Hz; (C) PC1 loadings profile corresponding to part A; (D) PC1 loadings profile corresponding to part B. Site A beers: (◆) A1; (■) A2; (▲) A3. Site B beers: (◇) B1; (□) B2; (△) B3. Site C beers: (∗) C1; (–) C2; (+) C3. Grouping shapes were drawn manually in the scores scatter plots to aid the eye. Peaks arising from lactic acid, pyruvic acid, and first derivative artifacts are indicated in part D.

Increased pyruvic acid levels : poor yeast quality

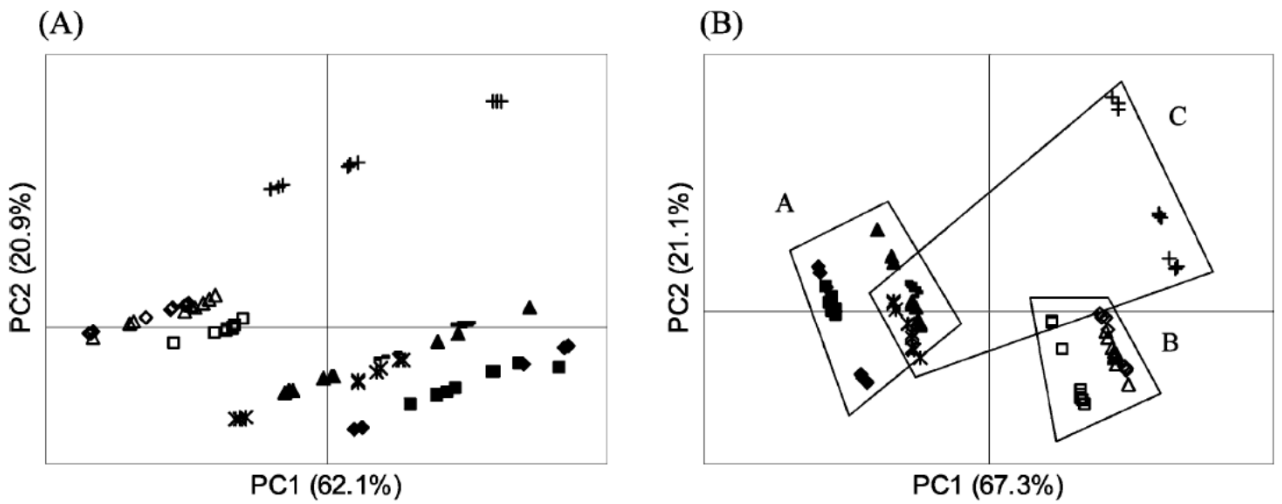


Figure 3. PCA of sugar NMR spectral regions (3.1–5.8 ppm): (A) scores scatter plot of PC1 vs PC2 for spectra processed with LB 0.3 Hz; (B) scores scatter plot of PC1 vs PC2 for spectra processed with LB 10 Hz. Site A beers: (◆) A1; (■) A2; (▲) A3. Site B beers: (◇) B1; (□) B2; (△) B3. Site C beers: (∗) C1; (–) C2; (+) C3. Grouping shapes were drawn manually in the scores scatter plots to aid the eye.

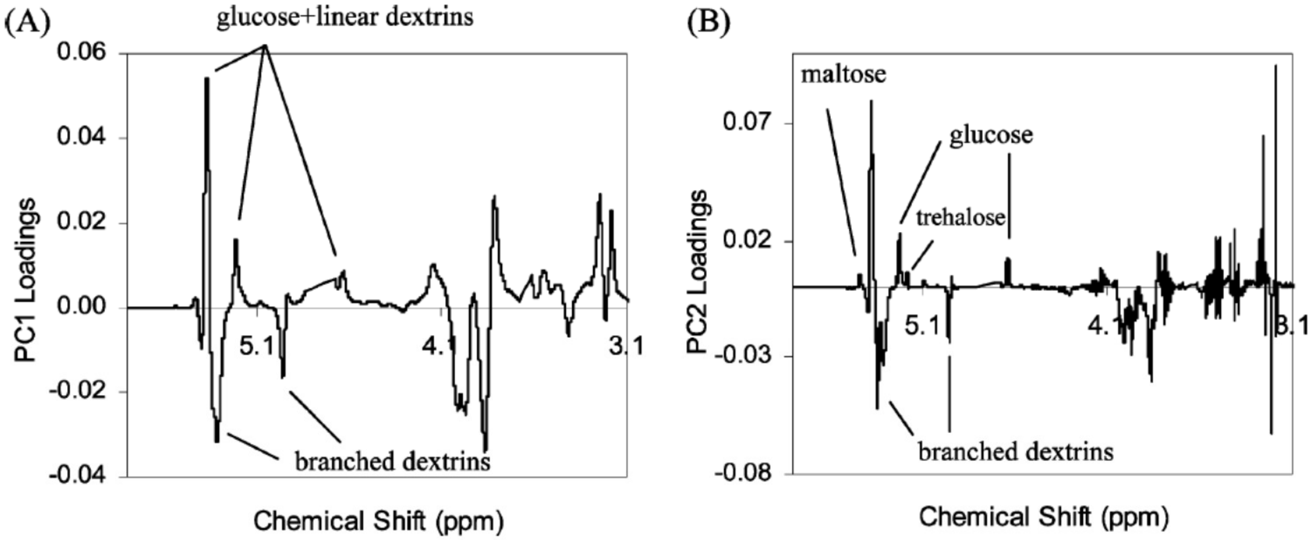
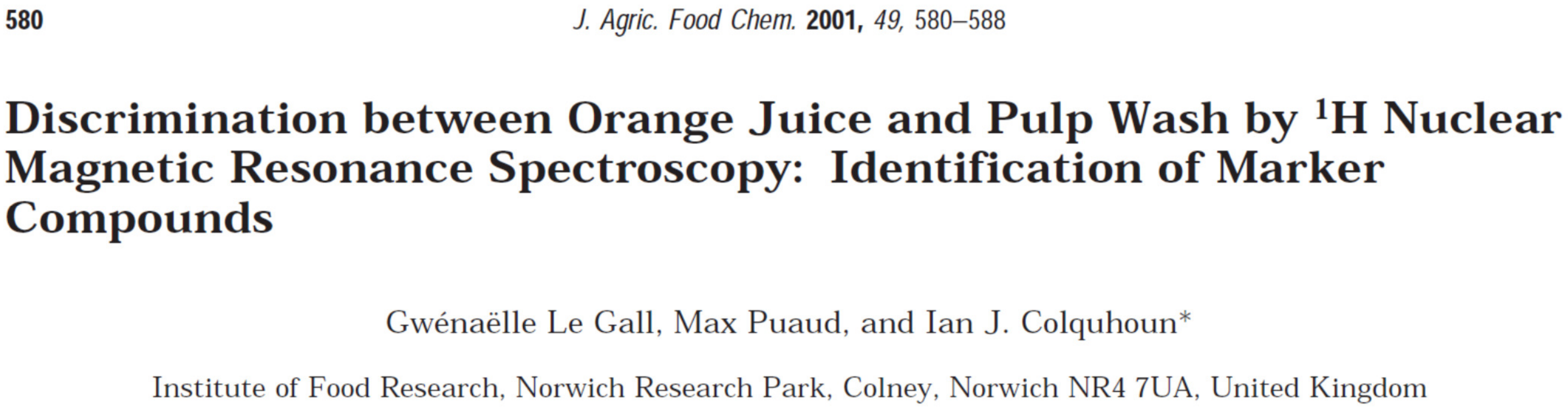


Figure 4. PCA of sugar NMR spectral regions (3.1–5.8 ppm): (A) PC1 loadings profile for spectra processed with LB 10 Hz; (B) PC2 loadings profile for spectra processed with LB 0.3 Hz. The main peaks responsible for variations in positive and negative PC1 and PC2 are indicated and assigned (glc, glucose; mal, maltose; tre, trehalose).

Linear vs branched dextrins correlate with fine conditions during malting and mashing

ORANGE JUICE



580 *J. Agric. Food Chem.* **2001,** *49,* 580–588

Discrimination between Orange Juice and Pulp Wash by ^{1}H Nuclear Magnetic Resonance Spectroscopy: Identification of Marker Compounds

Gwénaëlle Le Gall, Max Puaud, and Ian J. Colquhoun*

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom

According to U.S. Food and Drug Administration (FDA) investigations, some companies are known to have made millions of dollars selling fraudolent orange juice

Adulteration may be done by the addition of water, sugars, pulp wash, or other citrus fruit juices

Pulp wash is a second extract obtained by washing the separated pulp with water after the first pressing. Its chemical composition is similar to orange juice but paler, more bitter, and is regarded as lower quality

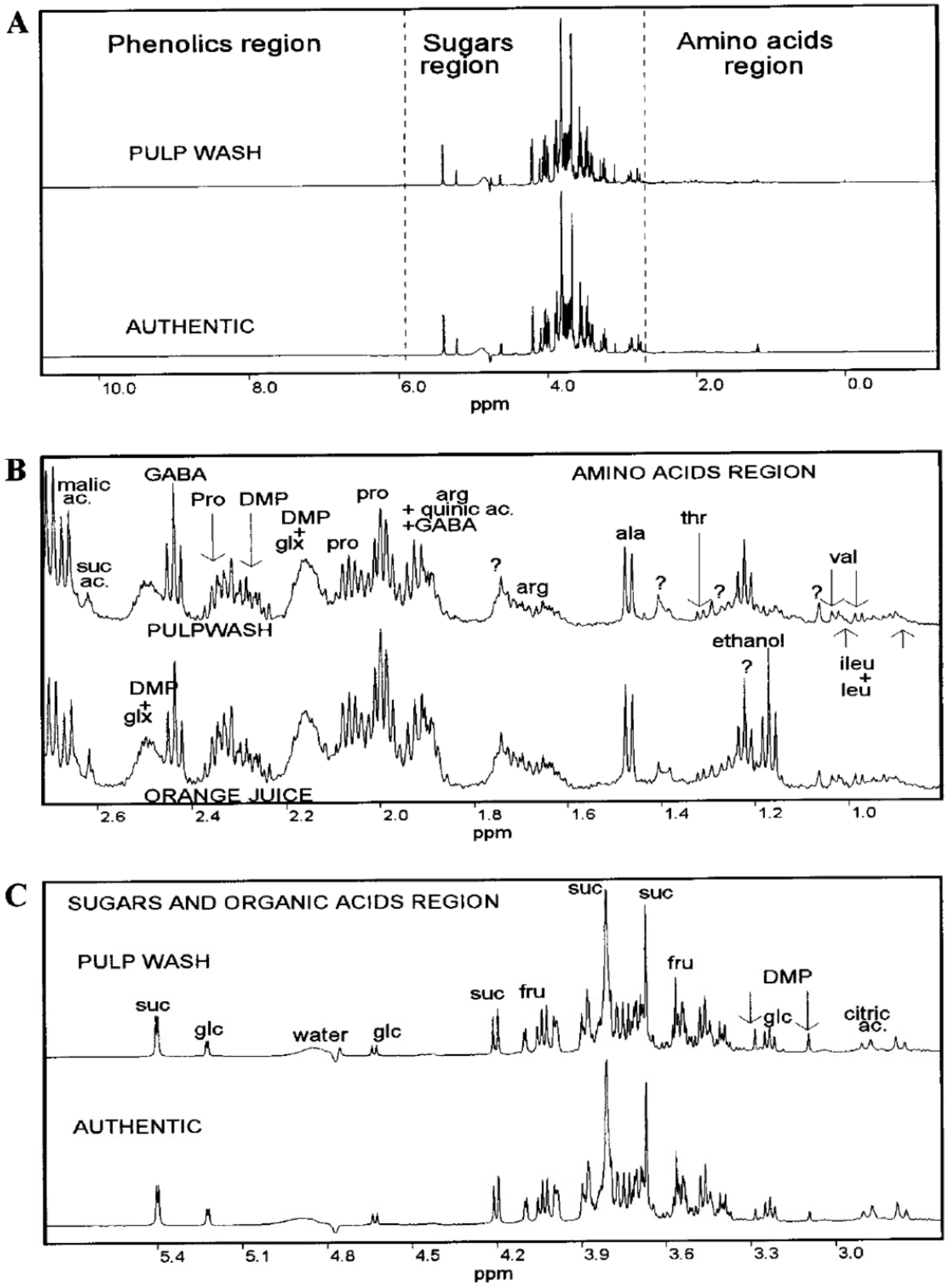


Figure 1. (A) 500 MHz ^{1}H NMR spectra of typical orange juice and pulp wash samples (overall view); (B) expanded view, high-field region (key: suc. ac., succinic acid; GABA, γ-aminobutyric acid; glx, glutamine/glutamic acid; DMP, dimethylproline); (C) midfield region (key: suc, sucrose; glc, glucose; fru, fructose).

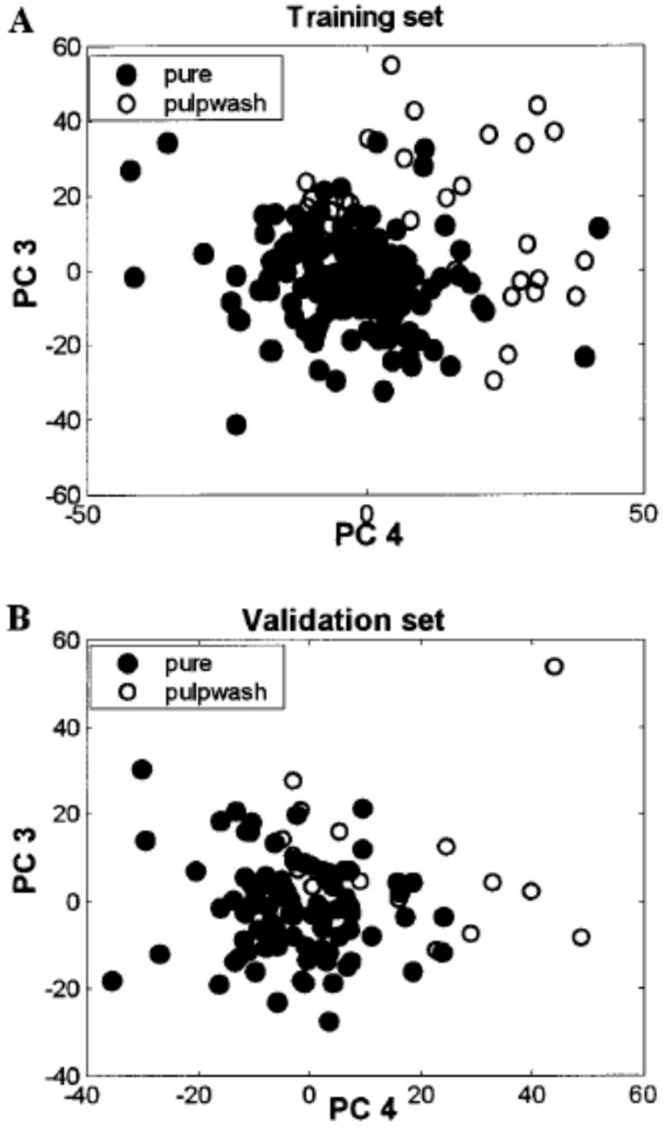


Figure 3. PC scores on the first two PC axes (PC4 and PC3) selected by the stepwise LDA procedure: (A) scores from PCA of the training set spectra; (B) scores for the validation set calculated using the training set PC loadings.

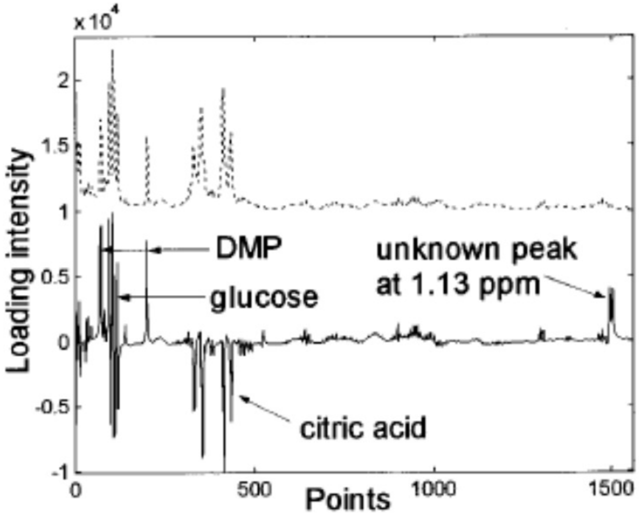
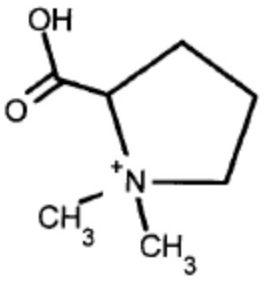


Figure 4. PC4 loading (lower solid line) and mean spectrum of all samples (upper dashed line), high-field region.



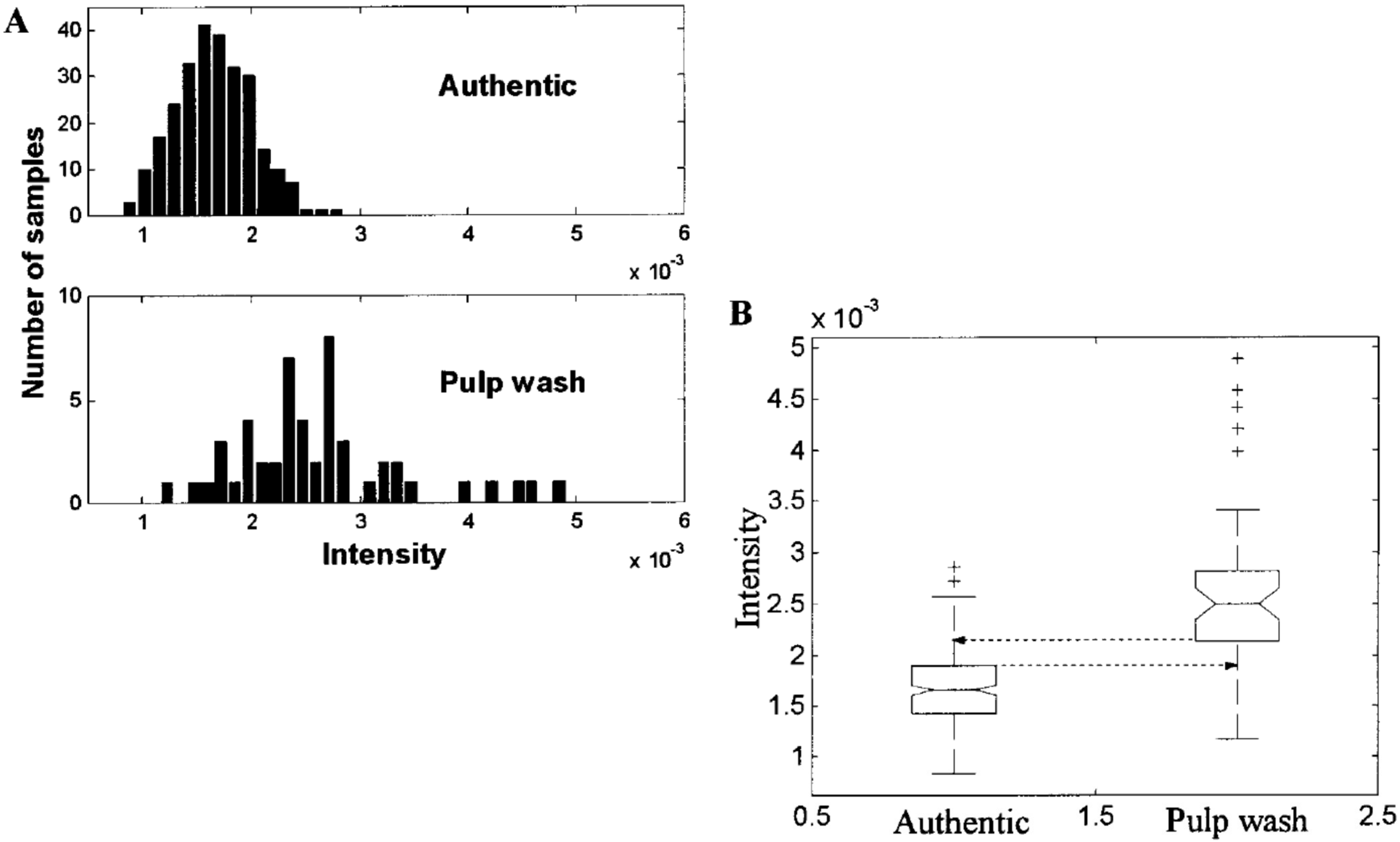


Figure 5. Signal intensity distributions for methyl signal (3.28 ppm) of DMP, comparison of orange juice ("authentic") and pulp wash groups: (A) histograms; (B) box plots showing medians, confidence intervals (notches), range, and outliers (+).

BLACK TEA

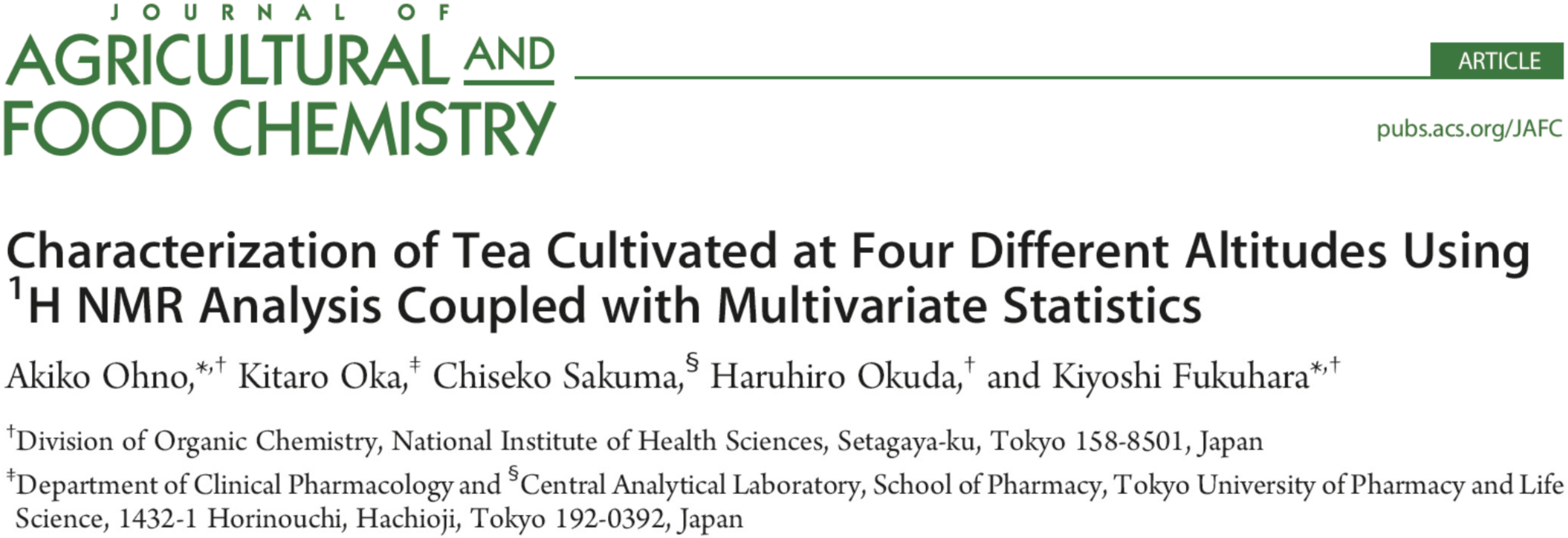
White, green, oolong, black tea differ in the fermentation process:

green: unfermented
white: lightly f.
oolong: partially
black: fermented

All derive from *Camellia sinensis*

Black tea is more oxidized, has stronger flavor, and contains more caffeine
Drinking black tea is associated with reduced cardiovascular risk

During manufacture, enzyme-catalyzed oxidation and partial polymerization of flavonols occur. As a result, theaflavins (TFs) and thearubigins characteristic of the black tea taste and color are produced.
Flavonoids constitute 10-12% of dry leaf weight.
The taste differs according to differences in growing environment.



JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Characterization of Tea Cultivated at Four Different Altitudes Using ^{1}H NMR Analysis Coupled with Multivariate Statistics

Akiko Ohno,*,[†] Kitaro Oka,[‡] Chiseko Sakuma,[§] Haruhiro Okuda,[†] and Kiyoshi Fukuhara*,[†]

[†]Division of Organic Chemistry, National Institute of Health Sciences, Setagaya-ku, Tokyo 158-8501, Japan

[‡]Department of Clinical Pharmacology and [§]Central Analytical Laboratory, School of Pharmacy, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan

Sri-Lanka tea-planting regions:
RAN (>1900m), UDA (1000-1500m), MEDA (600-1200m), YATA (<600m)

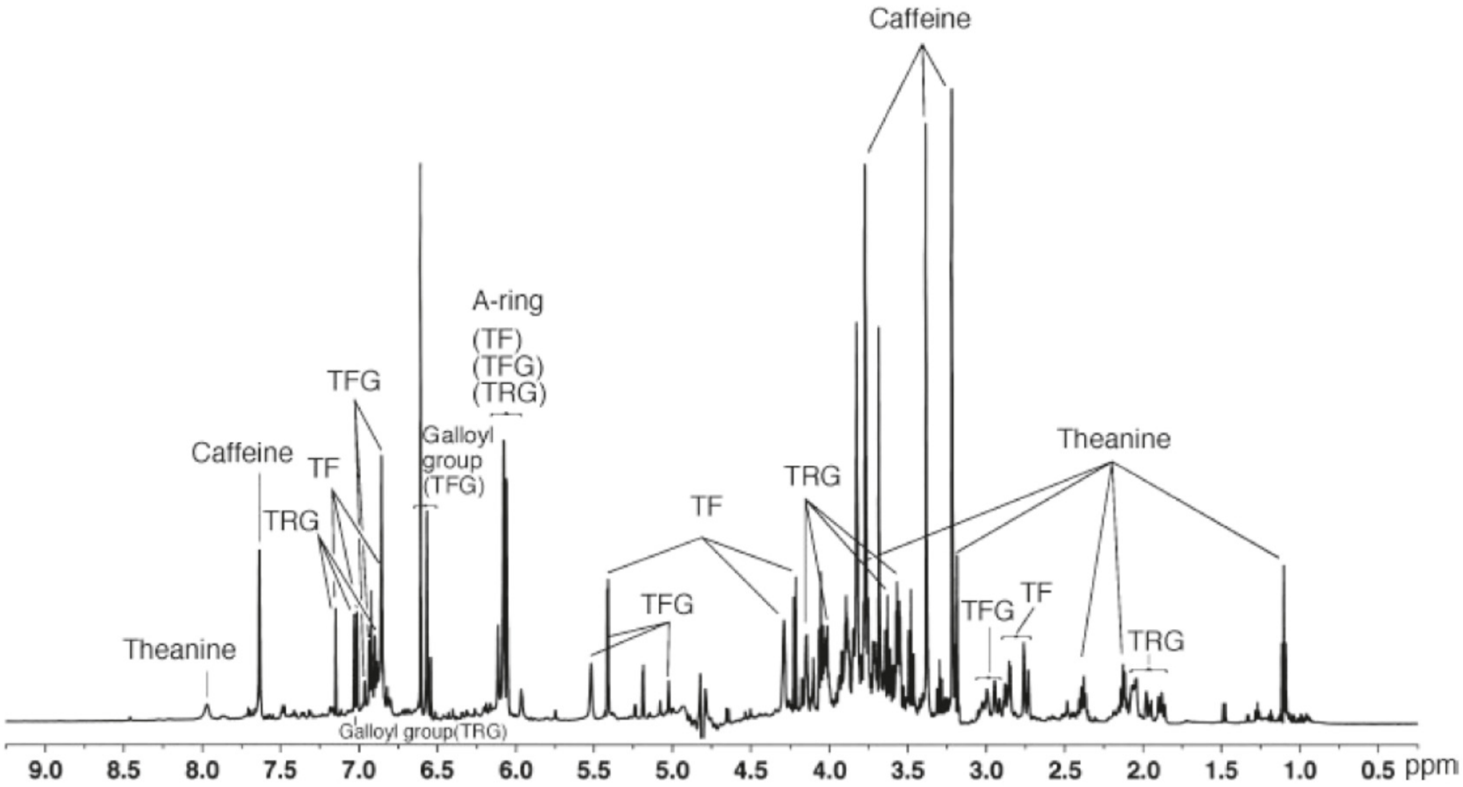


Figure 1. Representative ^{1}H NMR spectrum of black tea from RAN.

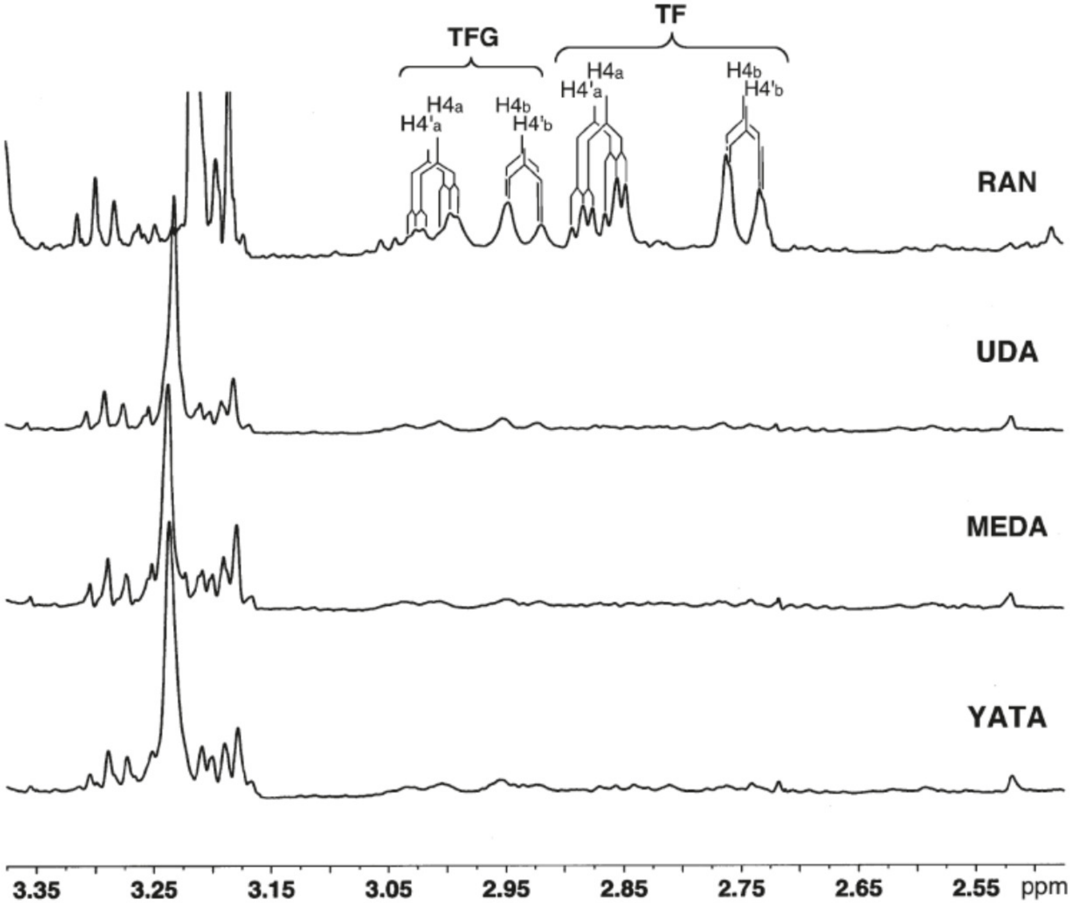
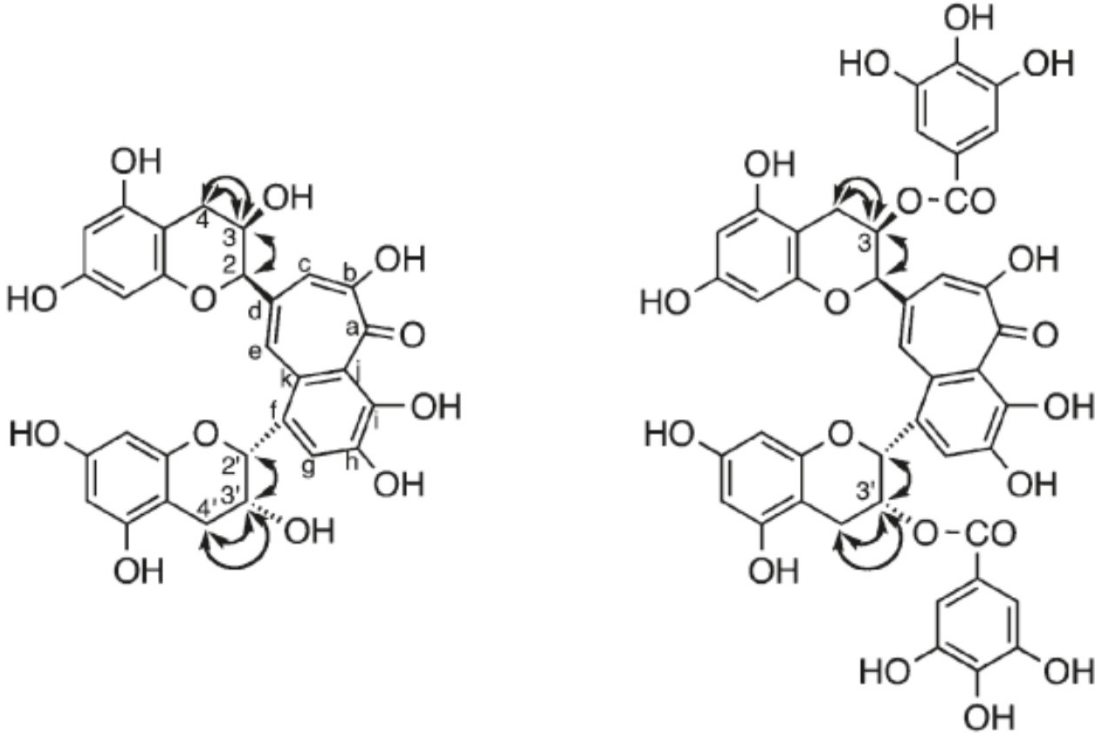


Figure 2. ^{1}H NMR spectrum expansion (2.55−3.35 ppm) of RAN, UDA, MEDA, and YATA.



Theaflavin (TF)

Theaflavin 3, 3'-digallate (TFG)

Figure 3. COSY and TOCSY (H ↔ H) correlations for TF and TFG.

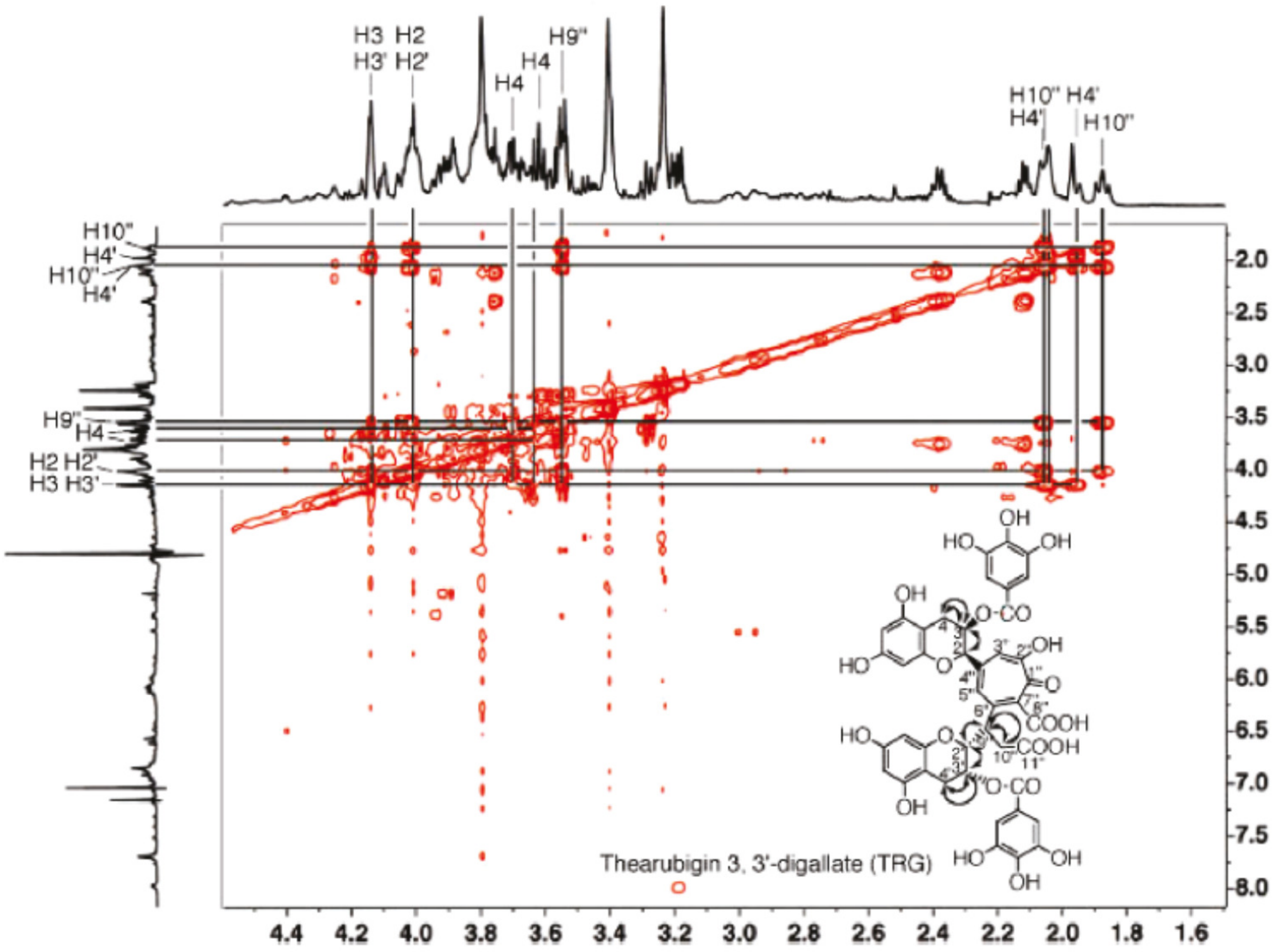


Figure 4. TOCSY (H ↔ H) correlations from the characteristic region of TRG in YATA.

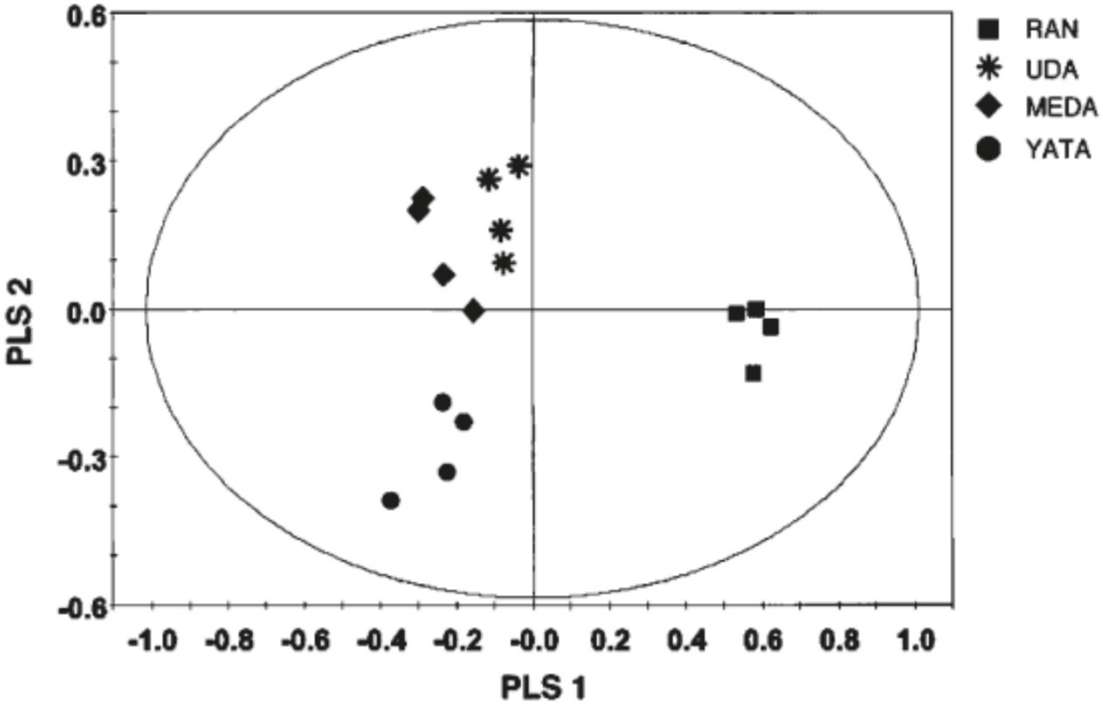


Figure 5. PLS-DA score plot derived from the ^{1}H NMR spectra of black teas from RAN (square), UDA (star), MEDA (diamond), and YATA (circle).

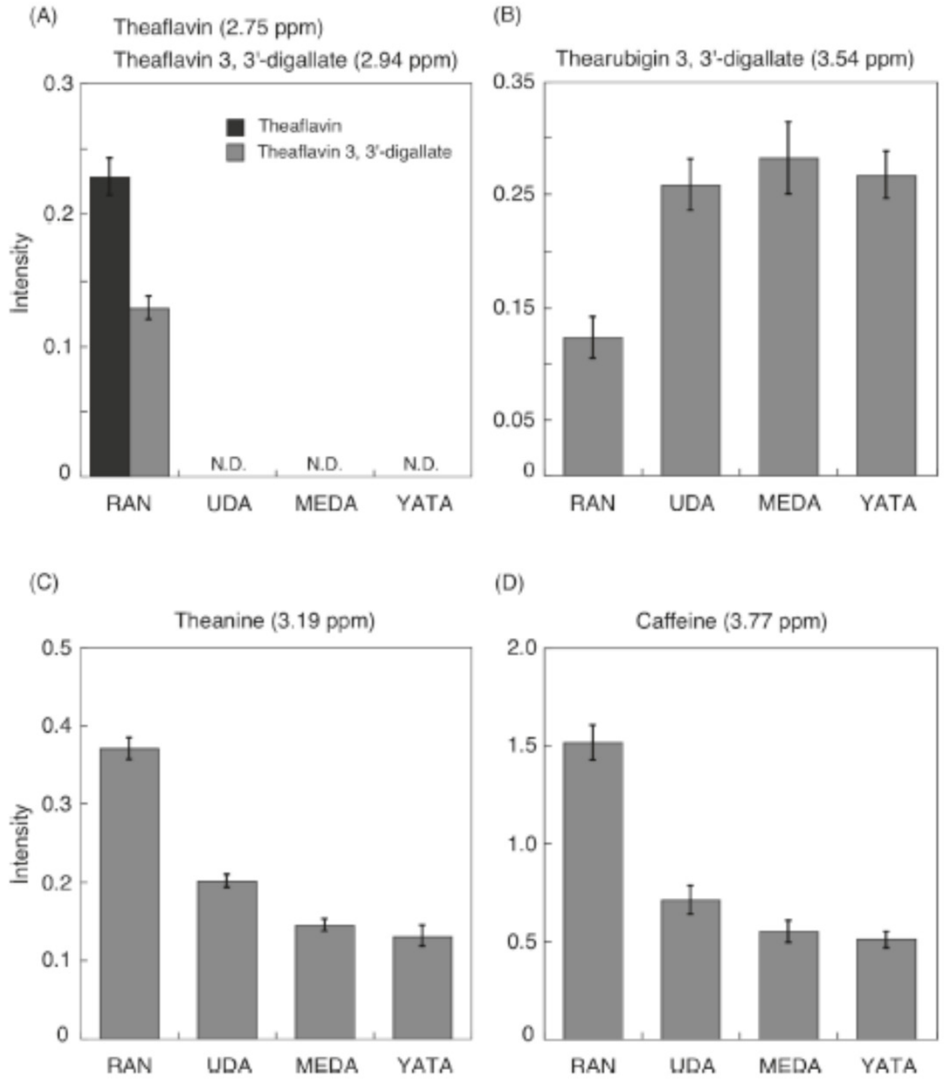


Figure 7. Different components from RAN, UDA, MEDA, and YATA. These components, identified from ^{1}H NMR spectra, are responsible for the differentiation in the PLS-DA model. (A) TF (2.75 ppm), TFG (2.94 ppm), (B) TRG (3.54 ppm), theanine (3.19 ppm), and caffeine (3.77 ppm).

OILS

OLII

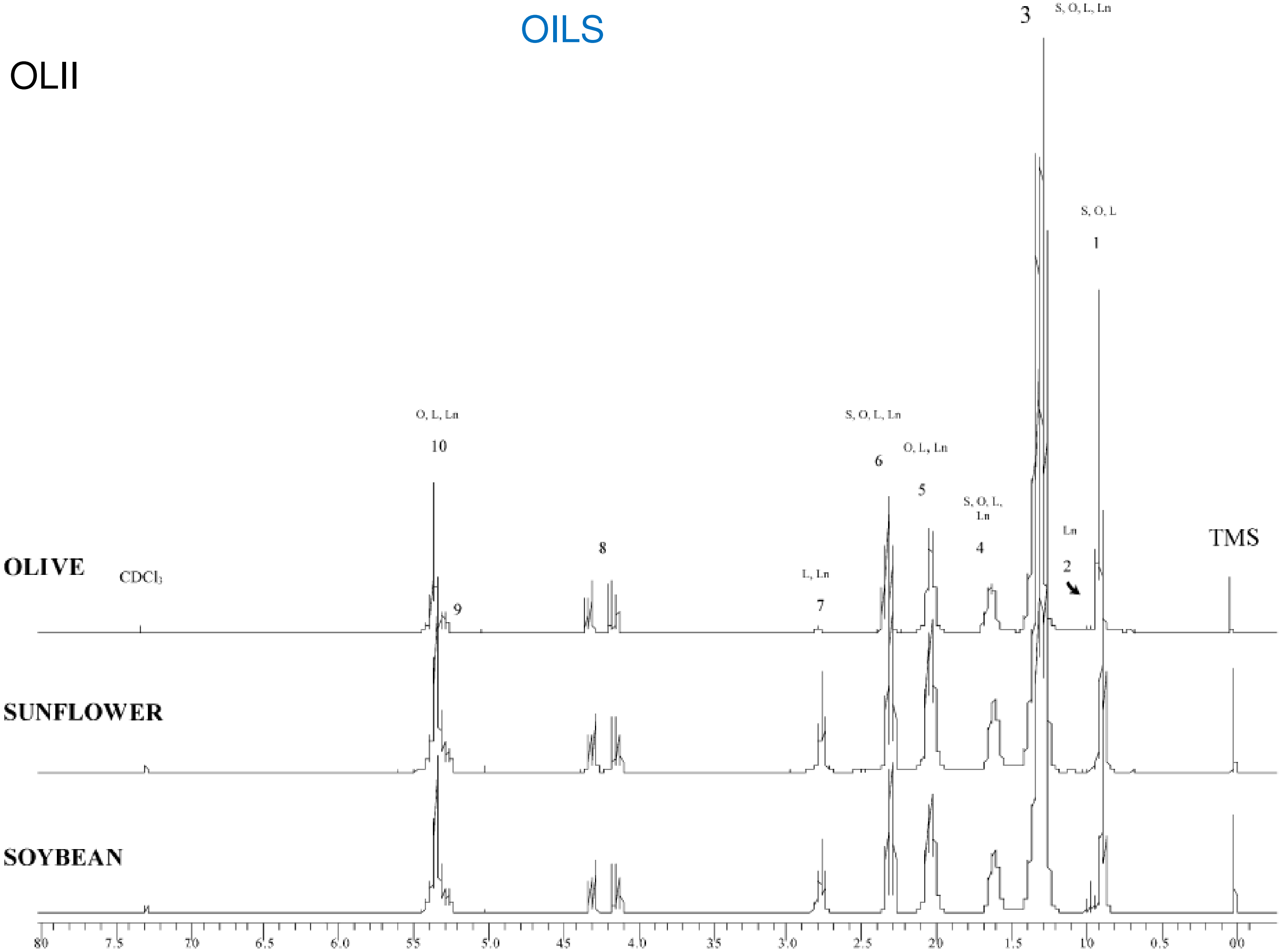


Fig. 1. ^{1}H NMR spectra of olive, sunflower and soybean oils. S, O, L and Ln refer to saturated, oleic, linoleic and linolenic acyl groups respectively. Signal numbering corresponds with that in Table 1.

OLII

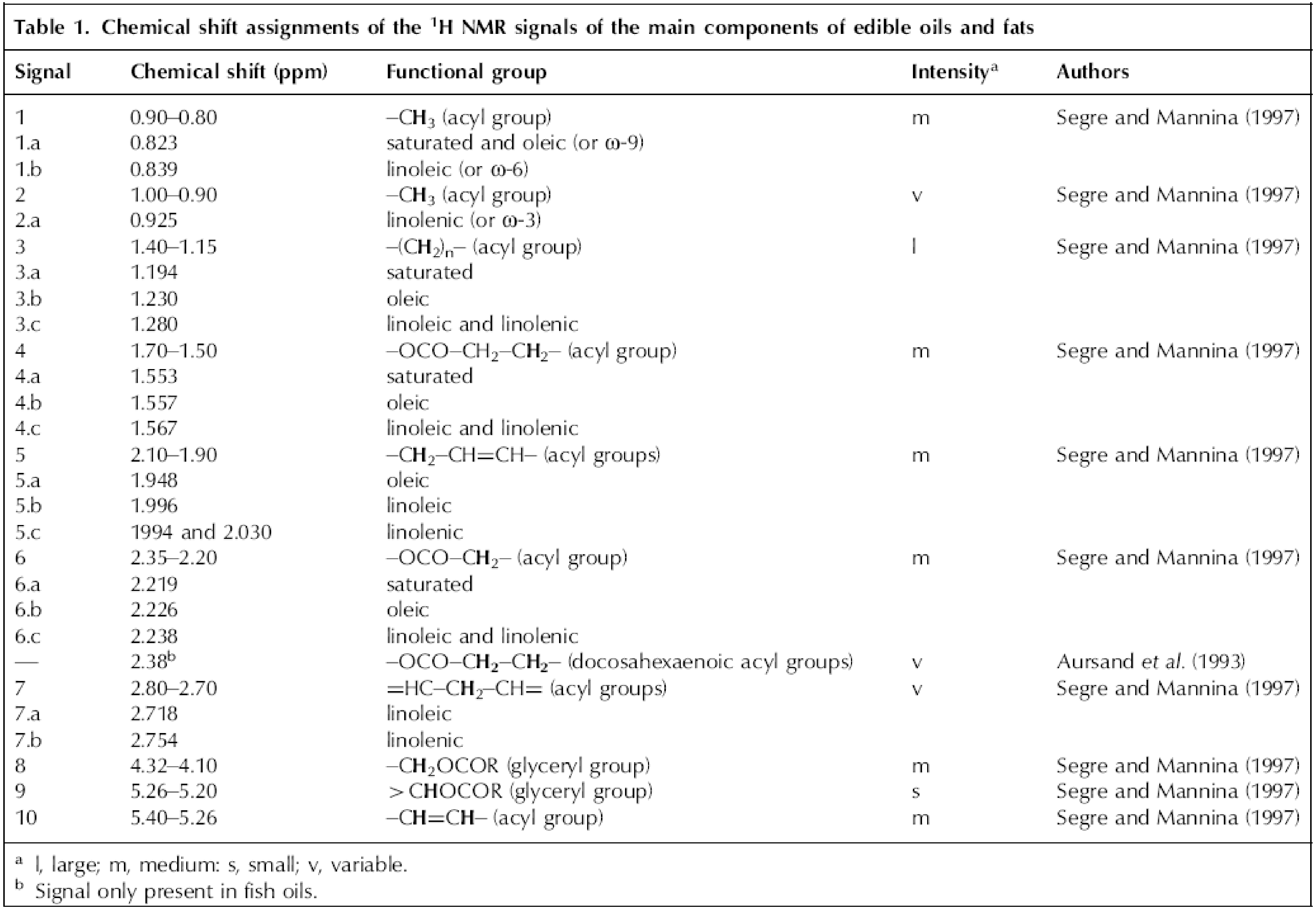


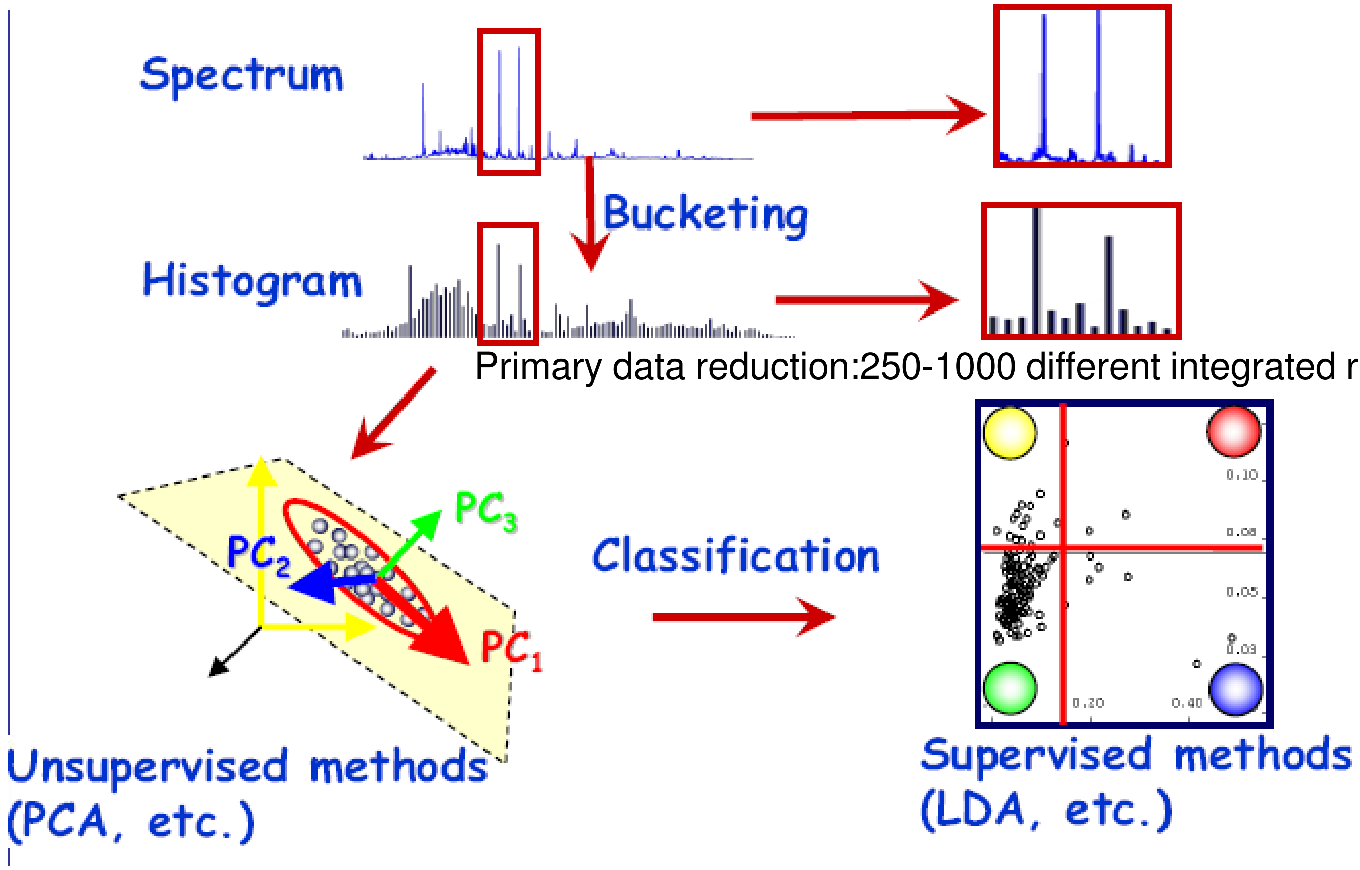
Table 1. Chemical shift assignments of the ^1H NMR signals of the main components of edible oils and fats

Signal	Chemical shift (ppm)	Functional group	Intensity[a]	Authors
1	0.90–0.80	$-CH_3$ (acyl group)	m	Segre and Mannina (1997)
1.a	0.823	saturated and oleic (or ω-9)		
1.b	0.839	linoleic (or ω-6)		
2	1.00–0.90	$-CH_3$ (acyl group)	v	Segre and Mannina (1997)
2.a	0.925	linolenic (or ω-3)		
3	1.40–1.15	$-(CH_2)_n-$ (acyl group)	l	Segre and Mannina (1997)
3.a	1.194	saturated		
3.b	1.230	oleic		
3.c	1.280	linoleic and linolenic		
4	1.70–1.50	$-OCO-CH_2-CH_2-$ (acyl group)	m	Segre and Mannina (1997)
4.a	1.553	saturated		
4.b	1.557	oleic		
4.c	1.567	linoleic and linolenic		
5	2.10–1.90	$-CH_2-CH=CH-$ (acyl groups)	m	Segre and Mannina (1997)
5.a	1.948	oleic		
5.b	1.996	linoleic		
5.c	1994 and 2.030	linolenic		
6	2.35–2.20	$-OCO-CH_2-$ (acyl group)	m	Segre and Mannina (1997)
6.a	2.219	saturated		
6.b	2.226	oleic		
6.c	2.238	linoleic and linolenic		
—	2.38[b]	$-OCO-CH_2-CH_2-$ (docosahexaenoic acyl groups)	v	Aursand *et al.* (1993)
7	2.80–2.70	$=HC-CH_2-CH=$ (acyl groups)	v	Segre and Mannina (1997)
7.a	2.718	linoleic		
7.b	2.754	linolenic		
8	4.32–4.10	$-CH_2OCOR$ (glyceryl group)	m	Segre and Mannina (1997)
9	5.26–5.20	$>CHOCOR$ (glyceryl group)	s	Segre and Mannina (1997)
10	5.40–5.26	$-CH=CH-$ (acyl group)	m	Segre and Mannina (1997)

[a] l, large; m, medium: s, small; v, variable.

[b] Signal only present in fish oils.

NMR-based metabolomics: the concept

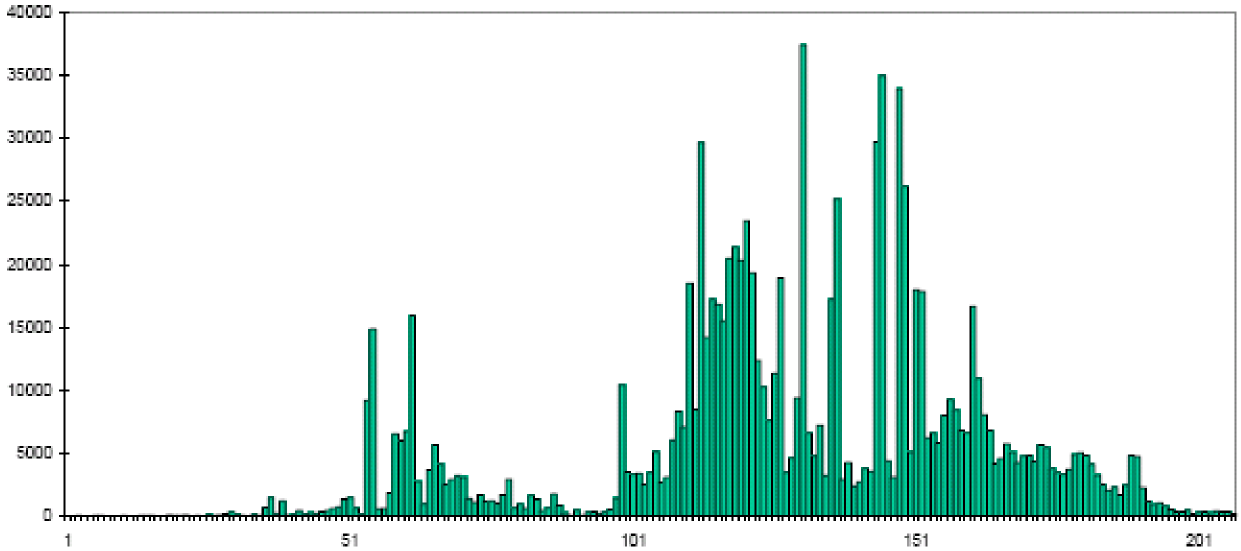


Primary data reduction:250-1000 different integrated regions

No *a priori* knowledge of the class of samples

Model for the prediction of independent data
Use class information to maximise separation among classes

Data pre-processing (NMR)

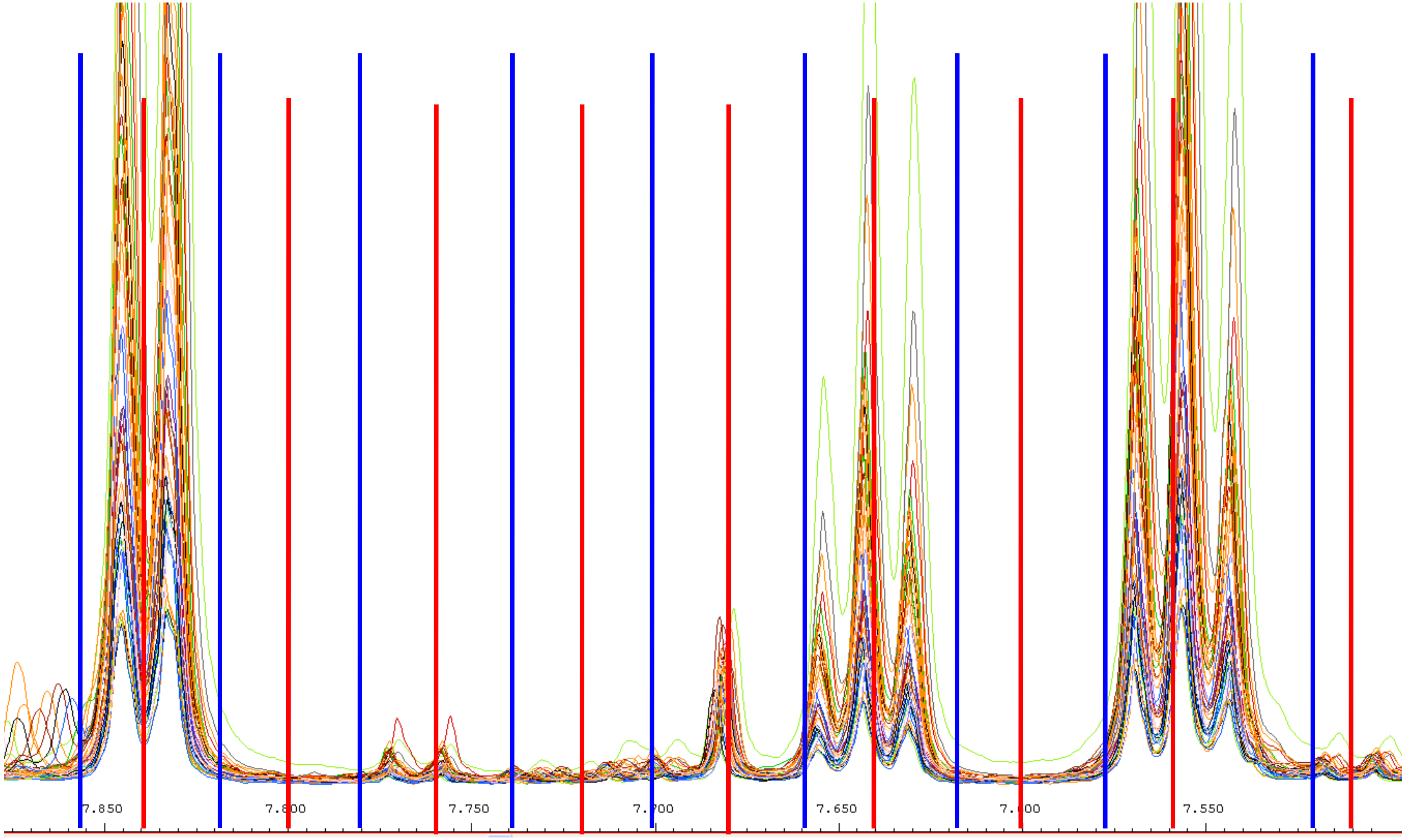


- Discretise x-axis into *n* equal sized bins, height = area under intensity (reduces impact of small variations in chemical shift e.g. due to pH)

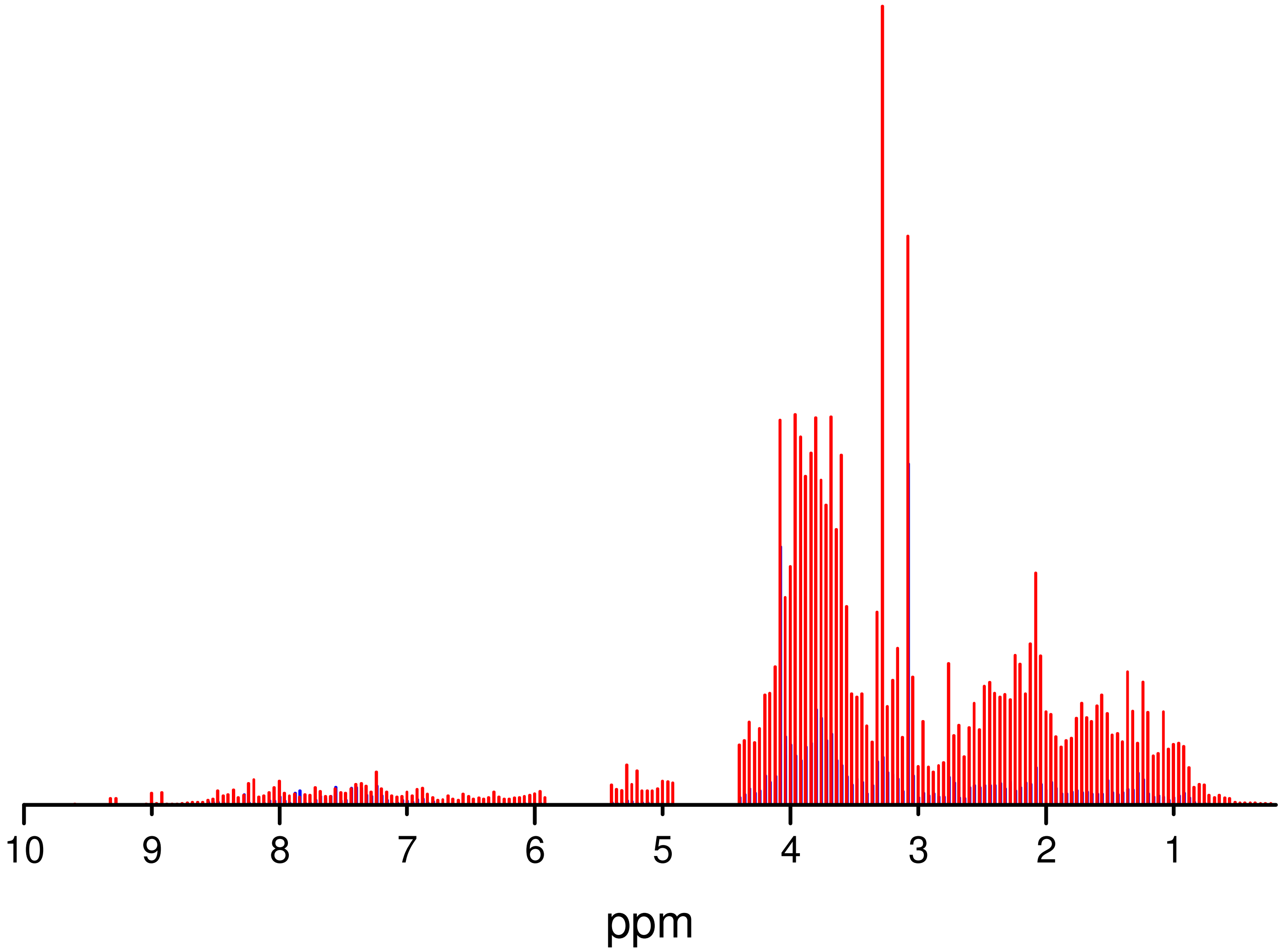
- Normalise bars for constant total area (removes effect of differences in concentration across samples)

- Remove insignificant regions (e.g. water and urea resonances in urine spectra)

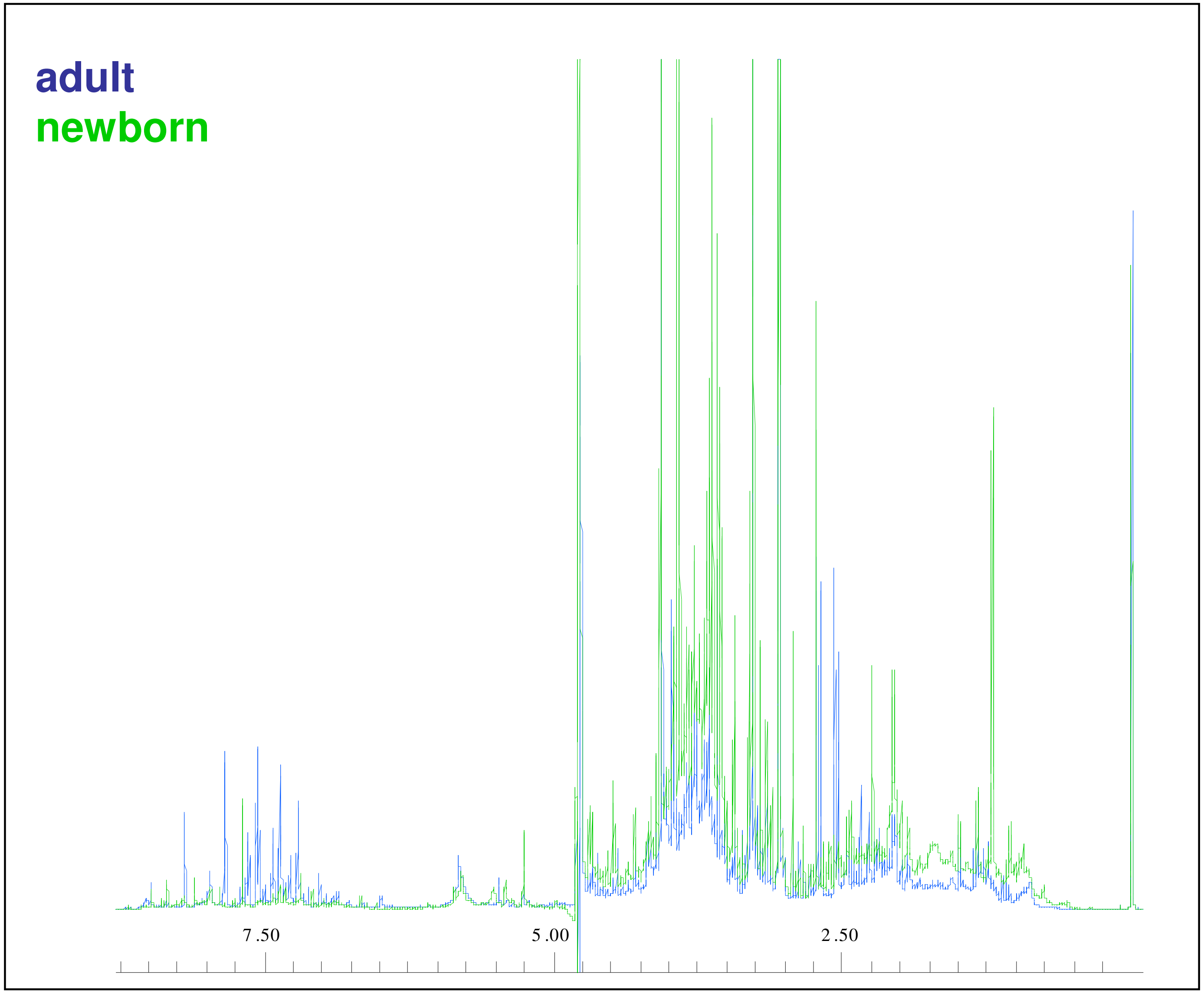
Fixed vs variable bucketing



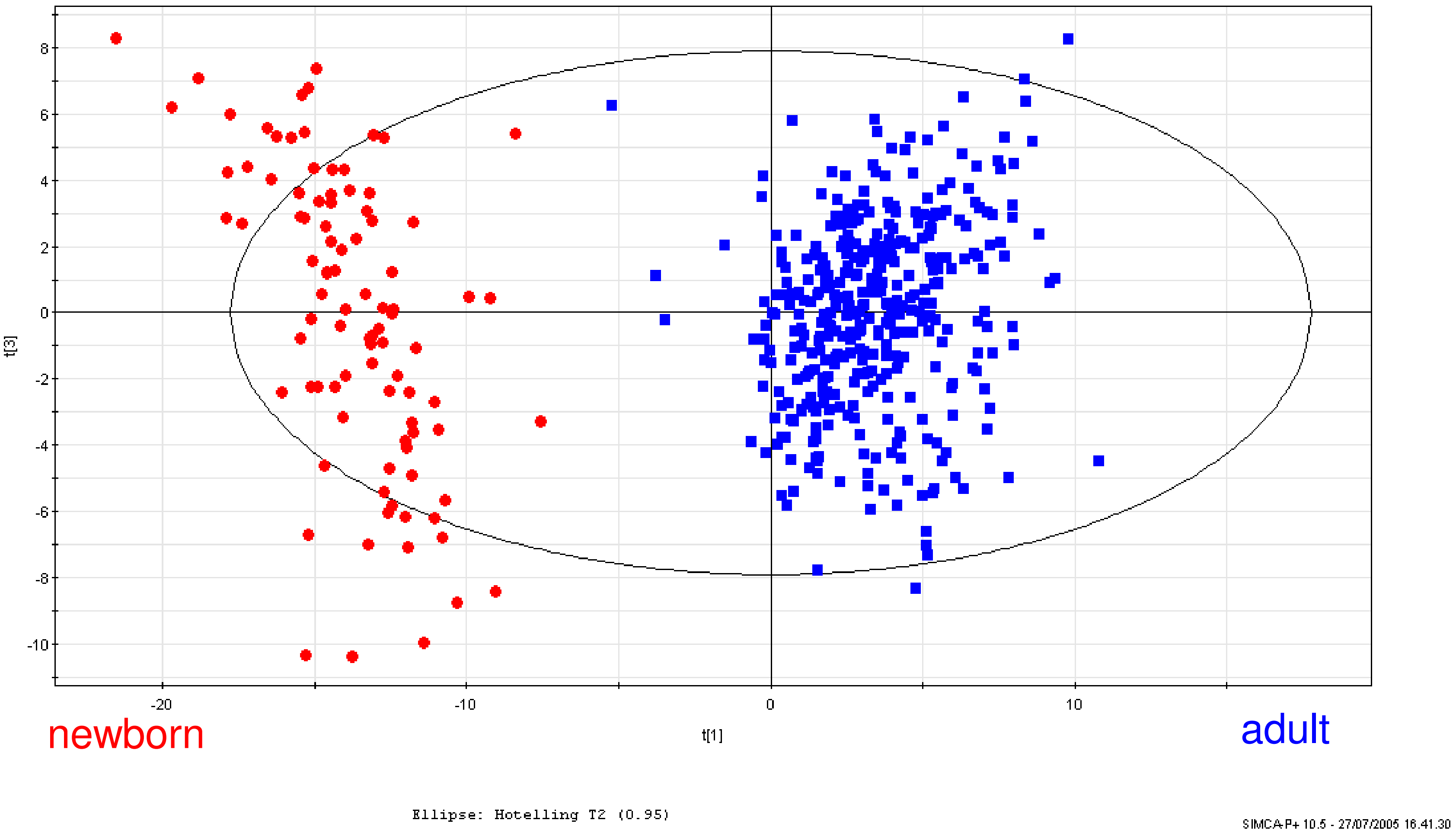
Normalization



Visualizing age-related differences



PCA newborns vs adults



Data pre-treatment (general for metabolomics)

Different data preprocessing steps are applied in order to generate 'clean' data in the form of normalized peak areas that reflect the (intracellular) metabolite concentrations. These clean data can be used as the input for data analysis. However, it is important to use an appropriate data pretreatment method before starting data analysis.

Besides induced biological variation, other factors are also present in metabolomics data:

1. Differences in orders of magnitude between measured metabolite concentrations; for example, the average concentration of a signal molecule is much lower than the average concentration of a highly abundant compound like ATP. However, from a biological point of view, metabolites present in high concentrations are not necessarily more important than those present at low concentrations.

2. Differences in the fold changes in metabolite concentration due to the induced variation; the concentrations of metabolites in the central metabolism are generally relatively constant, while the concentrations of metabolites that are present in pathways of the secondary metabolism usually show much larger differences in concentration depending on the environmental conditions.

3. Some metabolites show large fluctuations in concentration under identical experimental conditions. This is called uninduced biological variation.

Besides these biological factors, other effects present in the data set are:

4. Technical variation; this originates from, for instance, sampling, sample work-up and analytical errors.

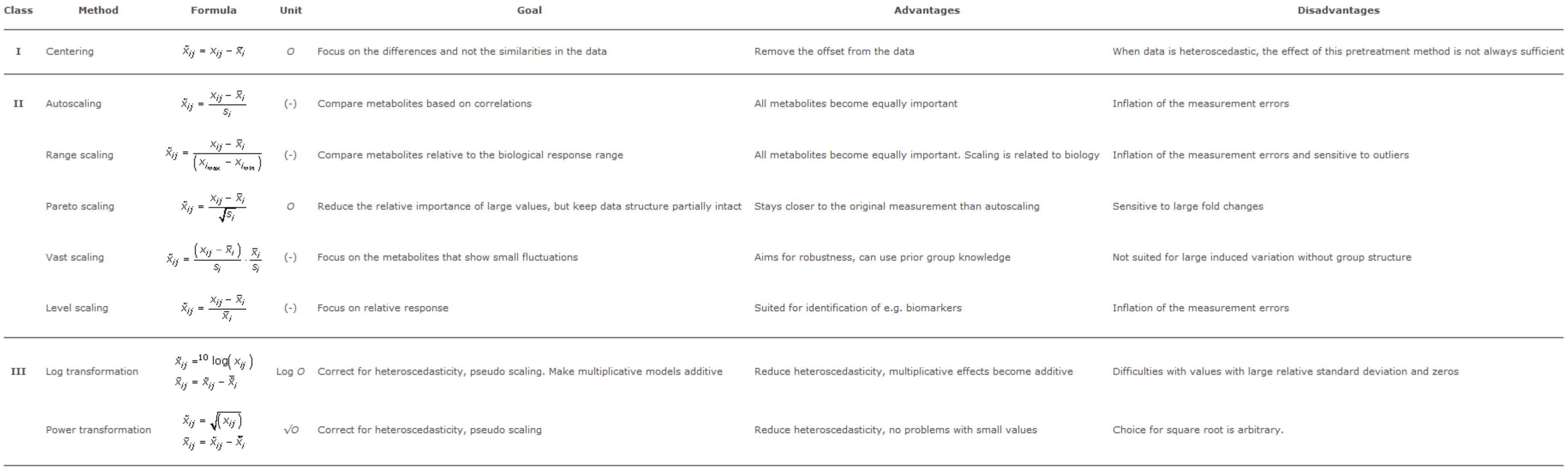
5. Heteroscedasticity; for data analysis, it is often assumed that the total uninduced variation resulting from biology, sampling, and analytical measurements is symmetric around zero with equal standard deviations. However, this assumption is generally not true. For instance, the standard deviation due to uninduced biological variation depends on the average value of the measurement. This is called heteroscedasticity, and it results in the introduction of additional structure in the data. Heteroscedasticity occurs in uninduced biological variation as well as in technical variation.

The variation in the data resulting from a metabolomics experiment is the sum of the induced variation and the total uninduced variation. The total uninduced variation is all the variation originating from uninduced biological variation, sampling, sample work-up, and analytical variation. Data pretreatment focuses on the biologically relevant information by emphasizing different aspects in the clean data.

… data pre-treatment (general for metabolomics)

Class I: Centering

Centering converts all the concentrations to fluctuations around zero instead of around the mean of the metabolite concentrations. Hereby, it adjusts for differences in the offset between high and low abundant metabolites. It is therefore used to focus on the fluctuating part of the data, and leaves only the relevant variation (being the variation between the samples) for analysis. Centering is applied in combination with all the methods described below.



Class	Method	Formula	Unit	Goal	Advantages	Disadvantages
I	Centering	$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$	*O*	Focus on the differences and not the similarities in the data	Remove the offset from the data	When data is heteroscedastic, the effect of this pretreatment method is not always sufficient
II	Autoscaling	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$	(-)	Compare metabolites based on correlations	All metabolites become equally important	Inflation of the measurement errors
	Range scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\left(x_{i_{max}} - x_{i_{min}}\right)}$	(-)	Compare metabolites relative to the biological response range	All metabolites become equally important. Scaling is related to biology	Inflation of the measurement errors and sensitive to outliers
	Pareto scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$	*O*	Reduce the relative importance of large values, but keep data structure partially intact	Stays closer to the original measurement than autoscaling	Sensitive to large fold changes
	Vast scaling	$\tilde{x}_{ij} = \frac{\left(x_{ij} - \bar{x}_i\right)}{s_i} \cdot \frac{\bar{x}_i}{s_i}$	(-)	Focus on the metabolites that show small fluctuations	Aims for robustness, can use prior group knowledge	Not suited for large induced variation without group structure
	Level scaling	$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{x}_i}$	(-)	Focus on relative response	Suited for identification of e.g. biomarkers	Inflation of the measurement errors
III	Log transformation	$\tilde{x}_{ij} = {}^{10}\log\left(x_{ij}\right)$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$	Log *O*	Correct for heteroscedasticity, pseudo scaling. Make multiplicative models additive	Reduce heteroscedasticity, multiplicative effects become additive	Difficulties with values with large relative standard deviation and zeros
	Power transformation	$\tilde{x}_{ij} = \sqrt{\left(x_{ij}\right)}$ $\hat{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$	√*O*	Correct for heteroscedasticity, pseudo scaling	Reduce heteroscedasticity, no problems with small values	Choice for square root is arbitrary.

van den Berg *et al. BMC Genomics* 2006 **7**:142 doi:10.1186/1471-2164-7-142

… data pre-treatment (general for metabolomics)

Class II: Scaling

Scaling methods are data pretreatment approaches that divide each variable by a factor, the scaling factor, which is different for each variable. They aim to adjust for the differences in fold differences between the different metabolites by converting the data into differences in concentration relative to the scaling factor.

There are two subclasses within scaling. The first class uses a measure of the data dispersion (such as, the standard deviation) as a scaling factor, while the second class uses a size measure (for instance, the mean).

Scaling based on data dispersion

Scaling methods tested that use a dispersion measure for scaling were autoscaling, pareto scaling, range scaling, and vast scaling (Table 1). Autoscaling, also called unit or unit variance scaling, is commonly applied and uses the standard deviation as the scaling factor. After autoscaling, all metabolites have a standard deviation of one and therefore the data is analyzed on the basis of correlations instead of covariances, as is the case with centering.

Pareto scaling is very similar to autoscaling. However, instead of the standard deviation, the square root of the standard deviation is used as the scaling factor. Now, large fold changes are decreased more than small fold changes, thus the large fold changes are less dominant compared to clean data.

Scaling based on average value

Level scaling falls in the second subclass of scaling methods, which use a size measure instead of a spread measure for the scaling. Level scaling converts the changes in metabolite concentrations into changes relative to the average concentration of the metabolite by using the mean concentration as the scaling factor. The resulting values are changes in percentages compared to the mean concentration. As a more robust alternative, the median could be used. Level scaling can be used when large relative changes are of specific biological interest, for example, when stress responses are studied or when aiming to identify relatively abundant biomarkers.

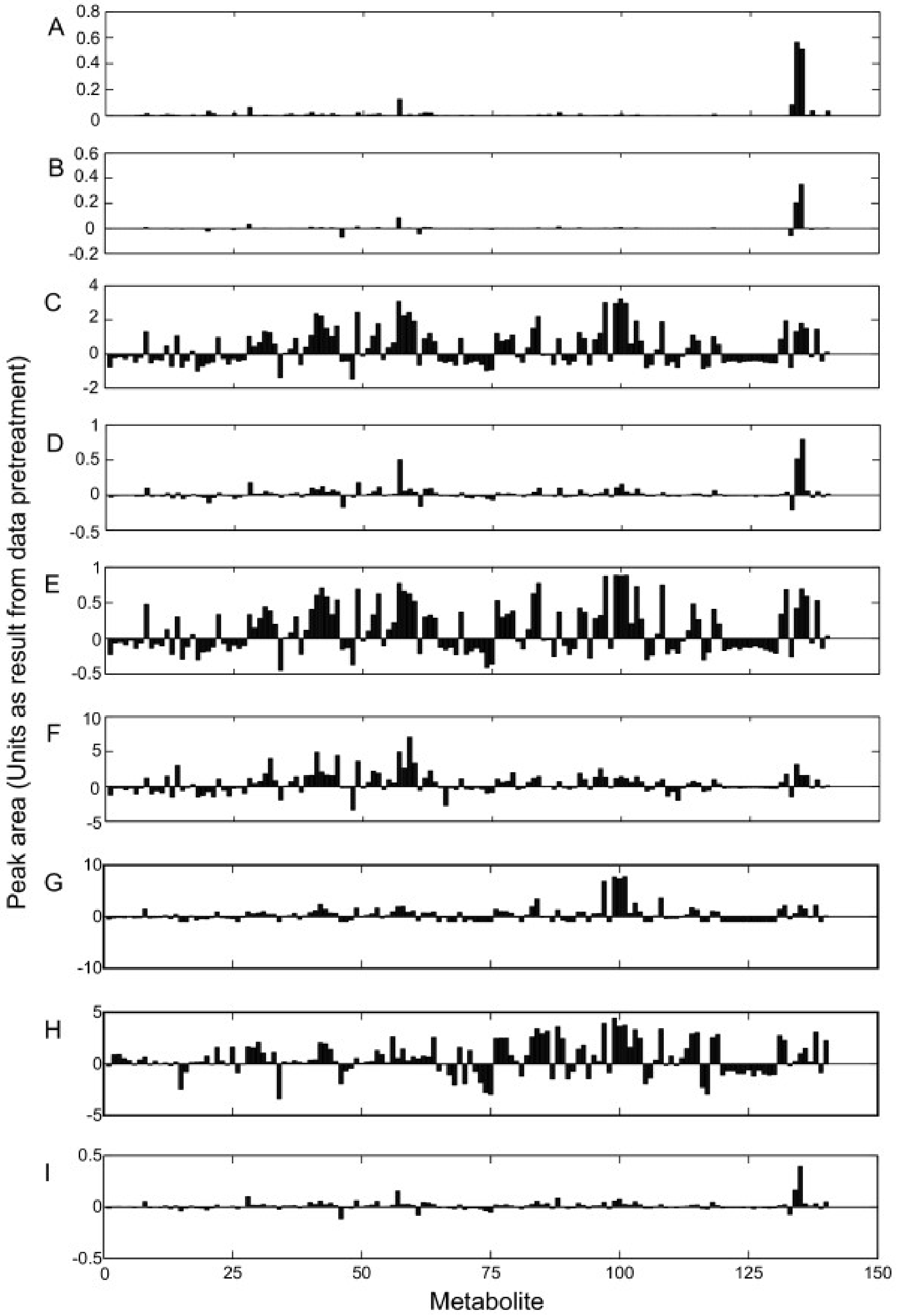
Class III: Transformations

Transformations are nonlinear conversions of the data like, for instance, the log transformation and the power transformation (Table 1). Transformations are generally applied to correct for heteroscedasticity, to convert multiplicative relations into additive relations, and to make skewed distributions (more) symmetric. In biology, relations between variables are not necessarily additive but can also be multiplicative. A transformation is then necessary to identify such a relation with linear techniques.

Since the log transformation and the power transformation reduce large values in the data set relatively more than the small values, the transformations have a pseudo scaling effect as differences between large and small values in the data are reduced. However, the pseudo scaling effect is not determined by the multiplication with a scaling factor as for a 'real' scaling effect, but by the effect that these transformations have on the original values. This pseudo scaling effect is therefore rarely sufficient to fully adjust for magnitude differences. Hence, it can be useful to apply a scaling method after the transformation. However, it is not clear how the transformation and a scaling method influence each other with regard to the complex metabolomics data.

A transformation that is often used is the log transformation (Table 1). A log transformation perfectly removes heteroscedasticity if the relative standard deviation is constant. However, this is rarely the case in real life situations. A drawback of the log transformation is that it is unable to deal with the value zero. Furthermore, its effect on values with a large relative analytical standard deviation is problematic, usually the metabolites with a relatively low concentration, as these deviations are emphasized. These problems occur because the log transformation approaches minus infinity when the value to be transformed approaches zero.

A transformation that does not show these problems and also has positive effects on heteroscedasticity is the power transformation (Table 1). The power transformation shows a similar transformation pattern as the log transformation. Hence, the power transformation can be used to obtain results similar as after the log transformation without the near zero artifacts, although the power transformation is not able to make multiplicative effects additive.



Effect of data pretreatment on the original data. Original data of experiment G2 (A), and the data after centering (B), autoscaling (C), pareto scaling (D), range scaling (E), vast scaling (F), level scaling (G), log transformation (H), and power transformation (I).

Signal/noise

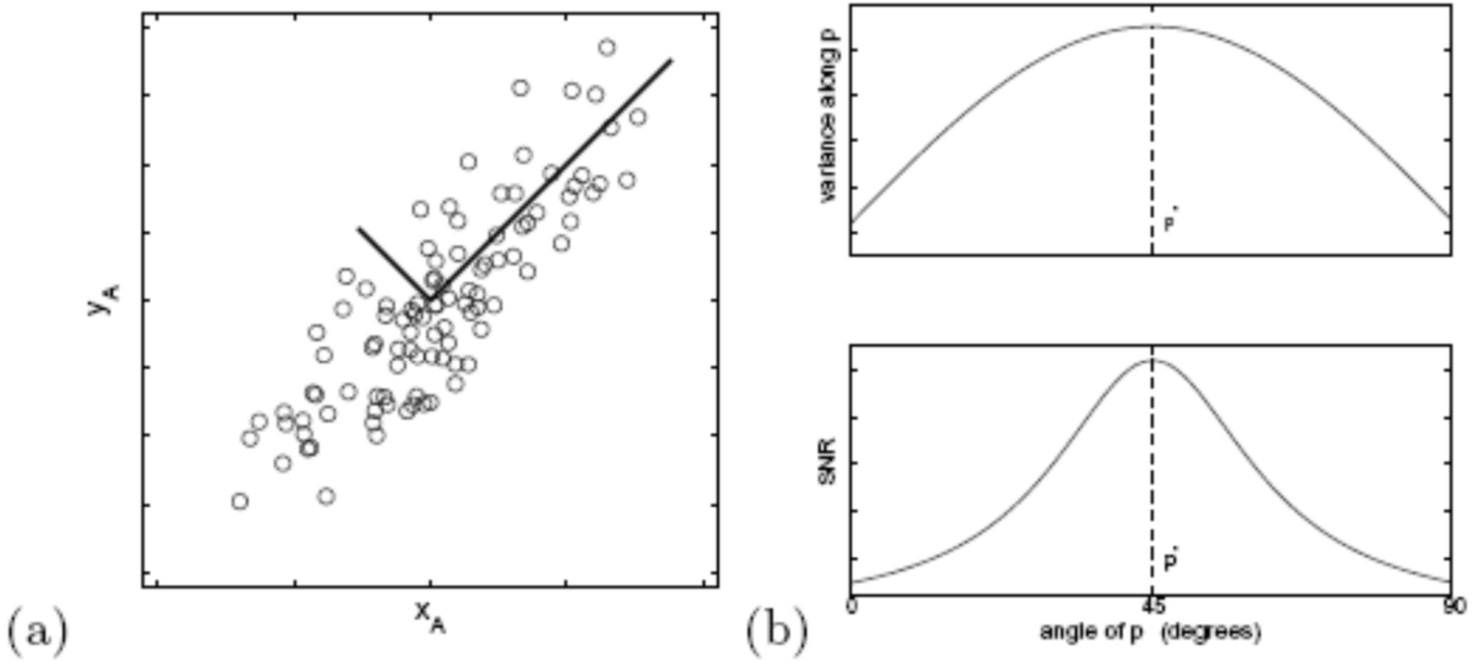


FIG. 2 (a) Simulated data of (x_A, y_A) for camera A. The signal and noise variances σ^2_{signal} and σ^2_{noise} are graphically represented by the two lines subtending the cloud of data. (b) Rotating these axes finds an optimal p^* where the variance and SNR are maximized. The SNR is defined as the ratio of the variance along p^* and the variance in the perpindicular direction.

Data redundancy

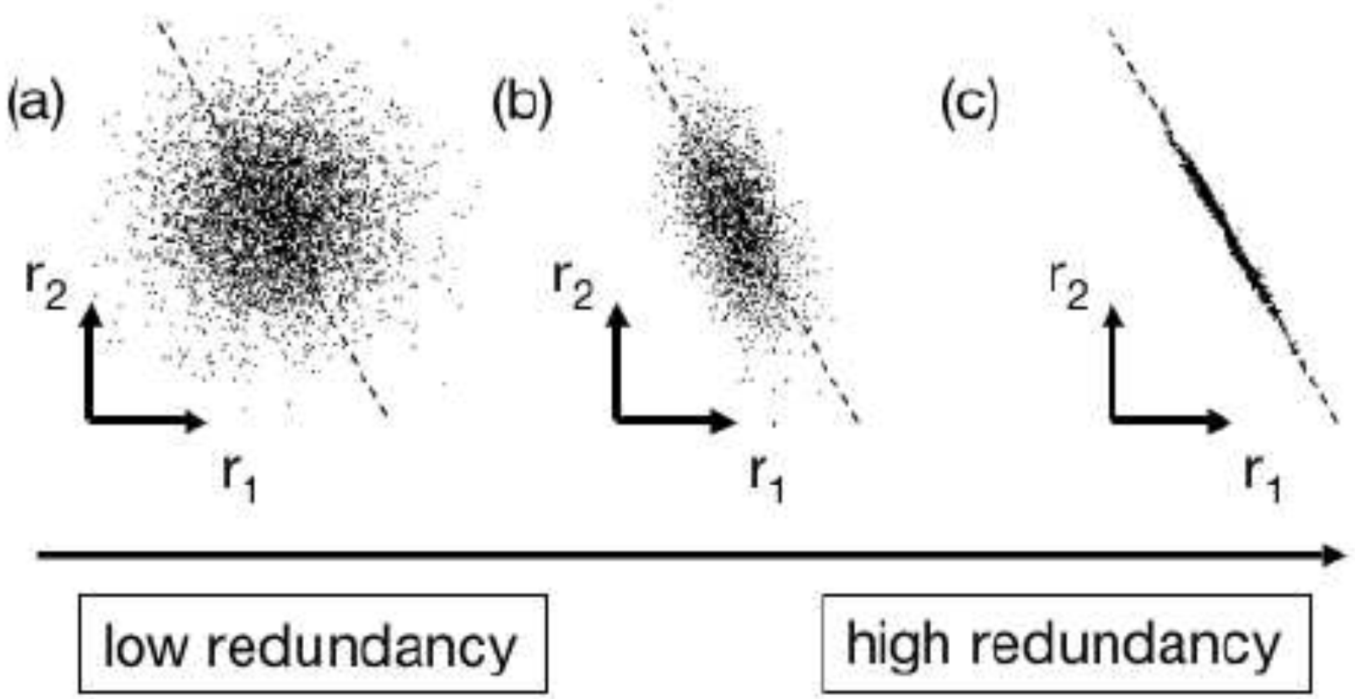


FIG. 3 A spectrum of possible redundancies in data from the two separate recordings r_1 and r_2 (e.g. x_A, y_B). The best-fit line $r_2 = kr_1$ is indicated by the dashed line.

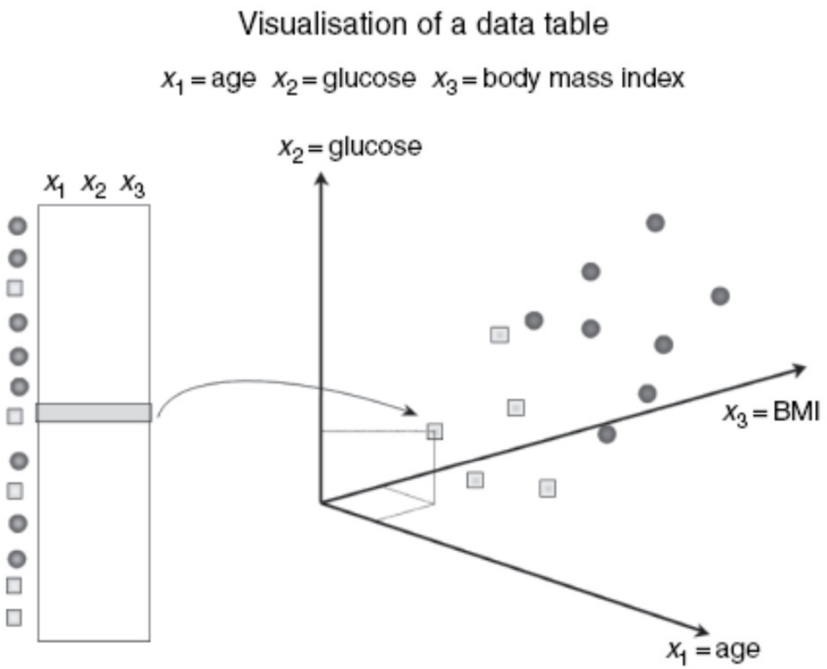

Visualisation of a data table
x_1 = age x_2 = glucose x_3 = body mass index
x_1 x_2 x_3
x_2 = glucose
x_3 = BMI
x_1 = age

Figure 6.1. Each row (e.g. object or observation) in a K-dimensional data table (here with $K = 3$ variables, designated x_1, x_2, x_3) can be represented as a point in a K-dimensional space (here one point in a three-dimensional space). The coordinates for each object in this multi-dimensional space are given by its three variables, that is a multivariate profile. A data table with N rows then corresponds to a swarm of points. Points that are close to each other have more similar properties than points that lie far apart.

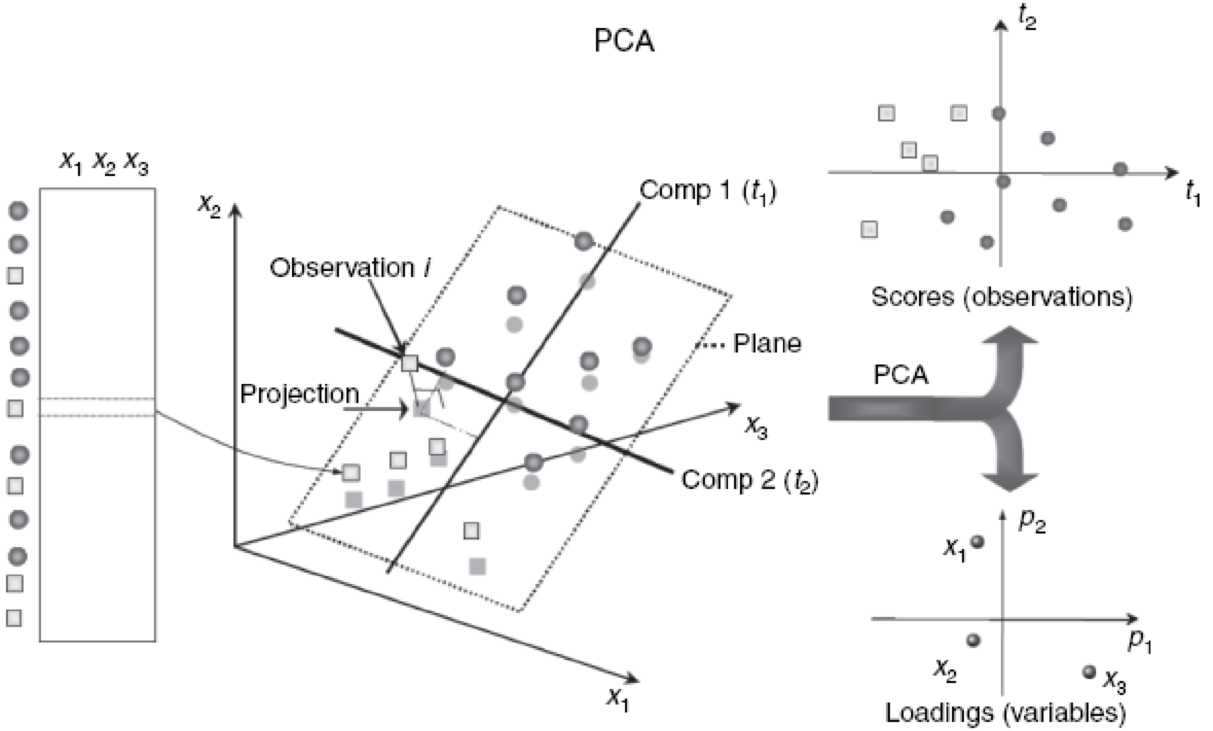


Figure 6.2. A principal component analysis (PCA) model approximates the variation in a data table by a low dimensional model plane. This model plane represents a two-dimensional projection of the multi-dimensional data and provides a score plot, where the relation among the observations or samples in the data table is visualized, for example if there are any groupings, trends or outliers. The loadings plot describes the influence of the variables and the relation among them. An important feature is that directions in the score plot correspond to directions in the loading plot, and vice versa.

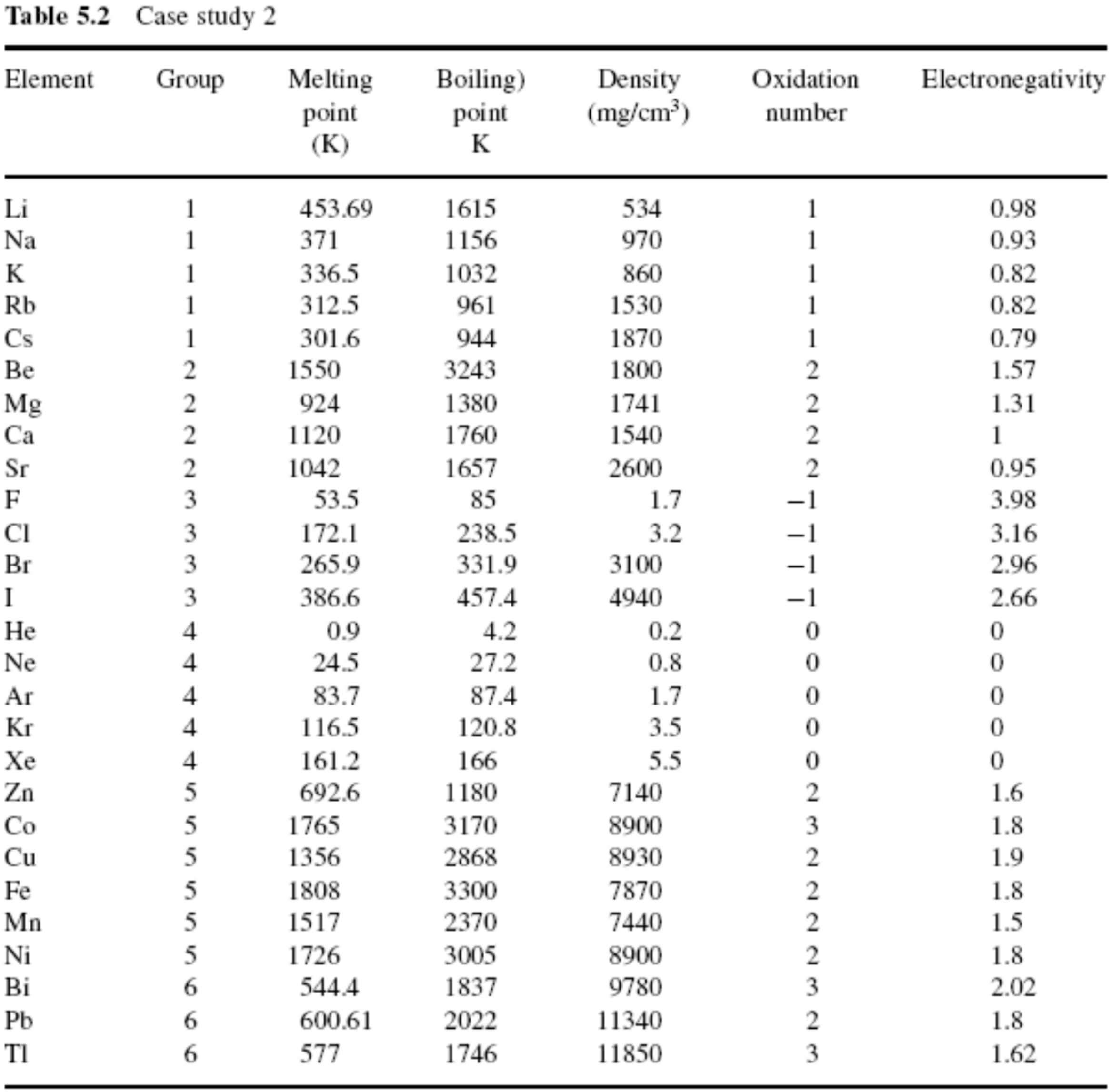


Table 5.2 Case study 2

Element	Group	Melting point (K)	Boiling) point K	Density (mg/cm^3)	Oxidation number	Electronegativity
Li	1	453.69	1615	534	1	0.98
Na	1	371	1156	970	1	0.93
K	1	336.5	1032	860	1	0.82
Rb	1	312.5	961	1530	1	0.82
Cs	1	301.6	944	1870	1	0.79
Be	2	1550	3243	1800	2	1.57
Mg	2	924	1380	1741	2	1.31
Ca	2	1120	1760	1540	2	1
Sr	2	1042	1657	2600	2	0.95
F	3	53.5	85	1.7	−1	3.98
Cl	3	172.1	238.5	3.2	−1	3.16
Br	3	265.9	331.9	3100	−1	2.96
I	3	386.6	457.4	4940	−1	2.66
He	4	0.9	4.2	0.2	0	0
Ne	4	24.5	27.2	0.8	0	0
Ar	4	83.7	87.4	1.7	0	0
Kr	4	116.5	120.8	3.5	0	0
Xe	4	161.2	166	5.5	0	0
Zn	5	692.6	1180	7140	2	1.6
Co	5	1765	3170	8900	3	1.8
Cu	5	1356	2868	8930	2	1.9
Fe	5	1808	3300	7870	2	1.8
Mn	5	1517	2370	7440	2	1.5
Ni	5	1726	3005	8900	2	1.8
Bi	6	544.4	1837	9780	3	2.02
Pb	6	600.61	2022	11340	2	1.8
Tl	6	577	1746	11850	3	1.62

PCA

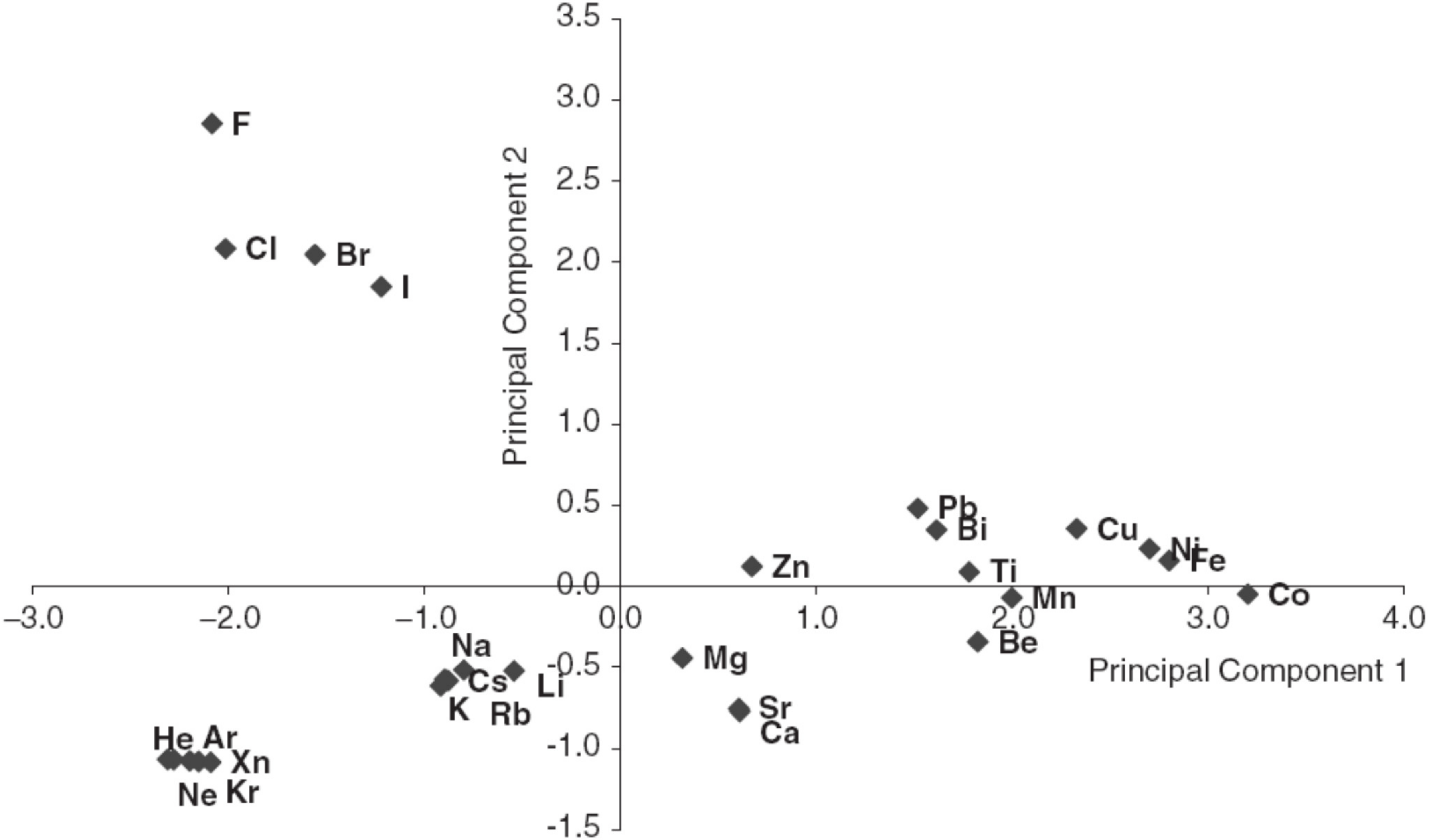


Figure 5.9 Scores plot of the first two PCs for case study 2

PCA

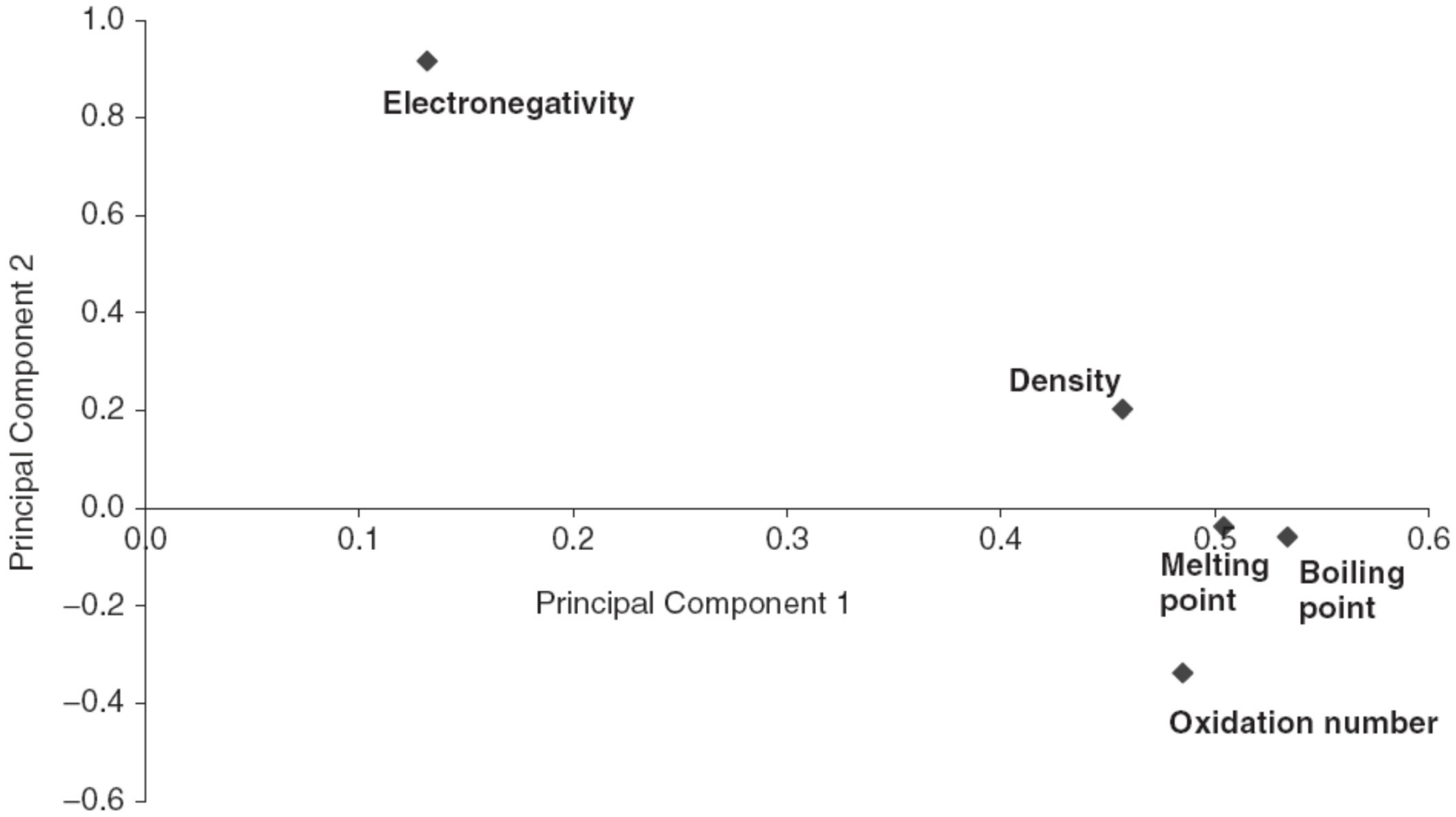


Figure 5.12 Loadings plot for case study 2

PCA

Cluster analysis

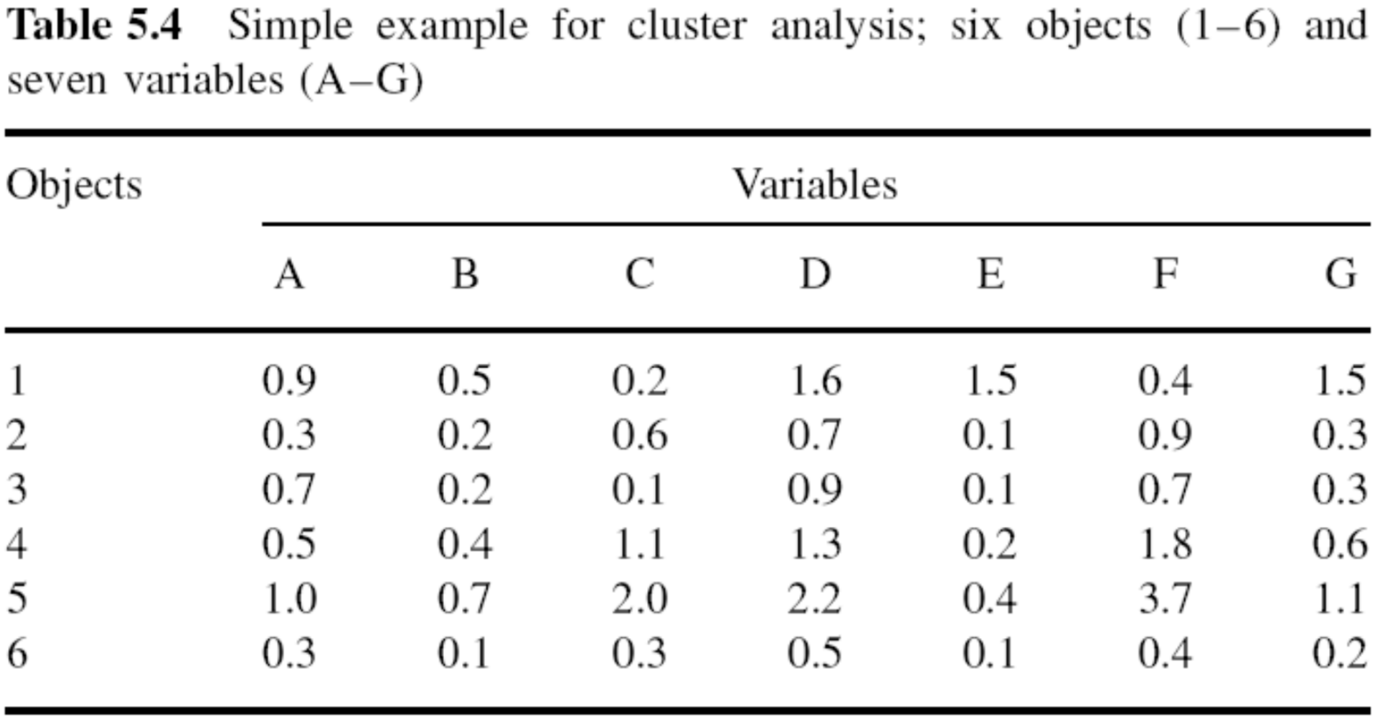


Table 5.4 Simple example for cluster analysis; six objects (1–6) and seven variables (A–G)

Objects	Variables						
	A	B	C	D	E	F	G
1	0.9	0.5	0.2	1.6	1.5	0.4	1.5
2	0.3	0.2	0.6	0.7	0.1	0.9	0.3
3	0.7	0.2	0.1	0.9	0.1	0.7	0.3
4	0.5	0.4	1.1	1.3	0.2	1.8	0.6
5	1.0	0.7	2.0	2.2	0.4	3.7	1.1
6	0.3	0.1	0.3	0.5	0.1	0.4	0.2

Table 5.5 Correlation matrix for the six objects in Table 5.4

	1	2	3	4	5	6
1	1					
2	–0.338	1				
3	0.206	0.587	1			
4	–0.340	$\boxed{0.996}$	0.564	1		
5	–0.387	0.979	0.542	0.990	1	
6	–0.003	0.867	0.829	0.832	0.779	1

CA

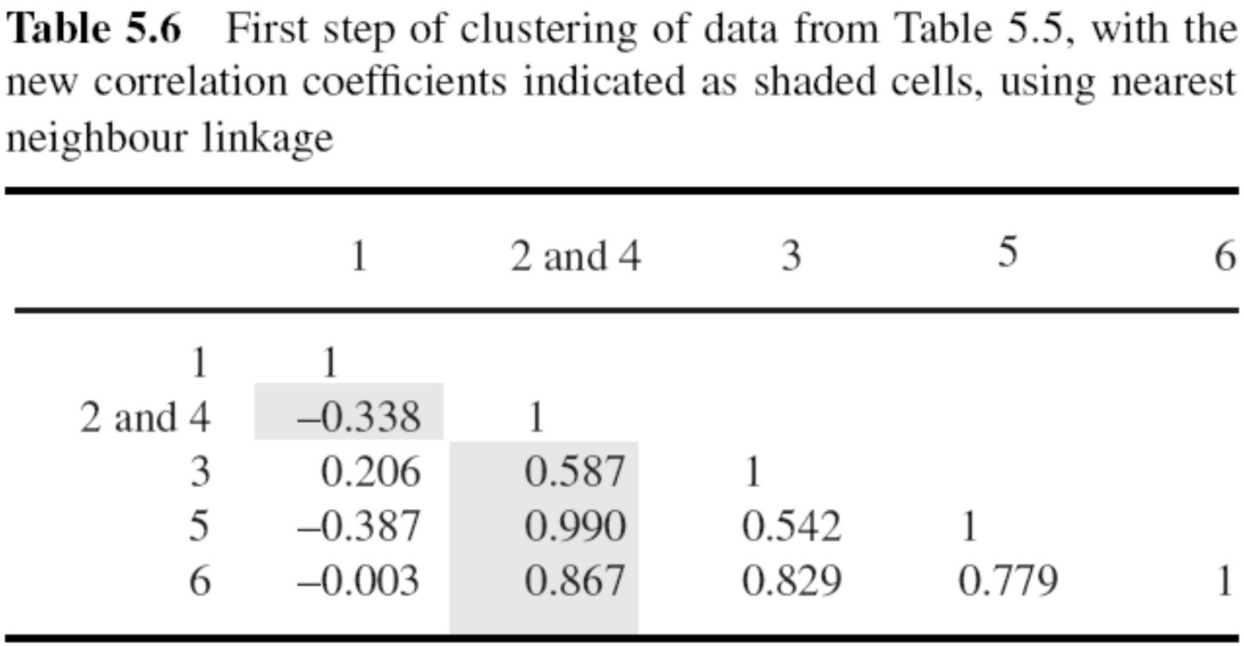


Table 5.6 First step of clustering of data from Table 5.5, with the new correlation coefficients indicated as shaded cells, using nearest neighbour linkage

	1	2 and 4	3	5	6
1	1				
2 and 4	–0.338	1			
3	0.206	0.587	1		
5	–0.387	0.990	0.542	1	
6	–0.003	0.867	0.829	0.779	1

CA

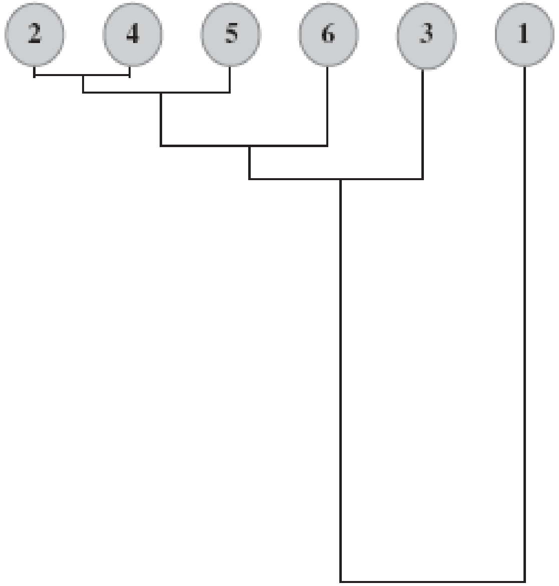


Figure 5.18 Dendrogram for data in Table 5.4, using correlation coefficients as similarity measures and nearest neighbour clustering

Discriminant analysis

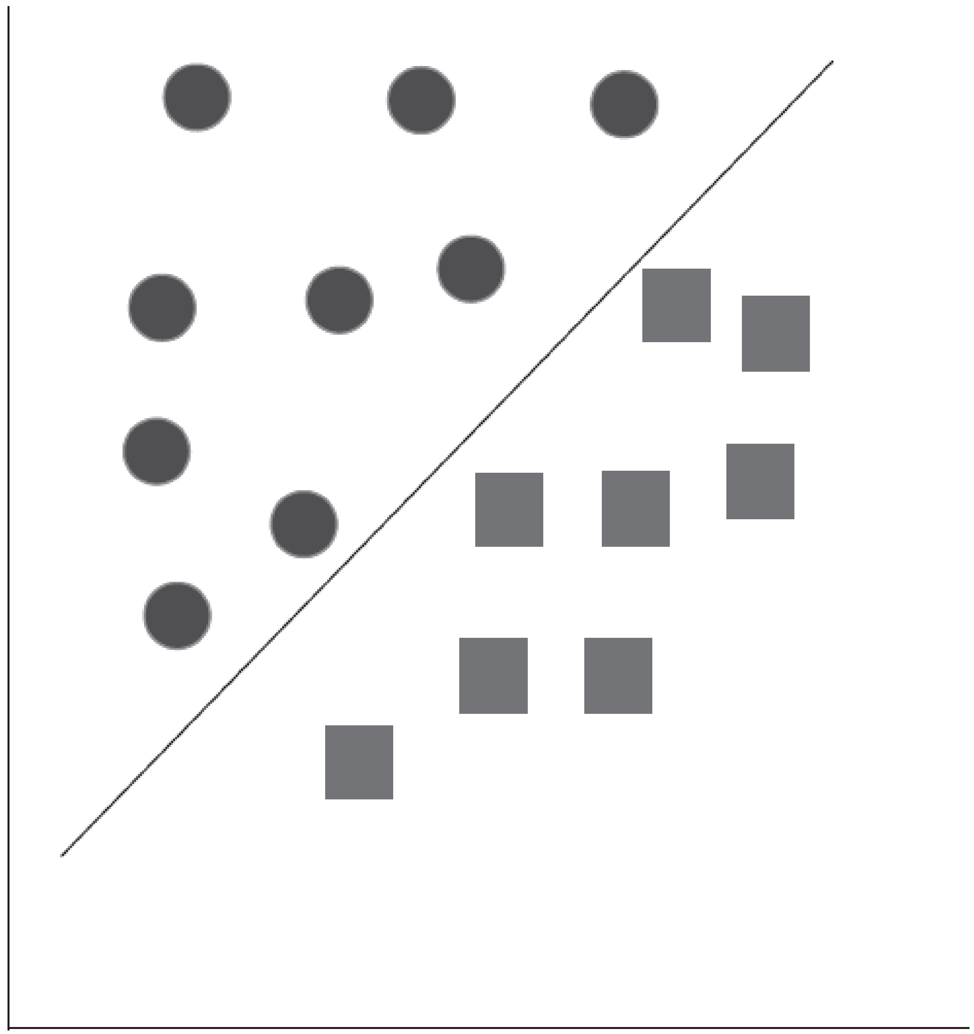


Figure 5.19 Bivariate classification where no measurement alone can distinguish groups

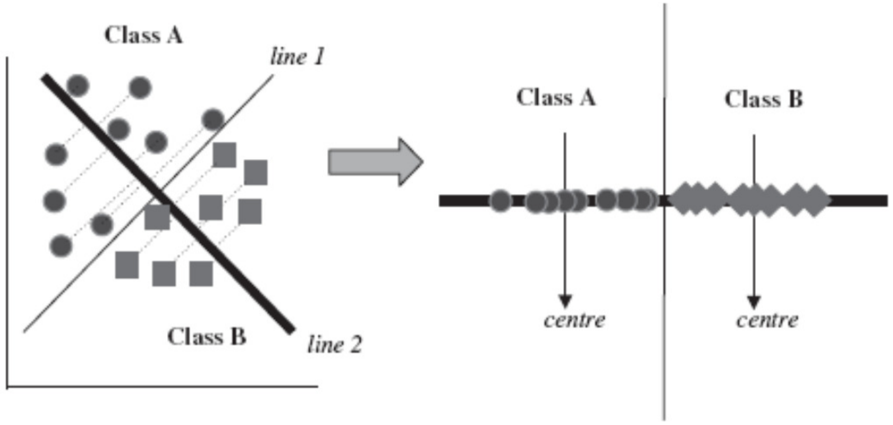


Figure 5.20 Projections

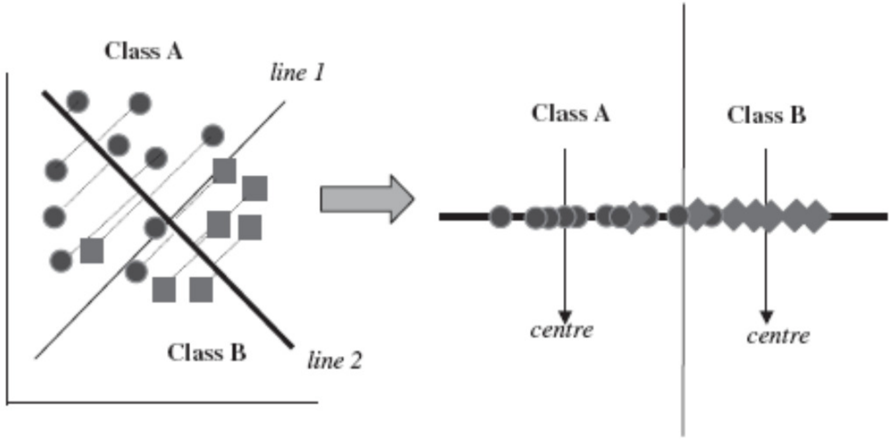


Figure 5.21 Projections where it is not possible to unambiguously classify objects

DA

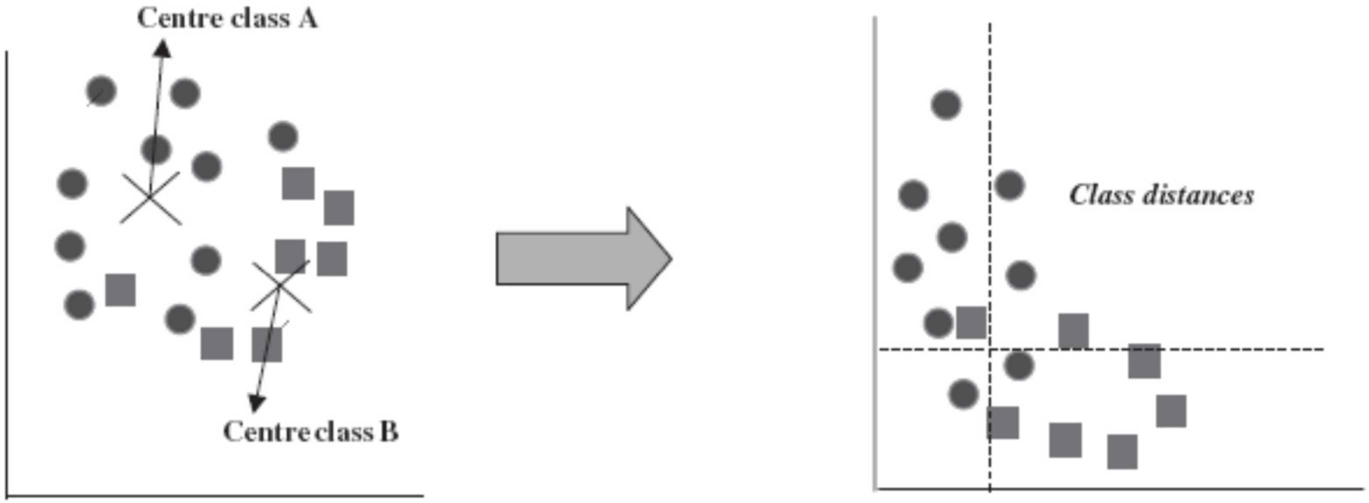


Figure 5.23 Class distance plot using two-dimensional information about centroids

DA