Journal of Pharmaceutical and Biomedical Analysis 48 (2008) 27–41

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Isotopic ratios to detect infringements of patents or proprietary processes of pharmaceuticals: Two case studies

E. Deconinck[a,b], A.M. van Nederkassel[a], I. Stanimirova[c], M. Daszykowski[c], F. Bensaid[d], M. Lees[d], G.J. Martin[d], J.R. Desmurs[e], J. Smeyers-Verbeke[a], Y. Vander Heyden[a,*]

[a] *Vrije Universiteit Brussel (VUB), Department of Analytical Chemistry and Pharmaceutical Technology, Laarbeeklaan 103, B-1090 Brussels, Belgium*
[b] *Laboratory for Pharmaceutical Technology and Biopharmacy, Katholieke Universiteit Leuven, O&N II Herestraat 49 Box 921, B-3000 Leuven, Belgium*
[c] *University of Silisia, Department of Chemometrics, Institute of Chemistry, 9 Szkolna Street, 40-006 Katowice, Poland*
[d] *Eurofins Scientific Analytics, Rue Pierre Adolphe Bobierre, BP 42301, 44323 Nantes Cedex 3, France*
[e] *CDP-Innovation SAS, Espace G2C, 75 Route de Gerland, 69307 Lyon Cedex 07, France*

ARTICLE INFO

Article history:
Received 3 March 2008
Received in revised form 17 April 2008
Accepted 18 April 2008
Available online 29 April 2008

Keywords:
Counterfeiting
Chemometrics
Ibuprofen
Acetyl salicylic acid
Isotopic ratios

ABSTRACT

Because of the increasing problem of drug counterfeiting and the potential danger related as well as the economic losses involved, the pharmaceutical industry and the regulatory instances are interested in the development of anti-counterfeiting and patent protection methodologies. In this paper, the evaluation of measured isotopic ratios by means of explorative chemometric techniques was performed to distinguish groups in two data sets containing samples of acetyl salicylic acid and ibuprofen, respectively. The samples in the data sets originated from different countries and manufacturers. For both compounds a clear distinction of groups of samples could be obtained. These groups could be explained based on the origin of the samples, both geographically as well as based on the manufacturer. Hypotheses were formulated concerning the synthetic pathways of the molecules and they were linked to the groups obtained with the chemometric tools.

© 2008 Published by Elsevier B.V.

1. Introduction

Counterfeit medicinal products are defined by the World Health Organisation (WHO) as products "deliberately and fraudulently mislabelled with respect to identity and/or source" [1]. This comprises products with wrong quantitative and/or qualitative composition, without active substances, with possible toxic substances or with correct composition in fake packaging [1].

Until recently, counterfeit drugs were only a problem of developing countries, due to a weak drug regulatory control and enforcement together with some other factors, such as scarcity, erratic supply of basic medicines, uncontrolled distribution chains, large price differences between genuine and counterfeit medicines, and lack of effective intellectual property right protection. However, counterfeiting products are increasingly becoming a serious global problem into the European and North American markets. This is possible through the bulk imports of raw materials or the bulk purchase of pills from abroad, but also through the rising popularity of so-called internet pharmacies, often selling counterfeit products originating from different developing countries [1,2].

Counterfeit products not only represent a financial drain for the pharmaceutical industry, but they are also a serious threat to public health, due to the lower quality of active substances, if at all present, excipients and the possible presence of toxic substances. The result of a counterfeit use could be that the medication is not effective or could lead to a long-term disease or injury [3–5]. Therefore, both the pharmaceutical industries and the regulatory instances are interested in the development of reliable methodologies of patent protection and anti-counterfeiting [6,7].

Most of the methodologies presented for the fight against counterfeiting of drugs are based on the determination and identification of trace impurities. In 1992, the Food and Drug Administration issued a report in which their approach in uncovering fraud in the generic drug industry was described [1]. The report focuses on the analysis of the excipients rather than the active ingredients, and the techniques used included Fourier-transform infrared spectrometry, thermogravimetric analysis, and liquid chromatography (LC) and gas chromatography (GC) [1]. In recent years, chromatographic techniques have become the methods of choice for obtaining drug impurity data. This is mostly due

* Corresponding author. Tel.: +32 2 477 47 34; fax: +32 2 477 47 35.
E-mail address: yvandh@vub.ac.be (Y. Vander Heyden).

0731-7085/$ – see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.jpba.2008.04.023

to considerable technological progress, contributing both to robustness and practicability. Also the use of hyphenated techniques such as GC-mass spectrometry, have contributed to the popularity of chromatographic techniques in this domain.

These chromatographic techniques are usually applied to identify specific compounds that can be linked to side reactions in the synthetic process of the active substance or to one of the excipients used [8]. A fingerprinting approach using pattern recognition on a part or the complete chromatogram of trace organic impurities of pharmaceuticals of different manufacturers has also been reported useful [2,9]. To identify counterfeit drugs also the use of near infrared spectrometry combined with multivariate modeling and classification is described in the literature [10].

However, these different approaches only lead to indirect evidence of fraud in the chemical and formulation processes, and are unable to screen the active substance itself. Stable isotope analysis, that probes into the atomic composition of the molecules themselves, has now appeared as a valuable and complementary technique for a number of applications in the pharmaceutical industry.

Stable isotopes are naturally occurring chemical tracers that are generally available in measurable concentrations. They are determined by isotope ratio mass spectrometry (IRMS) or by nuclear magnetic resonance (NMR) spectroscopy. These techniques are often used in the analysis of food and beverages. Repeatability and reproducibility of the data have been determined by several interlaboratory comparison studies, e.g. for ^{2}H and ^{18}O in fruit and vegetables by IRMS [11], for ^{2}H in fruit juices by NMR [12] and for ^{13}C in sugars and pulp from fruit juices by IRMS [13]. In a similar way to food products, where stable isotope content is used to differentiate between botanical origins, and between natural and synthetic sources, similar information can be obtained on pharmaceutical products. Compounds-specific isotopic information can provide an indication of the source of the raw material used in a reaction, and can help distinguish between two products that have been manufactured in different ways.

The goal of this study is to verify whether the isotopic ratios can be used as markers in order to identify counterfeiting and violations of the patented drugs production. This research was part of the EU-Counterpharm project. The idea is that stable isotope composition of the active molecule of a drug, determined either as an overall ratio by IRMS or as an isotopomeric profile obtained by NMR gives data, that is impossible to falsify for both compound and process recognition. Isotopic ratios were measured for two commonly used and commercially accessible drugs, i.e. ibuprofen and acetyl salicylic acid. Samples were purchased from different countries and manufacturers. Different chemometric tools were applied, i.e. principal component analysis (PCA) [14–16], projection pursuit (PP) [16] and multiple factor analysis (MFA) [17,18], in order to find patterns in the experimental data that could be linked to the origin and/or the manufacturing process of the drugs.

2. Theory

2.1. Synthetic pathways

For acetyl salicylic acid there is only one synthetic route, involving the acetylation of salicylic acid, obtained through carboxylation of phenol. Fig. 1 represents the synthetic pathway of acetyl salicylic acid.

Four different synthetic routes can be distinguished for ibuprofen. Fig. 2 shows the schematic representations for the different syntheses of ibuprofen.

2.2. PCA

PCA is a projection method [14,15]. It allows projecting high dimensional data into a low-dimensional space of new variables called principal components (PCs). Principal components are orthogonal and are constructed as linear combinations of the explanatory variables to maximise the description of the data variance. The projections of objects onto PCs are scores, and the projections of variables onto PCs are loadings. Therefore, scores inform about similarities among objects, while the loadings show the contributions of the different variables to a given PC and the correlation among the explanatory variables [14,15].

2.3. PP

PP is also a projection technique [15,16,19,20]. Similar to PCA, with PP the high-dimensional data are projected onto a low-dimensional space spanned by a few latent factors, called projection pursuit features (PPFs). Contrary to PCA, PPFs are obtained by maximising the projection index describing inhomogeneity of the data. Thus, with PP, a better insight into the data structure can be achieved and groups of similar samples that are not observed by PCA can be revealed. Eventually such groups of samples, if they exist, might be related to similar synthesis pathways. Therefore the kurtosis projection [21–23] index was chosen among the different projection indices available [20,22–24]. To obtain the PPFs, the Croux and Ruiz-Gazen algorithm [21] was used. In the first step of this algorithm, the data are sphered attributing to all variables the same importance in the further analysis (each variable has mean equal to zero and unit variance). After all objects are projected onto the normalized directions defined by the data origin

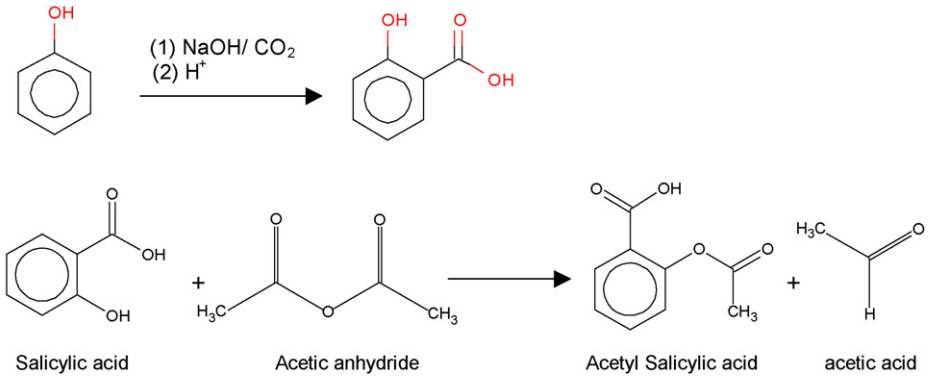

Fig. 1. Synthetic pathway for acetyl salicylic acid.

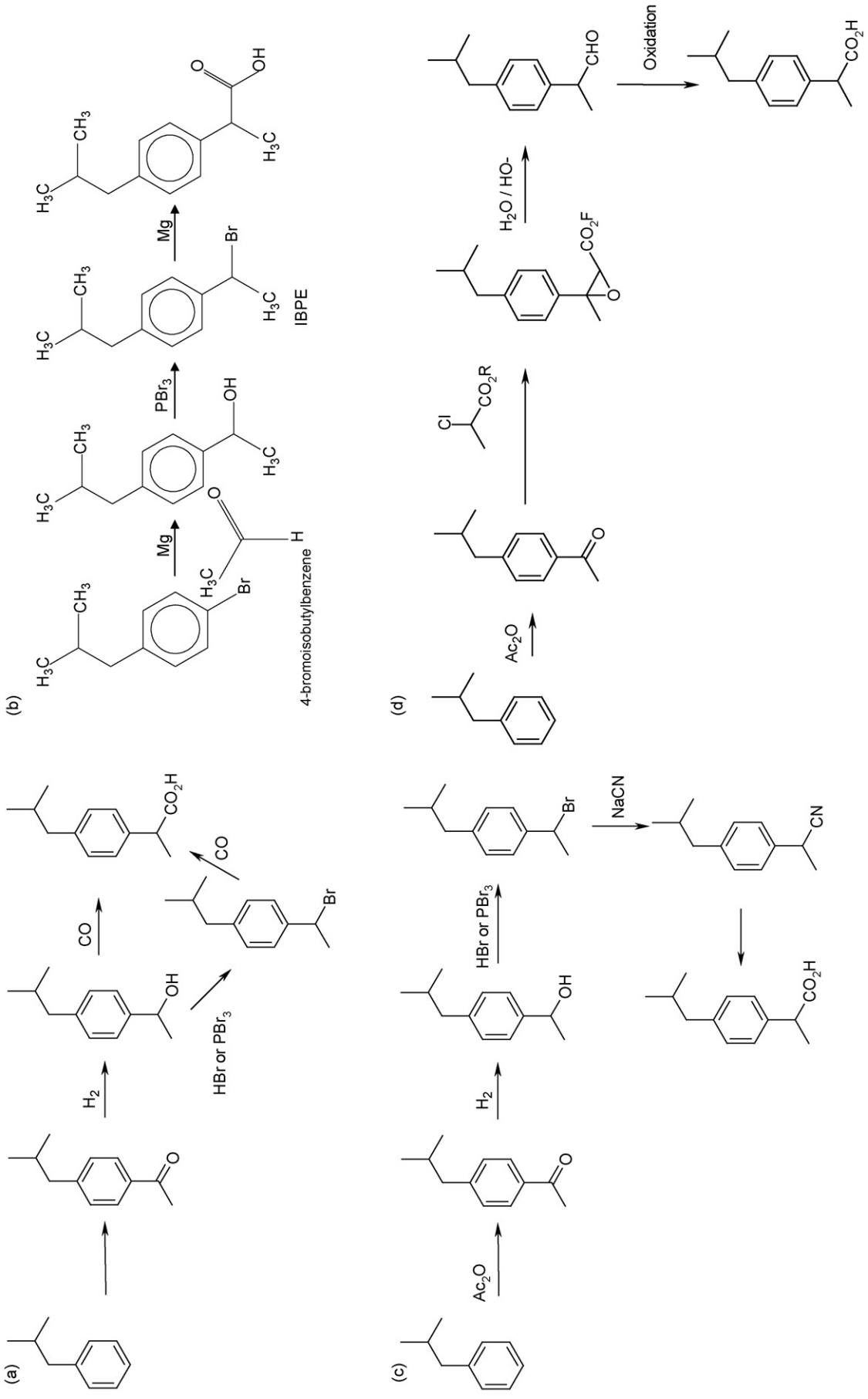

Fig. 2. Synthetic pathways for ibuprofen: (a) the carbonylation route; (b) the magnesium route; (c) the cyanure route; (d) Darzen's route.

Table 1
Identification of the different acetyl salicylic acid samples by country, manufacturer, brand name and batch number

Sample number	Country	Company name	Brand name	Batch number
1	Switzerland	Streuli	Asa	501803
2	Switzerland	Streuli	Asa	456803
3[a]	Switzerland	Roche	Aspro 500	L0B287
4[a]	Switzerland	Roche	Aspro 501	E28N3
5	Italy	Togal-Werk	ASS 500	15540
6	USA	Whitehall-Robins Healthcare	Anacin	3010213
7	USA	Walgreen's	Acetyl salicylic acid	P26110
8[b]	USA	Bayer	Acetyl salicylic acid	225261C
9	USA	SmithKline Beecham	Ecotrin	0J30A
10	USA	Walgreen's	Acetyl salicylic acid	1GE0435
11	USA	Walgreen's	Acetyl salicylic acid	1GE0162
12	Denmark	Nycomed	Kodimagnyl	10139051
13[b]	Germany	Bayer	Acetyl salicylic acid 500	EALZV00
14	Slovakia	Slovakofarma	Acylpyrin	190301
15	Slovakia	Slovakofarma	Acylpyrin	230,801
16[a]	Netherland	Roche	Aspro	L0D168
17	Netherland	Chefaro International	Chefarine	01H2201
18	UK	Boots	Caplets	4JJ
19[b]	Germany	Bayer	Acetyl salicylic acid 501	EATCL00
20[b]	Guatemala	Bayer	Acetyl salicylic acid	2022810
21[b]	Mexico	Bayer	Acetyl salicylic acid	1304MR
22[b]	Guatemala	Bayer	Acetyl salicylic acid	2050110
23	India	Reckitt Benckiser	Disprin	M13027
24	India	Reckitt Benckiser	Disprin	M13155
25	India	USV	Ecosprin	04001083
26	India	Nicholas Piramal	Micropyrin (contains cafein)	W2127
27	Portugal	SmithKline Beecham/Sterling Produtos Farmaceuticos	A-A-S	20440
28	Portugal	SmithKline Beecham/Sterling Produtos Farmaceuticos	A-A-S	20441
29[b]	Portugal	Bayer	Acetyl salicylic acid	23219
30[b]	Portugal	Bayer	Acetyl salicylic acid	BTA1HJ1
31	USA	Wal-Mart Stores	Equate, acetyl salicylic acid micro coated	3GE0627
32[b]	Japan	Bayer (Distributed by Meiji)	Acetyl salicylic acid	0006
33[b]	Spain	Bayer	Not available	R90101
34[a]	Thailand	Rhodia Thaï Industries	Rhodine	203802
35[a]	France	Rhodia	Rhodine	9910816
36	China	Nanjing	Not available	9913344
37[a]	USA	McNeil	Acetyl salicylic acid Saint Joseph	JCM018
38[a]	USA	McNeil	Acetyl salicylic acid Saint Joseph	HPM068
39[a]	USA	McNeil	Acetyl salicylic acid Saint Joseph	JCM067
40	USA	Wal-Mart Stores	Equate low strength acetyl salicylic acid	4CE0273
41	USA	Wal-Mart Stores	Equate low strength acetyl salicylic acid	4CE0108
42	USA	Wal-Mart Stores	Equate	5BE0599
43	USA	Wal-Mart Stores	Equate	4BE0109
44[b]	USA	Bayer	Original strength	282524A
45[b]	USA	Bayer	Original strength	280503P
46[b]	USA	Bayer	Original strength	278093L
47[b]	USA	Bayer	Original strength	280843P
48[b]	USA	Bayer	Extra strength	282704B
49[b]	USA	Bayer	Extra strength	280613P
50[b]	USA	Bayer	Extra strength	282684B
51	USA	Walgreen's	Acetyl salicylic acid	P30819
52	USA	Walgreen's	Acetyl salicylic acid	P31658
53	USA	Walgreen's	Acetyl salicylic acid	P32447
54	Belgium	Nycomed Christiaens	Acenterine	03A07
55[a]	France	Rhodia	Rhodine	0419912

[a] Nine samples (3, 4, 16, 34, 35, 37–39, 55) were manufactured by Rhodia (samples from Roche and McNeil), further referred to as manufacturer 2.
[b] Seventeen samples (8, 13, 19–22, 29, 30, 32, 33, 44–50) were manufactured by Bayer, further referred to as manufacturer 1.

and the individual data objects, the kurtosis index is evaluated for each projection. The projection with the maximum value of kurtosis is selected as the first PPF. The remaining PPFs are consecutively found in the residual space of the data. A detailed description of the algorithm can be found in Refs. [16,25].

2.4. Robust principal component analysis (RPCA)

In classical PCA the variance is maximised. Since the variance is very sensitive to the presence of outliers, the direction in which the principal components are drawn will be highly influenced by outliers. Therefore, the presence of outliers can make it very difficult to observe the general data structure. The applied criteria also can hamper the detection of objects with outlying behaviour.

In the first step of a PCA analysis the data are centred around the mean. Since the mean is not robust and thus influenced by the presence of outliers, in robust PCA the L_1-median estimator is used. This estimator is defined as the point that minimises the Euclidean distances to all data points [26,27]. In the following step the latent factors, called robust principal components (RPCs) have to be obtained. In order to make them less influenced by outliers a robust scale, called the Q_n estimator, is used instead of the variance. Q_n is defined as the first quartile of all pairwise differences between

Table 2
Identification of the different ibuprofen samples by country, manufacturer, brand name and batch number

Sample number	Country	Manufacturer	Brand name	Batch number
1	France	Laboratoires Elerte	Antarène	SX51
2	France	Wyeth	Advil	2DT041
3	France	Upsa	Upfen	lF6331
4	India	Cipla	Ibugesic	KR2005
5	India	Cipla	Ibugesic	E10088
6	India	Briocia Pharma	IPBrufen	NC018
7	India	Briocia Pharma	IPBrufen	NC032
8	India	Aventis	Combiflam	213341
9	India	Aventis	Combiflam	213407
10	Portugal	Abbott	Brufen	062808D
11	Portugal	Abbott	Brufen	062798D
12	Portugal	Medinfar	Arfen 400	2846
13	Portugal	Medinfar	Arfen 400	2848
14	Portugal	Ratiopharm	Ibuprofeno-ratiopharm	C22191
15	Portugal	Ratiopharm	Ibuprofeno-ratiopharm	D07480
16	USA	Equate	Wal-Mart Stores	3GE0329
17	China	Zuhai United Laboratories (Zhongshan)	Ibuprofen	030310
18	Japan	SS	Eve A	09S022
19	USA	Wyeth	Advil	A64082
20	USA	Wyeth	Advil	A64071
21	USA	Wyeth	Advil	A51288
22	USA	Wyeth	Advil	A54691
23	USA	Wal-Mart Stores	Equate	3ME0875
24	USA	Wal-Mart Stores	Equate	4BE0179
25	USA	Wal-Mart Stores	Equate	4CE0539
26	USA	Wal-Mart Stores	Equate	4BE0751
27	USA	McNeil	Motrin	JAA096
28	USA	McNeil	Motrin	HMA0196
29	USA	McNeil	Motrin	HSA220
30	USA	McNeil	Motrin	HSA097
31	USA	McNeil	Motrin	HSA195
32	USA	Walgreen's	Wal-Profen	P30939
33	USA	Walgreen's	Ibuprofen 200	P30349
34	USA	Walgreen's	Ibuprofen 200	P31016
35	USA	Walgreen's	Ibuprofen 200	P30509
36	Belgium	EOS Health Care	Dolofin	03A02
37	Belgium	Unicophar	Buprophar	02I21
38	Belgium	Unicophar	Buprophar I	03C05
39	Belgium	Eurogenerics/Nozik	Nofenal	299F09
40	Belgium	Abbott	Brufen	02G25.4
41	UK	Tesco	Ibuprofen	N4V0522
42	France	AJC Pharma	Ibuprofène Qualimed	1005
43	France	Upsa	Upfen	F6331
44	France	Elerté	Antarène	SX59
45	UK	Coop	Ibuprofen	BNW4V509
46	UK	Boots	Ibuprofen	N4K535
47	UK	Boots	Ibuprofen	10XX
48	UK	Boots	Ibuprofen	11XX
49	UK	Crookes Healthcare	Nurofen	10YY
50	UK	Coop	Ibuprofen	BNW4V519
51	Belgium	EOS Health Care	Dolofin	03A02
52	Belgium	Abbott	Brufen	03J14043
53	France	AJC Pharma	Ibuprofène Qualimed	1004
54	USA	Albermarle International Corporation	Ibuprofen 40 microns FBDR 70 KG	4070-2296/577112

two data points [28]. In order to obtain the directions of the RPCs, Q_n was maximised using the algorithm presented by Croux and Ruiz-Gazen [21,29]. As a result, an easier detection of objects with outlying behaviour is possible.

Outlier detection is based on the evaluation of the distance–distance plot, which represents for each object its robust distance versus its orthogonal distance to the robust PCA space [30]. Objects with high robust distances, have values for the majority of the variables that fit the general pattern, but have extreme values for some of the variables. Orthogonal outliers or objects with high orthogonal distances are objects that do not fit the robust PCA model. In order to decide whether an object has high robust or high orthogonal distances, cut-off values are defined as $\sqrt{\chi^2_{p,0.975}}$. This corresponds to the square root of the chi-square distribution for p significant RPCs and 97.5% confidence level.

2.5. MFA

In some data sets, groups or blocks of related variables can be distinguished. In PCA it is then possible that some variables have too high a weight on the principal component due to the fact that they belong to a block that is more represented. Therefore, in MFA the different blocks of variables are weighted [17,18]. More precisely, each variable in a block is weighted by the inverse of the first eigenvalue of a block. Because all variables of a block have the same weight, the first eigenvalue of each block becomes one and the influence of blocks is balanced on the first PC. Since MFA is in fact PCA in which the variables are weighted, information about the similarity of objects and the contributions of the different variables can also be obtained. Due to the specific weighing scheme of MFA additional information about the similarity of the different blocks of variables can be gathered. This is done by projecting the

so-called cross-product matrices ($\mathbf{X}_k\mathbf{Q}_k\mathbf{X}_k$) of the different blocks k, where $\mathbf{X}_k$ represents the data matrix of the variables in block k and $\mathbf{Q}_k$ contains the weights of the different variables in block k, on a certain PC. Each block of variables is then represented by one point in the common MFA space [17,18].

3. Experimental

3.1. Data collection

For acetyl salicylic acid, 55 samples from 18 different countries were collected. Table 1 shows all the samples with their country of origin, manufacturer, brand name and batch number. In this data set 17 samples were manufactured by Bayer (manufacturer 1). For another nine samples it was, based on the obtained results, assumed that Rhodia (manufacturer 2) was the manufacturer. These samples are indicated in Table 1.

Fifty-four samples of ibuprofen were collected in eight different countries. Table 2 shows the different samples defined by their county of origin, manufacturer, brand name and batch number.

3.2. Purification

Purification of all samples was performed according to the procedure described by Remaud et al. [31]. For acetyl salicylic acid a number of tablets, equivalent to 1 g of active compound, were first powdered and then dissolved in 20 ml of methanol. The obtained liquid was then filtered under vacuum through a fritted glass funnel to remove the major excipients. After elimination of the solvent, using a rotary evaporator, the sample was dissolved in a minimum amount of eluent and adsorbed on 1 g of silica gel. The sample was then eluted through a silica gel (70–200 μm) column (500 mm × 40 mm i.d., Batailler, Nantes, France). As eluent dichloromethane/ethanol (85/15, v/v) was used, each fraction of 100 μl eluents is controlled by thin-layer chromatography (TLC). For the TLC silica plates (Plaque CCM Polygram SIL G/UV, Roth, Germany) and a dichloromethane/ether (95/5, v/v) mobile phase were used. The fractions were spotted using a thin capillary (Roth, Germany) and detection was performed with an ultraviolet lamp at 254 nm.

All fractions, containing the compound of interest were collected in a 2-l round-bottomed flask and the solvent was removed using a rotary evaporator. A recrystallisation with both hot water and hot ethanol was applied. The crystals were dried overnight at 70 °C and stored under vacuum over P_2O_5. The purity of the commercial samples was then checked by ^{1}H NMR.

For ibuprofen a number of tablets, equivalent to 1 g of active compound, were first powdered and then dissolved in 20 ml of methanol. The obtained liquid was then filtered under vacuum through filter paper (number 2) to get rid of the major excipients. The solvent was reduced to a maximum of 5 ml using a rotary evaporator. The sample was loaded onto the head of the column and eluted through a silica gel (35–70 μm) column (500 mm × 30 mm i.d.). The eluent used was pentane/ether (75/25, v/v). 50 ml fractions were collected and controlled by TLC. All fractions containing the product were collected in a 2-l round-bottomed flask and the solvent was removed using a rotary evaporator. The compound was dried overnight at 40 °C and stored under vacuum over P_2O_5. The purity of the commercial samples was checked by ^{1}H NMR.

3.3. Isotopic measurements

For the acetyl salicylic acid samples ^{13}C, ^{18}O and ^{2}H measurements were performed by IRMS on bulk material and/or the purified product.

An IRMS measurement is generally expressed as follows:

$$\delta(‰) = 1000\left[\left(\frac{R_{\text{product}}}{R_{\text{Standard}}}\right) - 1\right]$$

where R_{product} is the isotopic ratio (^{13}C/^{12}C, ^{18}C/^{16}C) of the sample and R_{Standard} is the known isotopic ratio of the international standard (VPDB, VSMOW). These measurements led to the following relative isotopic ratios: $^{13}C_{raw}$ stands for the ^{13}C IRMS measurement on bulk material, expressed in ‰; $^{13}C_p$ stands for the ^{13}C IRMS measurement on purified material, expressed in ‰; $^{18}O_{raw}$ stands for the ^{18}O IRMS measurement on bulk material, expressed in ‰; and D/H stands for the ^{2}H IRMS measurement on bulk material expressed in ppm.

The mass spectrometric determinations of the carbon isotope (^{13}C) ratios were carried out by on-line analysis using a Carlo-Erba (Milano, Italy) NA 1500 II elemental analyser fitted to a Finnigan MAT DELTA E mass spectrometer (Bremen, Germany). Determinations of the oxygen isotope (^{18}O) ratios were carried out by on-line analysis using a Carlo-Erba NA 1500 II elemental analyser (NA 1500 series, Fisons Instruments SpA, Milano, Italy) fitted to a micromass spectrometer (Withenshawe, Manchester, UK). Hydrogen isotope ratios were determined using a Euro PYR-OH elemental analyser (Eurovector, 1300 °C, SpA, Milano, Italy) coupled to an Isoprime isotope ratio mass spectrometer (Manchester, UK).

Five site-specific isotopic ratios, $(D/H)_i$, and molar fractions f_i were also measured using the SNIF-NMR(R) method, a technique that uses nuclear magnetic resonance to quantitatively determine deuterium content in specific sites of a molecule. These specific ratios and fractions are associated to the deuterium contents of the different sites of the acetyl salicylic acid molecule. They are defined as:

$$\left(\frac{\text{D}}{\text{H}}\right)_i = \frac{S_i P^{\text{ref}} N_{\text{H}}^{\text{ref}}}{S^{\text{ref}} P_i N_{\text{H}}} \left(\frac{\text{D}}{\text{H}}\right)^{\text{ref}} \quad (1)$$

$$f_i = \frac{S_i}{\sum_{i=1}^{5} S_i} \quad (2)$$

with S_i and S^{ref} associated with the peak area of site i of the acetyl salicylic acid molecule and the peak area of the only site of the internal reference, respectively. P_i and P^{ref} the stoichiometric numbers of the hydrogen atoms on site i and on the internal reference. N_{H} and $N_{\text{H}}^{\text{ref}}$ are associated with the number of moles of acetyl salicylic acid and internal reference, respectively. $(D/H)^{\text{ref}}$ represents the isotopic deuterium ratio of the internal reference chosen. The experiments were carried out on a DPX400 Brüker NMR spectrometer (Wissembourg, France) at a temperature of 308 K. The spectra were recorded at 61.4 MHz using a specific 10 mm (o.d.) probe equipped with a 19F locking device. An internal referencing procedure was used. The mean values of three spectra gives the final $(D/H)_i$ ratios. The reference used is *tert*-butyl disulfide (TBDS). The $(D/H)_i$ ratios are expressed in ppm. The molar fractions of the monodeuterated isotopomers of acetyl salicylic acid, f_i, are directly calculated from the signal areas.

For the ibuprofen samples $^{13}C_{raw}$, $^{13}C_p$, D/H and seven molar fractions, f_i, were measured in analogy with the procedures described for acetyl salicylic acid.

3.4. Chemometric exploration

All algorithms used in the chemometric exploration were programmed in house for Matlab 6.5 (The Mathworks, Matick, MA). Programming of the projection pursuit and the robust PCA algorithms was done according to the algorithms described by Croux and Ruiz-Gazen [21,29].

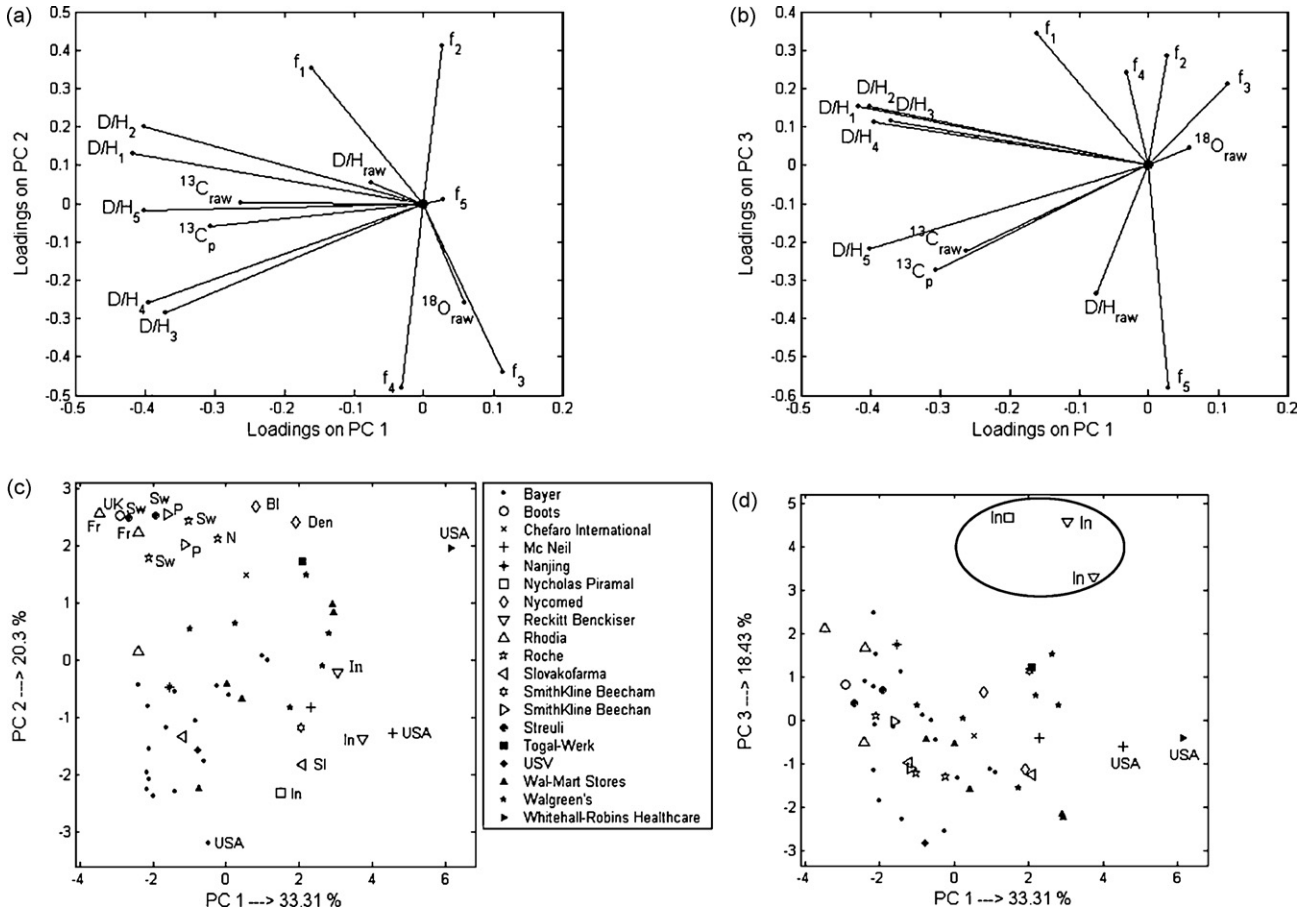

Fig. 3. Principal component analysis on the acetyl salicylic acid data set: (a) PC1–PC2 loading plot; (b) PC1–PC3 loading plot; (c) PC1–PC2 score plot; (d) PC1–PC3 score plot.

4. Results and discussion

4.1. Data pre-processing

For the acetyl salicylic acid data set, three variables, $^{13}C_{raw}$ product, D/H raw product and D/H_5, were not normally distributed as evaluated from Kolmogorov–Smirnov and Shapiro–Wilk tests [32]. A log transformation did not result in a normal distribution for these variables and maintained the normal distribution of the other variables.

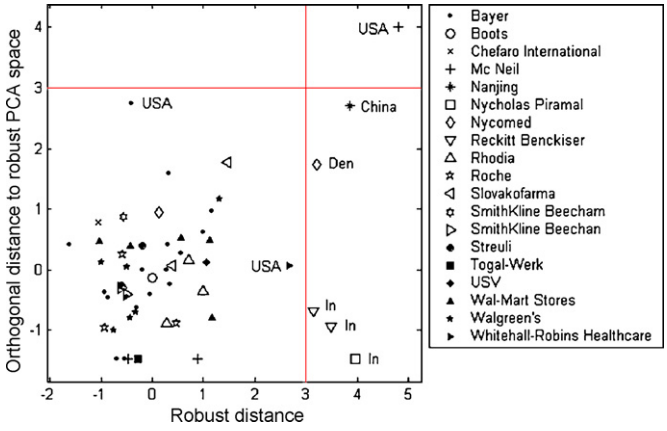

Fig. 4. Robust principal component analysis on the acetyl salicylic acid data set: distance–distance plot for 6 RPCs.

For the ibuprofen data set, $^{13}C_p$, f_4, f_6 and f_7 were not normally distributed. Again a log transformation did not result in a normal distribution for these variables and maintained the normal distribution of the others. Therefore, it was decided not to use the log transformation to process the data in the further analysis.

Except for robust PCA, all calculations were performed on the autoscaled data. For variables exhibiting negative values the autoscaling was calculated using their absolute values. This autoscaling avoids the effects of scale that could lead to some variables having higher importances than others. For robust PCA the data are standardized in a robust way [16]. Instead of using mean and standard deviation as is the case in the autoscaling procedure, median and robust scale estimates [28] were applied. In a first step the median of the column is subtracted from each data element in the column and then divided by the corresponding robust scale value.

4.2. The acetyl salicylic acid data set

4.2.1. Chemometric exploration

First a PCA analysis was performed. Since sample 39 (McNeil, USA), had an extreme value along PC2 and PCA is very sensitive to the presence of extreme objects, it was decided to reanalyse the data without this sample. Three principal components then explain about 72% of the variance in the data. The compression is not very efficient, due to the low correlation of several variables. On the loading plots (Fig. 3a and b), PC1 reflects the samples according to the D/H_i ratios and to a lesser extent to $^{13}C_{raw}$ and $^{13}C_p$. PC2 contrasts between the molar fractions f_1 and f_2 on the one hand, and

f_3 and f_4 on the other. On PC3 the variables with the highest loadings are the molar fractions f_1 and f_5, and D/H_{raw}. Some variables always are close together in the loading plots and thus show correlation. When the pairwise absolute correlation coefficients are investigated it follows that $^{13}C_{raw}$, $^{18}O_{raw}$ and D/H_{raw} mutually do not show high correlation and therefore provide different information. $^{13}C_{raw}$ and $^{13}C_p$ are highly correlated ($R = 0.86$), and probably one of both variables can be deleted from the data set. The group of the site-specific ratios (D/H_i) is rather highly correlated. Almost each pair of D/H_i has a correlation coefficient above 0.6. The highest absolute correlations are obtained for the pairs D/H_1 and D/H_2 ($R = 0.88$), D/H_3 and D/H_4 ($R = 0.83$) and D/H_4 and D/H_5 ($R = 0.73$). In the group of the molar fractions, only f_1 and f_5 ($R = 0.61$) show a relatively high correlation.

The score plots (Fig. 3c and d) show that some groups of samples can be distinguished. Fig. 3c allows the discrimination of two groups. The first group situated at the top of the plot (high PC2 values), contains two samples purchased from France (Rhodia), four from Switzerland (Roche and Streuli), two from Portugal (SmithKline Beecham) and one from the UK (Boots), The Netherlands (Roche), Belgium (Nycomed) and Denmark (Nycomed). All other samples belong to the second group, except for two extreme samples, along PC1, originating from the USA (McNeil and Whitehall-Robins Healthcare). In Fig. 3d, three Indian samples (Nicholas Piramal and Reckitt Benckiser) can be considered as extreme along PC3.

When this discrimination is explained in terms of variables, it can be stated that the first group is characterised by high f_1 and f_2 values, and low f_3 and f_4. For group 2 the opposite is true. The two extreme American samples along PC1 are characterised by low D/H_i, f_3 and f_4 values and high f_1 and f_2 values. The discrimination of the three Indian samples along PC3 is due to the fact that they have lower f_5 and higher f_1 values compared to the other samples. Fig. 3c also allows some discrimination between the samples of manufacturers 1 (Bayer) and 2 (Rhodia, Roche and McNeil). The samples of manufacturer 1 are grouped in the centre of the PC1–PC2 score plot while the samples of McNeil have high PC1 and low PC2 values and those of Rhodia and Roche low PC1 and high PC2 values.

In a second step, robust PCA was used to identify extreme observations. The data were standardized in a robust way before applying the method. The number of significant robust PCs is chosen in such a way that the cumulative percentage of explained variability is above 80%. For the data set used, 6 RPCs explain 89% of the information. A diagnostic plot, showing robust versus orthogonal distances to the robust PCA space, was made considering these 6 RPCs and is shown in Fig. 4. It shows that sample 39 produced by McNeil (USA) can be considered as a bad leverage observation. The samples from Reckitt and Nicholas (India), Whitehall-Robins (USA), Nycomed (Denmark) and Nanjing (China) are good leverage observations. One sample produced by Bayer (USA) can be considered as orthogonal outlier.

These results confirm that the elimination of sample 39 in the above PCA analysis was necessary. However, its elimination is not required for projection pursuit nor for multiple factor analysis, the results of which are further presented. The score plots of the

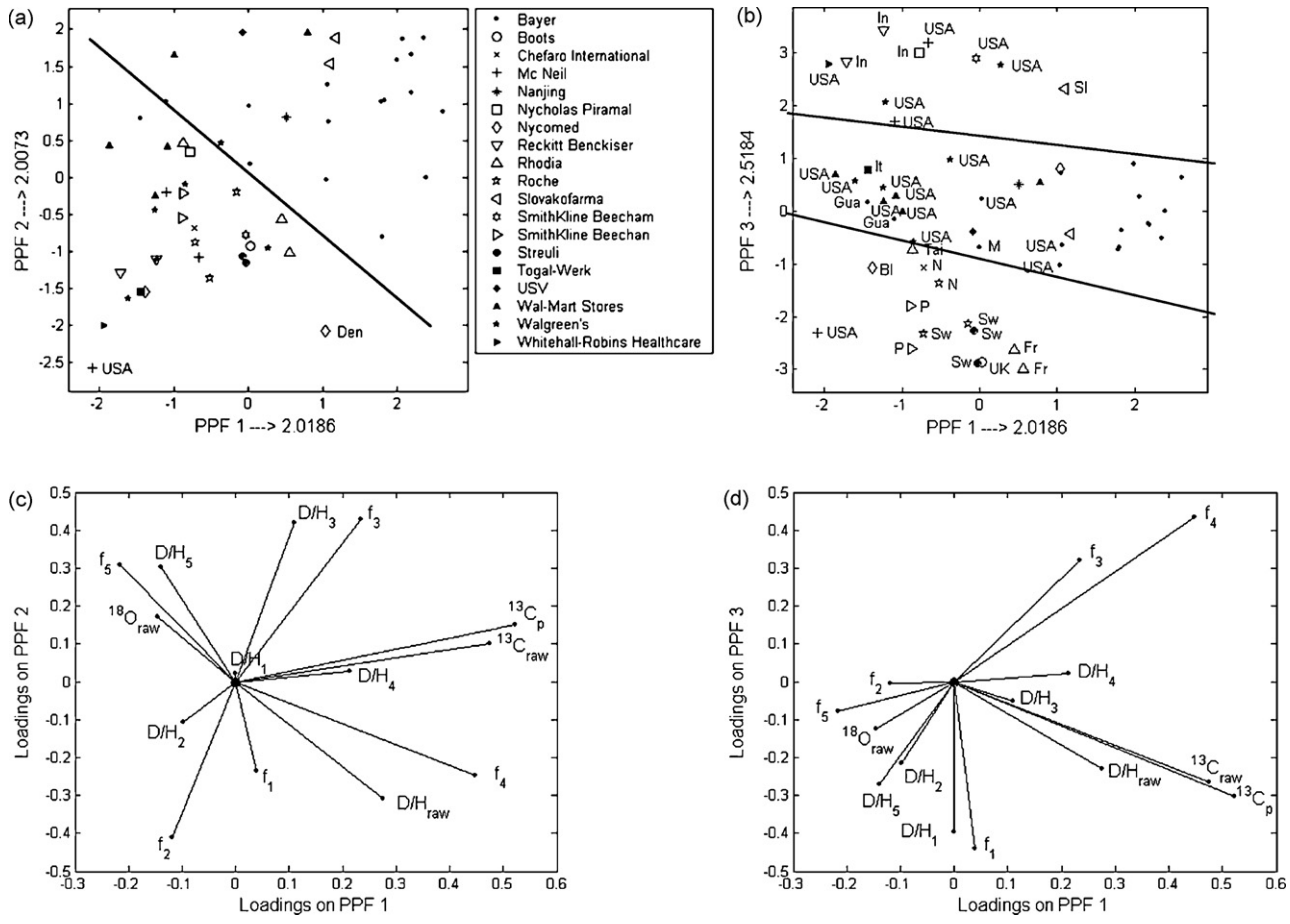

Fig. 5. Projection pursuit analysis on the acetyl salicylic acid data set with PPFs obtained by minimising the kurtosis projection index: (a) PPF1–PPF2 score plot; (b) PPF1–PPF3 score plot; (c) PPF1–PPF2 loading plot; (d) PPF1–PPF3 loading plot.

different RPCs did not allow discrimination of groups of samples and are not shown.

The projection pursuit approach reveals inhomogeneities in the data structure. In this study, the kurtosis was used as projection index [22,23]. PPFs obtained by minimising the kurtosis permit revealing clusters in the data (Fig. 5). Fig. 5a does not reveal a clustering tendency. Only two samples, from McNeil (USA) and Nycomed (Denmark) can be distinguished from the others. The plot allows a distinction between the samples of manufacturers 1 and 2. All samples of manufacturer 1, characterised by higher PPF 1 and PPF 2 values, are situated above the cut-off line drawn in Fig. 5a, while those of manufacturer 2 are situated below this line. The PPF1–PPF3 plot (Fig. 5b) allows distinction of three domains. The first is situated at the upper part of the plot and contains 6 USA samples (2 from Walgreen's, 1 of Whitehall-Robins, 2 from McNeil and 1 of SmithKline Beecham), 3 from India (2 of Reckitt and 1 of Nicholas) and 1 from Slovakia (Slovakofarma). The second group, situated at the lower part of the plot contains all French samples from Rhodia, all samples from Switzerland (Roche, Streuli), two samples from Portugal (SmithKline Beecham), one from the UK (Boots) and one from the USA (McNeil). The rest of the samples belong to the third domain situated intermediate to the previous two in the middle of the plot. This differentiation in groups can be explained in terms of loadings on the respective PPFs (Fig. 5c and d). Most samples of manufacturer 1 are characterised by high $^{13}C_p$, $^{13}C_{raw}$ and f_4 values (Fig. 5c). Fig. 5d shows that the first group of Fig. 5b has high values for f_4 and low for f_1, while group 3 in the middle exhibits lower f_4 and higher f_1 values. The second group of Fig. 5b is characterised by low values for f_4 and high for f_1.

PPFs obtained by maximising the kurtosis allow identification of extreme samples and possible outliers. The results are shown in Fig. 6. Four objects are far away from the majority. Three, i.e. two from the USA (McNeil and Whitehall-Robins) and one from Denmark (Nycomed) are distinguished in the PPF1–PPF2 plot (Fig. 6a). One, originating from China (Nanjing), appears as extreme along PPF3 (Fig. 6b). The extreme location of the McNeil sample is mainly due to the loadings of f_2 and f_5 on PPF1 (Fig. 6c). The two American samples have low $^{18}O_{raw}$ values while the Chinese sample has high f_4 but low f_1 values (Fig. 6d).

In a last step MFA was applied to the data set. MFA is PCA, in which blocks of variables are weighted to avoid a too high influence of some groups of variables in the analysis. For instance, in the data set there are five f_i-variables and only two $^{13}C_{raw}$. Therefore the information of the f_i-variables will mask that of the other variables. Therefore the groups of variables are weighted in MFA. In this study, the variables of a group are weighted by the first eigenvalue obtained after PCA, applied to this group. Four blocks of variables were distinguished: block 1 with $^{13}C_{raw}$ and $^{13}C_p$; block 2 with $^{18}O_{raw}$; block 3 with the six D/H_i and block 4 with the five f_i. The results of the MFA analysis are shown in Fig. 7.

Three principal components of MFA explain around 67% of the total data variance. PC1 is mainly associated with $^{13}C_{raw}$, $^{13}C_p$ and to a smaller degree with D/H_1, D/H_2, D/H_4 and D/H_5. PC2 contrasts variables f_1 and f_2 on the one hand and f_4 and f_5 on the other. PC3 is mainly related with variables D/H_3, D/H_4, $^{18}O_{raw}$ and to a smaller degree D/H_1, D/H_2 and f_5 (Fig. 7c and d).

Along PC 1, one cluster and a group of extreme observations can be distinguished (Fig. 7a). The extreme samples are originating

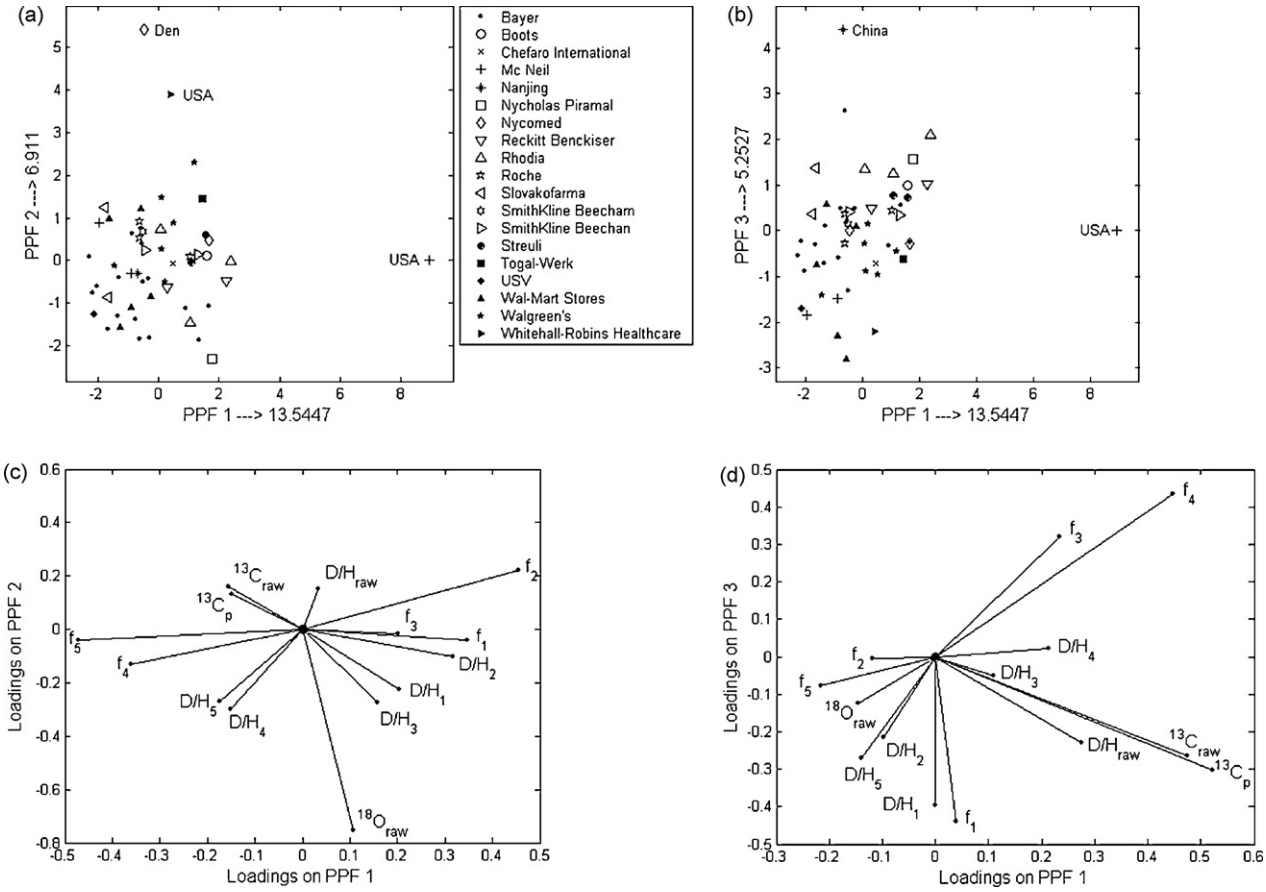

Fig. 6. Projection pursuit analysis on the acetyl salicylic acid data set with PPFs obtained by maximising the kurtosis projection index: (a) PPF1–PPF2 score plot; (b) PPF1–PPF3 score plot; (c) PPF1–PPF2 loading plot; (d) PPF1–PPF3 loading plot.

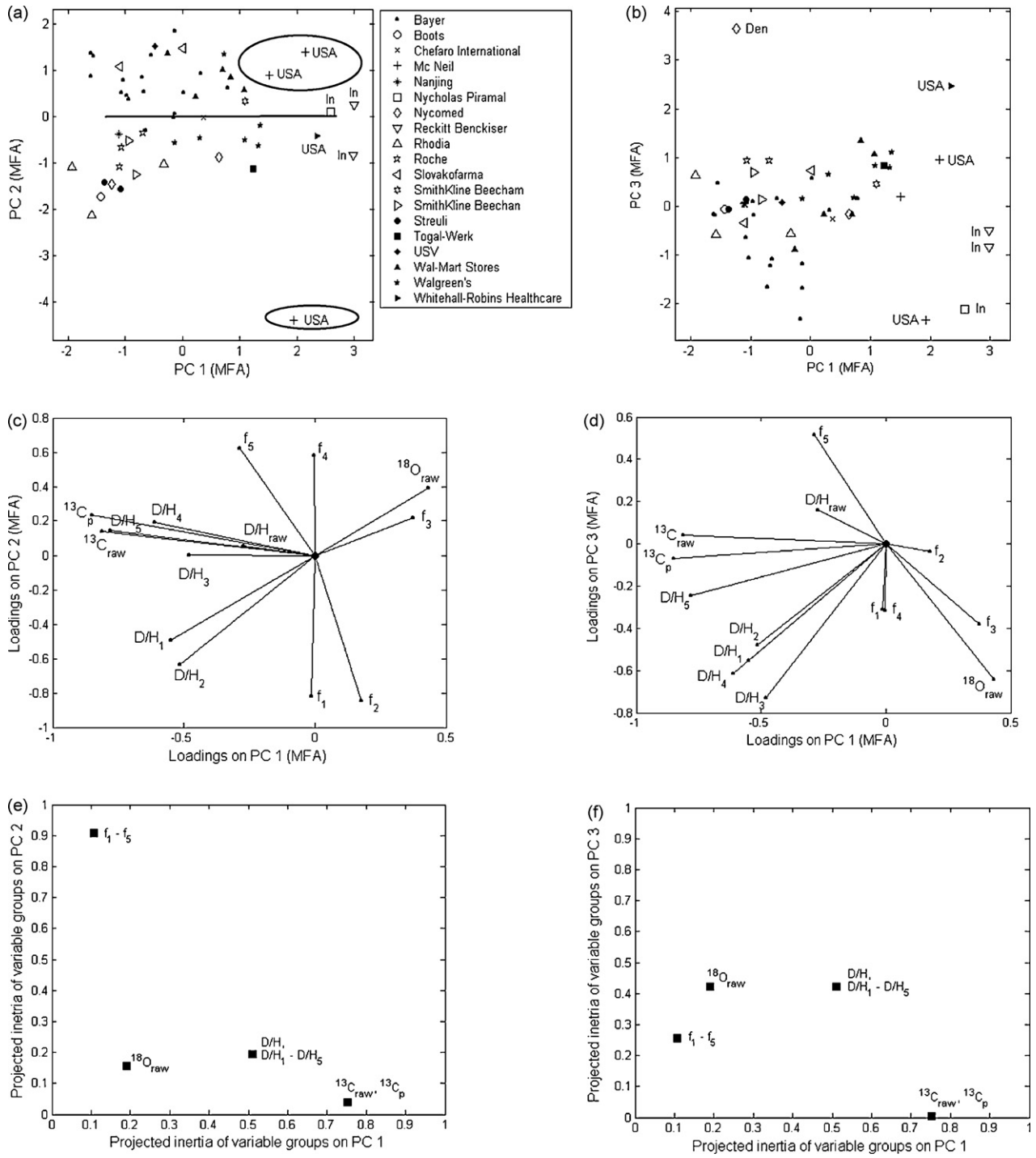

Fig. 7. Multiple factor analysis on the acetyl salicylic acid data set: (a) PC1–PC2 score plot; (b) PC1–PC3 score plot; (c) PC1–PC2 loading plot; (d) PC1–PC3 loading plot; (e) projection of groups' inertia on PC1 and PC2; (f) projection of groups' inertia on PC1 and PC3.

from India (Nicholas and Reckitt) and the USA (Whitehall-Robins and McNeil). They exhibit low values for $^{13}C_p$, $^{13}C_{raw}$, D/H_1, D/H_2, D/H_4 and D/H_5. One USA sample (McNeil) can also be distinguished along PC2, due to its somewhat higher values for f_1 and f_2 as well as lower values for f_4 and f_5. Another extreme sample can be detected along PC3 (Fig. 7b). It is characterised by high f_5 and low D/H_3, D/H_4, $^{18}O_{raw}$, D/H_1 and D/H_2 values compared to all other samples.

Fig. 7a also allows distinguishing the samples of manufacturers 1 and 2. The samples of manufacturer 1 are situated above the cut-off line drawn in Fig. 7a and are characterised by high values along PC2. The samples from manufacturer 2 are situated below this line, with exception of the three extreme encircled McNeil (USA) samples.

The first principal component is associated with the variables from the first and the third block, whereas the second PC is related with the fourth block (Fig. 7e). The third PC explains the variance of the samples according to the second and the third block (Fig. 7f).

4.2.2. Interpretation and general trends

Above it was shown that the different chemometric techniques are able to distinguish groups in the data. In this section, an attempt

Table 3
Calculations of $\delta^{13}C$ values for acetyl salicylic acid

	Phenol	CO_2 carboxylation	Acetyl anhydride	Theoretical $\delta^{13}C$ (‰)
Origins	Petrochemical (−26‰)	Natural gas (−45‰)	Natural gas (−45‰)	−33
	Petrochemical (−26‰)	Petrochemical (−26‰)	Natural gas (−45‰)	−30
	Petrochemical (−26‰)	Natural gas (−45‰)	Petrochemical (−26‰)	
	Petrochemical (−26‰)	Petrochemical (−26‰)	Petrochemical (−26‰)	−26

is made to explain the observed groups in terms of synthesis pathways and production processes.

Since for acetyl salicylic acid there is only one synthetic pathway, the repartition of the groups in the different plots points to the problem of traceability. The explanation of the different groups can be found in the origin of the phenol and the carboxylation conditions.

The projection pursuit analysis distinguished three groups (Fig. 5b). The $^{13}C_p$ values play an important role in determining these groups (see Fig. 5d). A thorough investigation of the $^{13}C_p$ values revealed the existence of three groups: a first group with a mean $\delta^{13}C$ value below −33‰, a second below −30‰ and a third below −26‰.

The calculation of the $\delta^{13}C$ values for acetyl salicylic acid is based on the following equation:

$$\delta^{13}C_{\text{theoretical}} = \frac{6 \times \delta^{13}C_{\text{aromatic cycle}} + \delta^{13}C_{CO_2\ \text{carboxylation}} + 2 \times \delta^{13}C_{\text{acetyl anhydride}}}{9} \quad (3)$$

Table 3 shows the calculations for acetyl salicylic acid, based on the origin of the primary substances. The theoretically calculated ^{13}C values confirm the PP analysis results. The samples originating from McNeil (USA), manufacturers 1 and 2, clearly occupy different regions in the PPF-score plot. These regions can be associated with clearly different ^{13}C contents, related to the origin of the primary substances, used in the synthesis. Fig. 5b also highlights some atypical samples. The first are the Bayer samples originating from Mexico and Guatemala, which are associated with high ^{13}C contents (−26‰). The hypothesis is that these production sites are supplied with salicylic acid from petrochemical origin, like all Bayer production sites, but they are free to choose the origin of the CO_2 necessary for carboxylation (possibly petrochemical origin) and of the acetyl anhydride. The acetyl anhydride probably originates from a synthesis from plant residues, with a ^{13}C content higher than the acetyl anhydride commonly used in Europe and the USA.

Another atypical sample is a Rhodia sample originating from Thailand, with a high ^{13}C content (−25‰). The production site in Thailand is only responsible for the acetylation step. Probably they perform the carbonylation by using acetic acid (petrochemical ori-

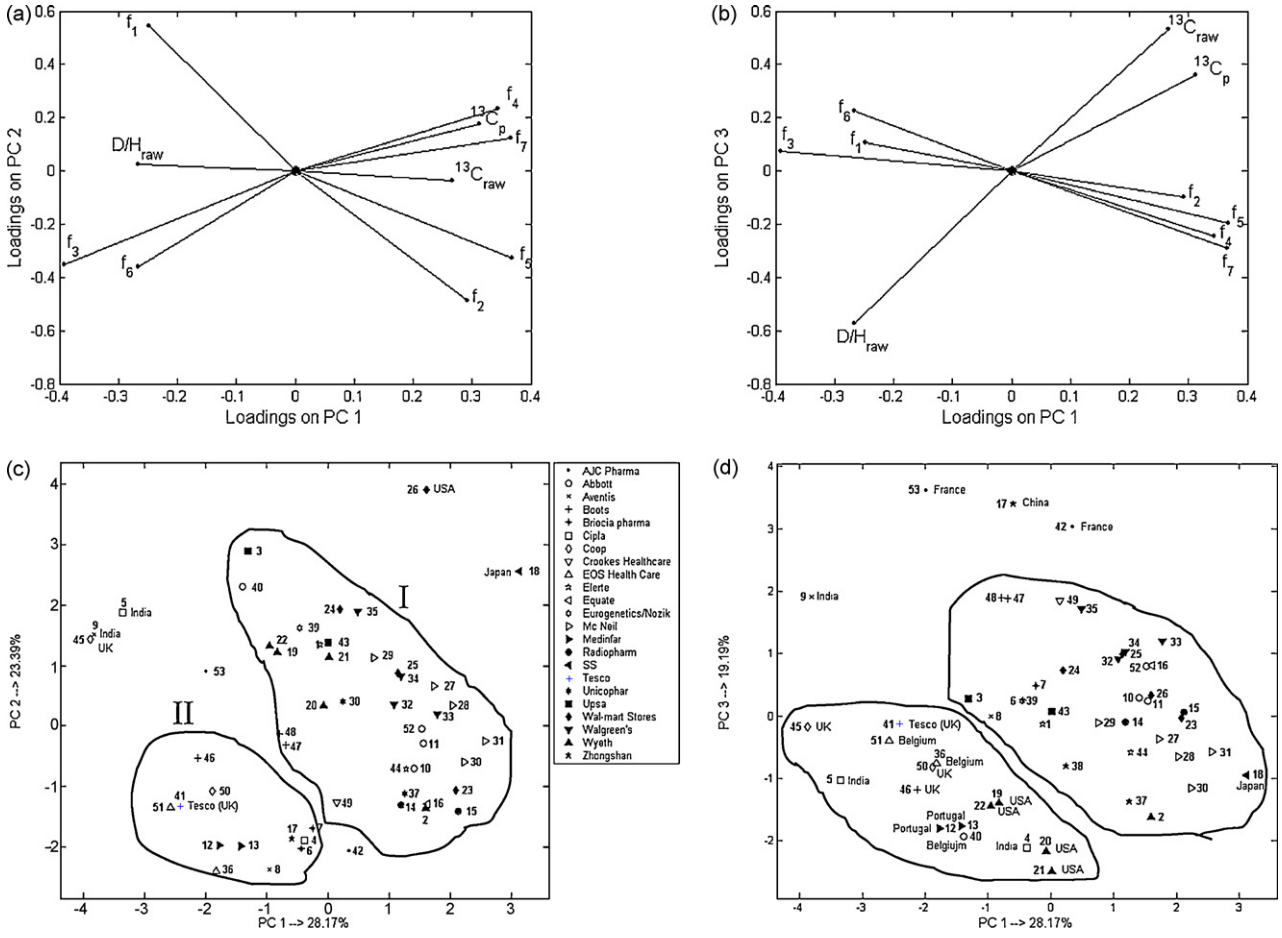

Fig. 8. Principal component analysis on the ibuprofen data set: (a) PC1–PC2 loading plot; (b) PC1–PC3 loading plot; (c) PC1–PC2 score plot; (d) PC1–PC3 score plot.

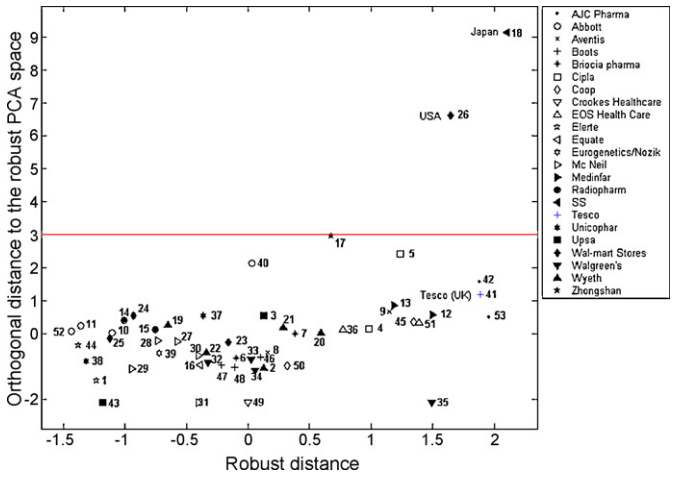

Fig. 9. Robust principal component analysis on the ibuprofen data set: distance–distance plot for 4 robust PCs.

gin) and not by using methanol (natural gas origin), as is commonly done in Europe.

4.3. The ibuprofen data set

4.3.1. Chemometric exploration

The same approach as for the acetyl salicylic acid data set was applied to the ibuprofen data set. The first three principal components of the PCA analysis explained about 71% of the variance in the data set. The loading plots (Fig. 8a and b) show that PC1 is associated with two groups of variables. The first includes f_4, f_5, f_7 and to a smaller extent f_2, $^{13}C_{raw}$ and $^{13}C_p$, whereas the second concerns mainly f_3 and to a smaller degree D/H_{raw}, f_1 and f_6. PC2 clearly contrasts between f_1 and f_2, f_3, f_5 and f_6, while PC3 is associated with D/H_{raw}, $^{13}C_{raw}$ and $^{13}C_p$. These loading plots also show that some variables are correlated. The pairwise absolute correlation coefficients show that the variables are in general less correlated than those obtained for acetyl salicylic acid. $^{13}C_{raw}$ and D/H_{raw} show a absolute correlation coefficient of 0.73, which means that they give rather similar information. Among the molar fractions, f_1 and f_2 are highly correlated (0.80), while f_5 shows some correlation with f_1 and f_2 ($R = 0.60$ and $R = 0.63$, respectively).

The PC1–PC2 score plot (Fig. 8c) shows a distinction in two groups. However several extreme samples can be identified. They originate from India (Cipla and Aventis), the UK (Coop), the USA (Wal-mart Stores) and Japan (SS). The two samples from India and the one from the UK are grouped on the left hand side of the score plot and are characterised by high f_3, D/H, f_1 and f_6 values and by low f_4, f_5, f_7, f_2, $^{13}C_{raw}$ and $^{13}C_p$ isotopic ratios. The PC1–PC3 plot (Fig. 8d) also reveals two groups in the data. The first contains four samples purchased in the USA (Wyeth), four in the UK (Coop, Boots, Tosco), three in Belgium (EOS Health Care, Abbott), two in India (Cipla) and two in Portugal (Medinfar). All other samples, except four originating from France (AJC Pharma), China (Zhongshan) and India (Aventis), which can be considered as extreme objects, belong to the second group. The first group is less dispersed than the sec-

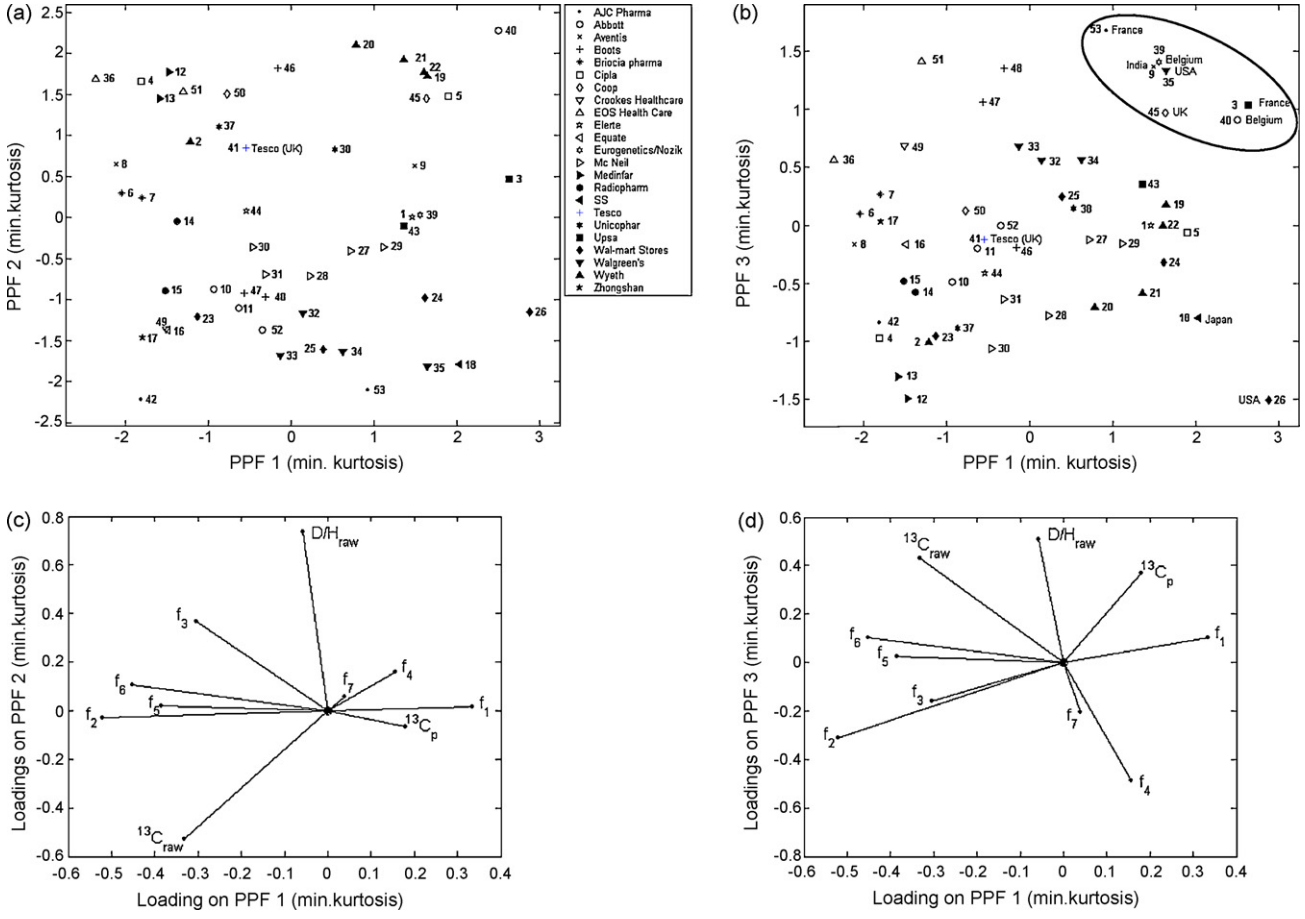

Fig. 10. Projection pursuit analysis on the ibuprofen data set with PPFs obtained by minimising kurtosis: (a) PPF1–PPF2 score plot; (b) PPF1–PPF3 score plot; (c) PPF1–PPF2 loading plot; (d) PPF1–PPF3 loading plot.

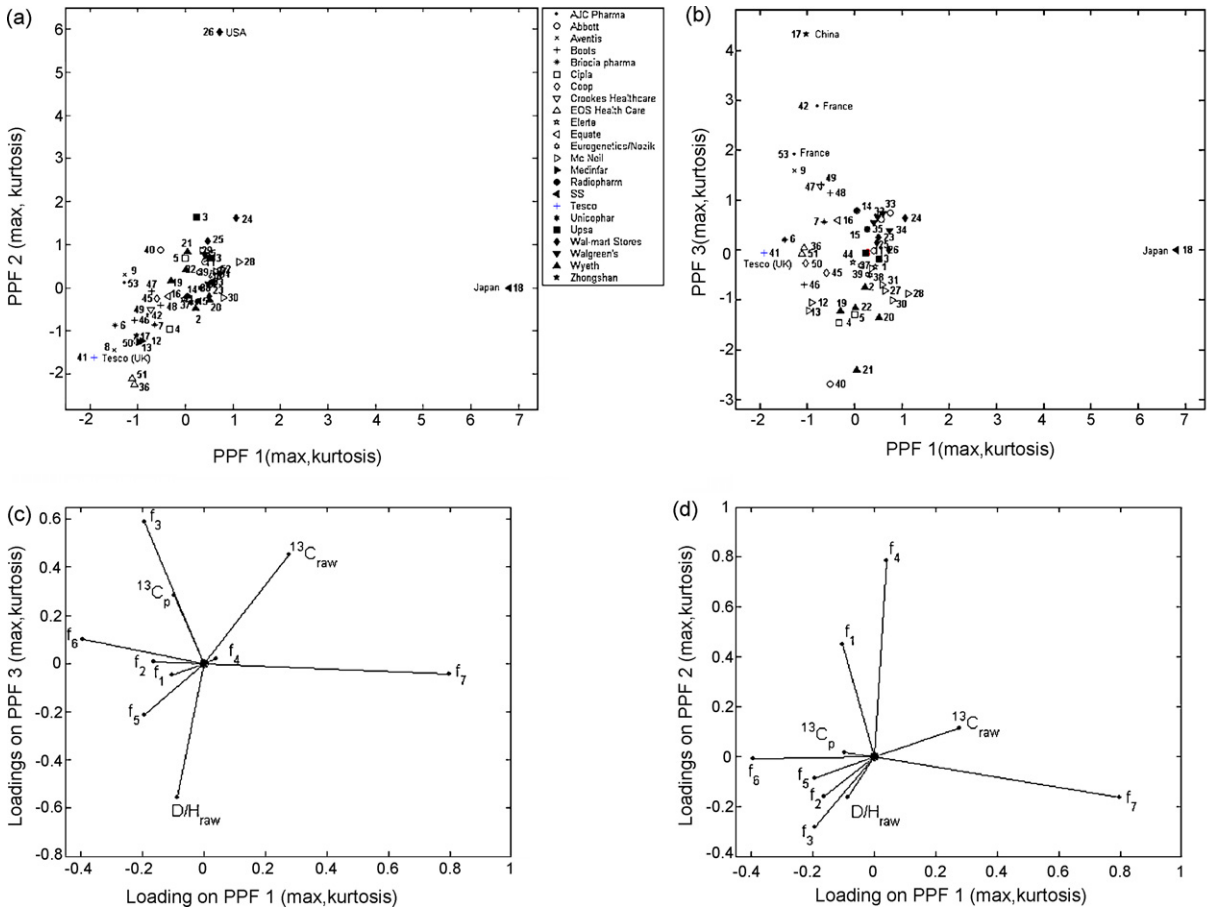

Fig. 11. Projection pursuit analysis on the ibuprofen data set with PPFs obtained by maximising kurtosis: (a) PPF1–PPF2 score plot; (b) PPF1–PPF3 score plot; (c) PPF1–PPF3 loading plot; (d) PPF1–PPF2 loading plot.

ond, indicating less similarities between samples in the latter. The segregation between both groups is due to somewhat higher f_3, D/H_{raw}, f_1, f_6 and lower f_4, f_5, f_7, f_2, $^{13}C_{raw}$, $^{13}C_p$ isotopic ratios for the second group. The extreme samples along PC3 originating from France and China exhibit higher $^{13}C_{raw}$ and $^{13}C_p$ values as well as lower D/H_{raw} values compared to the others. The extreme sample from India is characterised by high $^{13}C_{raw}$, $^{13}C_p$, f_1, f_3 and f_6 values and low f_2, f_4, f_5, f_7 and D/H values.

When RPCA was applied, four principal components explain 81% of the information. The diagnostic plot was created with these four RPCs and presented in Fig. 9. Samples 26 (Wal-mart Stores, USA) and 18 (SS, Japan) are orthogonal outliers, i.e. they have high residuals from the robust PCA model. Their elimination is, however, not required for projection pursuit and multiple factor analyses, the results of which are presented further. In the score plots no groups could be distinguished (not shown).

The score plots obtained by applying projection pursuit with minimising kurtosis are shown in Fig. 10. The PPF1–PPF2 score plot (Fig. 10a) does not reveal any clear clustering tendency, while on the PPF1–PPF3 plot (Fig. 10b), one segregated group of samples can be observed. It contains samples purchased in India (Aventis), France (AJC Pharma, Upsa), Belgium (Abbott, Eurogenerics/Nozik), the USA (Walgreen's) and the UK (Coop). The group can be interpreted in terms of the loadings on PPF1 and PPF3 (Fig. 10c and d). Its members are characterised by low f_2, f_4, f_5 and f_6 values, and high $^{13}C_{raw}$, $^{13}C_p$ and D/H values.

Fig. 11 shows the results obtained by maximising the kurtosis. Two samples, far away from the majority of the data, can be distinguished in Fig. 11a. One originates from Japan (SS) and is detected along PPF1 and the other from the USA (Wal-mart Stores), along PPF2. One Chinese (Zhongshan) and one French (AJC Pharma) sample appear as extreme ones along PPF3 (Fig. 11b).

The extreme location of the Japanese (SS) sample is mainly due to the high loading of f_7 on PPF1 (Fig. 11c and d). The American (Wal-mart Stores) sample is high in f_4, whereas the Chinese (Zhongshan) and French (AJC Pharma) samples have high values for f_3 and $^{13}C_{raw}$, but low for D/H.

The ibuprofen data set revealed three blocks of variables: block 1 with $^{13}C_{raw}$ and $^{13}C_p$; block 2 with D/H_i; and block 3 with f_i. The results of the MFA analysis are shown in Fig. 12. Three principal components of MFA explain around 70% of the total data variance. PC1 contrasts $^{13}C_{raw}$ and D/H_{raw} on the one hand and $^{13}C_p$ on the other (Fig. 12c). PC2 is related to two groups of variables. The first contains variables f_4 and f_7, and the second f_3 and f_6. PC3 differentiates the samples according to f_1 on the one hand, and f_2 and f_5 on the other (Fig. 12d).

The PC1–PC2 score plot allows the distinction between six groups of similar samples (Fig. 12a). Group I contains all Wyeth (USA) samples and one Abbott sample (Belgium). Group II contains 2 Medinfar (Portugal), 2 EOS Health Care (Belgium), 2 Coop (UK), 1 Boots (UK), 1 Tesco (UK) and 2 Cipla (India) samples. Group III contains 4 Indian samples (Briocia Pharma, Aventis). One Aventis sample is rather far away from the other Indian samples along PC2. Group IV contains 3 samples from UK (Boots, Crookes Healthcare). Group V contains 2 AJC samples (France) and 1 Zongshan (China) sample. Group VI contains the rest of the samples.

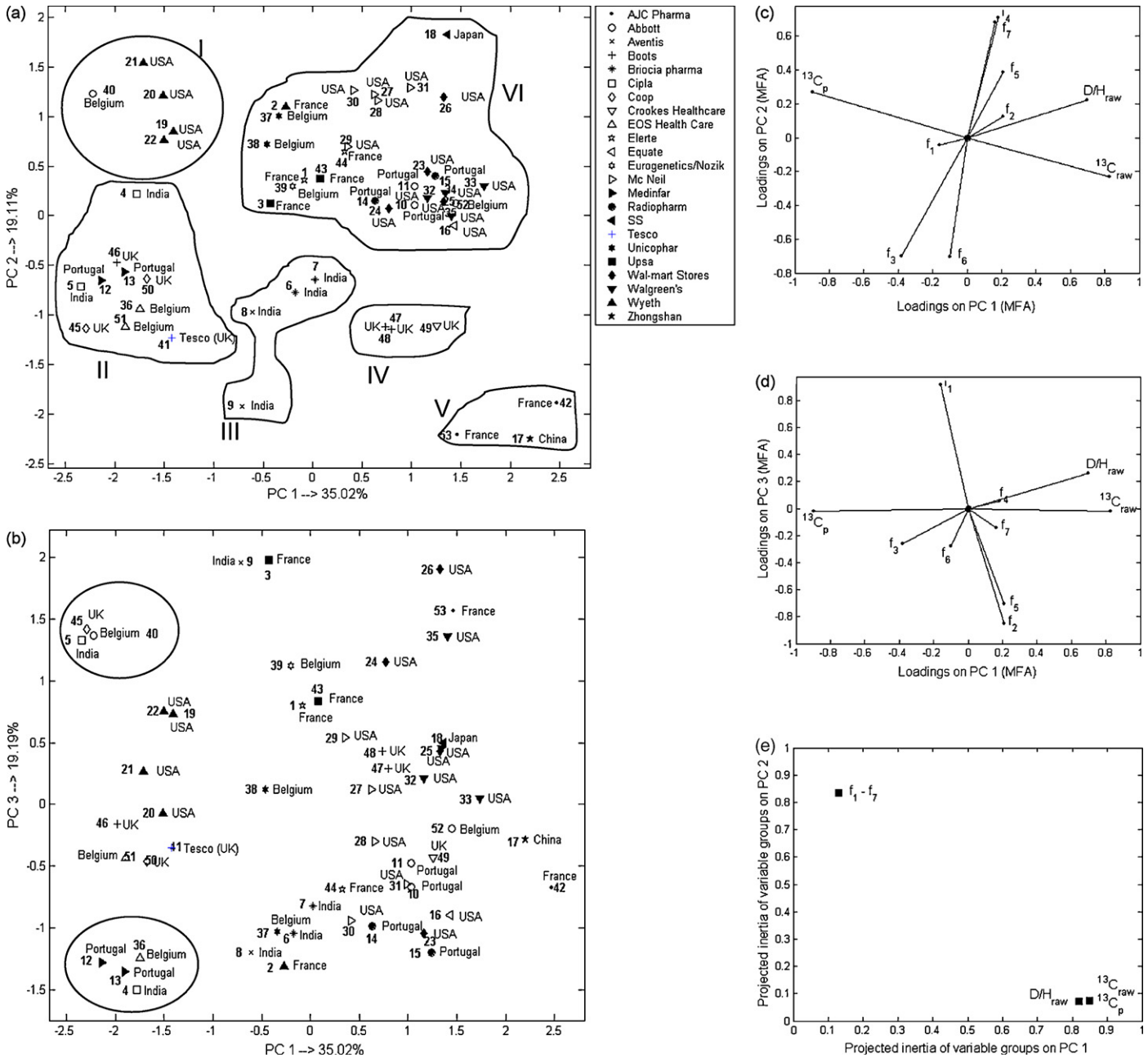

Fig. 12. Multiple factor analysis on the ibuprofen data set: (a) PC1–PC2 score plot; (b) PC1–PC3 score plot; (c) PC1–PC2 loading plot; (d) PC1–PC3 loading plot; (e) projection of groups' inertia on PC1 and PC2.

Generally, samples belonging to groups II, III, IV and V are segregated from the rest of the samples along PC2, due to higher values for f_3 and f_6, and lower for f_4 and f_7. Along PC1, groups I and II are distinguished and show large $^{13}C_p$ values. The other samples have high D/H_{raw} and $^{13}C_{raw}$ values.

It is important to notice that samples originating from a certain country and manufacturer are close to each other in the plot. For instance, all the Wyeth samples purchased in the USA can be distinguished from the other USA samples along PC1 due to their higher $^{13}C_p$ and lower D/H_{raw} and $^{13}C_{raw}$. McNeil samples are also found very close to each other. Another example is demonstrated with the Portuguese samples originating from Medinfar. They can be distinguished from all the other Portuguese samples by higher f_3, f_6, $^{13}C_p$ and lower f_4, f_7, D/H_{raw}, $^{13}C_{raw}$ ratios. The same conclusion holds true for EOS Health Care samples from Belgium.

An additional distinction between samples in group II can be made along PC3 (see Fig. 12b). Samples produced by Medinfar (Portugal), EOS Health Care (Belgium) and Cipla (India) can be distinguished from those of Coop (UK), Abbott (Belgium) and another Cipla (India) sample by their higher f_2, f_5 and lower f_1 values.

Fig. 12e shows that the first PC is associated with D/H_{raw}, $^{13}C_{raw}$ and $^{13}C_p$ whereas the second and the third (Fig. 12d) PC is mainly related with the f_i group.

4.3.2. Interpretation and general trends

For the ibuprofen data set the problem is situated in the characterisation of the synthetic pathways and of the primary substances used. Two groups can be distinguished from the PCA analysis. This discrimination is based on the variables f_7 and $^{13}C_{raw}$ on the one hand and f_6 and f_3 on the other. A thorough study of the variation

Table 4
Recalculated D/H_3 and D/H_6 values for the ibuprofen data set

	D/H_3 recalculated	D/H_6 recalculated	^{13}C (‰)
Group I, $n = 34$	94.8, S.D. = 15.2	125.8, S.D. = 4.7	−29
Group II, $n = 20$	125.5, S.D. = 19.0	137.8, S.D. = 7.0	−28

of the $^{13}C_{raw}$ values and the D/H_3 and D/H_6 (recalculated based on the f_3 and f_6 fractions as well as on a D/H_{tot} value estimated at 144.8 ppm) reveals the presence of two groups as can be seen from Table 4. Group I can be associated with the first synthetic pathway shown in Fig. 2, and denoted as the carbonylation process. This is confirmed by a low value for D/H_3, that corresponds to a hydrogenation with H_2 from gaseous origin (D/H_{gas} = 70–100 ppm). The value of D/H_6 can be associated with the ^2H content of the acetyl anhydride. The ^{13}C content of the carboxyl group would be originating from gas, as is normal when the carbonylation group is followed.

Since the second and the third synthetic pathway (Fig. 2b and c) are very labour intensive, they are probably not used anymore. Therefore, it is probable that group II corresponds to the fourth pathway (Fig. 2b), called the Darzen synthesis. The high D/H_6 values obtained for the samples in group II seem to show that exchanges took place between the methyl group and the aqueous environment (D/H_{water} = 150 ppm). This probably took place in the basic stage, characteristic for the Darzen synthesis route. The D/H_3 values for the samples of group II are higher than those of group I, pointing to the fact that the origin of the carboxyl group is different. Indeed in the Darzen's route the carboxyl group has a combined origin, petrochemical–gaseous, while in the carbonylation route it only has a gaseous origin.

Even if no immediate, synthesis related explanation can be given for the six groups distinguished in the MFA analysis (Fig. 12a), the clustering can be compared to that obtained with PCA. A similar distinction is observed. It seems that group I in the PC1–PC2 score plot corresponds to groups I, IV and VI in the MFA analysis, and group II to groups II and III. The samples that were extreme in the PCA plot (Fig. 8c) were assigned to a group in MFA (Fig. 12a). The samples of group V are dispersed in the PCA plot, since one belongs to group I in PCA, another to group II and one was unclassified (sample 42). Therefore, group V could not be assigned to a synthetic pathway.

5. Conclusions

In general, it can be stated that there is a good indication that isotopic ratios might be useful for recognising the origin of drugs, both geographically as well as based on synthetic process.

For both data sets it was possible to distinguish samples from a very different origin, like the Indian samples, from the others. Samples from the same company and origin are usually found close together, which indicates that the reproducibility of the measurements is good enough and that the influence of the between batch variation can be neglected. This also confirms the potential of the used techniques to discriminate groups and in the detection of counterfeiting and infringements of patents in the pharmaceutical industry.

An attempt was also made to explain the observed groups in relation to the possible synthetic processes. A good correlation between the hypotheses and the observations was found.

Acknowledgements

The PhD studies of E. Deconinck were funded with a specialization grant from the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT) and those of A.M. van Nederkassel from the Fund for Scientific Research (Flanders).

This publication contains results from the EU Project Counterpharm, a shared-cost RTD project funded under the Fifth Framework Program of the European Community, within the Competitive and Sustainable Growth Program, Measurement and Testing Activity (G6RD-CT-2002-00849). The authors are solely responsible for the content of the publication. It does not represent the opinion of the Community, nor is the Community responsible for any use that might be made of the data herein.

The authors want to thank the Rhodia Company and especially Dr. J. Cavezzan for furnishing samples of salicylic acid and acetyl salicylic acid.

References

[1] R.D. Kirchhoefer, J. AOAC Int. 75 (1992) 577–580.
[2] I.V. Tetko, T.I. Aksenova, A.A. Patiokha, A.E.P. Villa, W.J. Welsh, W.L. Zielinski, D.J. Livingstone, Anal. Chem. 71 (1999) 2431–2439.
[3] Y. Juillet, A.P. Vlasto, Fundam. Clin. Pharmacol. 19 (2005) 621–624.
[4] M. ten Ham, Drug Safe. 26 (2003) 991–997.
[5] D. Taylor, Brit. Med. J. 304 (1992) 334.
[6] J. Wechsler, Biopharm. Int. 19 (2006) 30–33.
[7] A. Ault, Lancet 363 (2004) 714.
[8] M.C. Gaudiano, E. Antoniella, P. Bertocchi, L. Valvo, J. Pharm. Biomed. Anal. 42 (2006) 132–135.
[9] T.I. Aksenova, I.V. Tetko, A.G. Ivakhnenko, A.E.P. Villa, W.J. Welsh, W.L. Zielinski, Anal. Chem. 71 (1999) 2423–2430.
[10] S.H. Scafi, C. Pasquini, Analyst 126 (2001) 2218–2224.
[11] J. Koziet, A. Rossmann, G.J. Martin, P. Johnson, Anal. Chim. Acta 302 (1995) 29–37.
[12] G.J. Martin, J. Koziet, A. Rossmann, M.J. Dennis, Anal. Chim. Acta 321 (1996) 137–146.
[13] A. Rossmann, J. Koziet, G.J. Martin, M.J. Dennis, Anal. Chim. Acta 340 (1997) 21–29.
[14] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics—Part B, Elsevier Science, Amsterdam, 1998.
[15] M. Daszykowski, B. Walczak, D.L. Massart, Chemom. Intell. Lab. Syst. 65 (2003) 97–112.
[16] I. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, Chemom. Intell. Lab. Syst. 71 (2004) 83–95.
[17] B. Escofier, J. Pages, Comput. Stat. Data Anal. 18 (1994) 121–140.
[18] I. Stanimirova, B. Walczak, D.L. Massart, Anal. Chim. Acta 545 (2005) 1–12.
[19] P. Huber, Ann. Stat. 13 (1985) 435–475.
[20] M.C. Jones, R. Sibson, J. Roy. Stat. Soc. A 150 (1987) 1–36.
[21] C. Croux, A. Ruiz-Gazen, COMPSTAT: Proceedings in Computational Statistics 1996, Physica-Verlag, Heidelberg, 1996, pp. 211–217.
[22] M. Daszykowski, I. Stanimirova, B. Walczak, D. Coomans, Chemom. Intell. Lab. Syst. 78 (2005) 19–29.
[23] D. Pena, F. Prieto, Technometrics 43 (2001) 286–300.
[24] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York, 2001.
[25] I. Stanimirova, M. Daszykowski, E. Van Gyseghem, F.F. Bensaid, M. Lees, J. Smeyers-Verbeke, D.L. Massart, Y. Vander Heyden, Anal. Chim. Acta 552 (2005) 1–12.
[26] B.M. Brown, J. Roy. Stat. Soc. Ser. B 45 (1983) 25–30.
[27] C.G. Small, Int. Stat. Rev. 58 (1990) 263–277.
[28] P.J. Rousseeuw, C. Croux, J. Am. Stat. Assoc. 88 (1993) 1273–1283.
[29] C. Croux, A. Ruiz-Gazen, J. Multivariate Anal. 95 (2005) 206–226.
[30] P. Rousseeuw, M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, 1987.
[31] G.S. Remaud, Y.L. Martin, G.G. Martin, G.J. Martin, J. Agric. Food Chem. 45 (1997) 859–866.
[32] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics—Part A, Elsevier Science, Amsterdam, 1997.