Journal of Chromatography A, 1218 (2011) 2963–2975

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Comparison of different statistical approaches to evaluate the orthogonality of chromatographic separations: Application to reverse phase systems

Ramia Al Bakain[a], Isabelle Rivals[b], Patrick Sassiat[a], Didier Thiébaut[a], Marie-Claire Hennion[a], Guillaume Euvrard[b], Jérôme Vial[a,*]

[a] Laboratoire Sciences Analytiques, Bioanalytiques et Miniaturisation UMR – CNRS–UPMC – PECSA, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, ParisTech, 10 Rue Vauquelin, 75005 Paris, France
[b] Équipe de Statistique Appliquée, Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris, ParisTech, 10 Rue Vauquelin, 75005 Paris, France

ARTICLE INFO

Article history:
Received 22 September 2010
Received in revised form 9 March 2011
Accepted 15 March 2011
Available online 13 April 2011

Keywords:
RPLC
Orthogonality
ANOVA
Informational theory
Statistical tests of independence

ABSTRACT

Selectivity of phase system is of primary concern when designing a bidimensional chromatographic system and looking for the highest degree of orthogonality between the two separations. Several statistical or geometrical criteria can potentially be used to measure the degree of orthogonality. A comparison of eight candidate criteria has been carried out in this study. Analysis of variance (ANOVA) was used to evaluate the relevance of each criterion and its ability to reveal the significance of the influence of factors like pH, stationary phase, and organic modifier. Experimentally, a set of 32 chromatographic systems was evaluated by the same generic gradient with 63 probe solutes, likely to be present in biological and/or environmental samples and covering a wide range of physico-chemical properties: acidic, basic and neutral compounds with different p*K*a, molecular mass and hydrophobicity (log *P*). Each chromatographic system was defined by the nature of the stationary phase (8 different silica or grafting chemistries), the pH of the aqueous fraction of the mobile phase (2.5 or 7.0) and the nature of the organic modifier (acetonitrile or methanol). The orthogonality of the 496 couples of chromatographic systems was evaluated and ranked using the eight different approaches: the three correlation coefficients (Pearson, Spearman and Kendall), two geometric criteria characterizing the coverage of the 2D separation space, Slonecker's information similarity and two chi-square statistics of independence between normalized retention times. In fact, there were only seven distinct criteria, since we established the analytical equivalence between the rankings with the likelihood ratio statistics and Slonecker's information similarity. Kendall's correlation coefficient appeared to be the best measure of orthogonality since, according to ANOVA, it exhibited the highest sensitivity to all experimental factors. The chi-square measures, and hence Slonecker's information similarity, performed equally well provided the discretization of the separation space was carried out appropriately. Finally, from the compared study of the factors acting upon orthogonality carried out by ANOVA, it is possible to draw the conclusion that the pH of the mobile phases has the highest impact on the selectivity followed by the type of stationary phase and finally by the organic modifier.

1. Introduction

The demand for characterization of complex samples, *i.e.* containing several hundreds of compounds, is stronger than ever before and requires analytical tools to meet this increasing difficulty. Despite the recent progresses in column and instrument technology, the limitations of traditional one-dimensional analytical techniques such as liquid chromatography or gas chromatography are now reached since they only allow the separation of a hundred compounds in a reasonable time. Increasing the separation capacity is possible, but at the cost of a longer analysis time [1]. Because of their unequalled resolving power, multidimensional separations have received a great attention during the past few years for the detailed characterization of complex samples in the field of biology, pharmaceutical analysis, proteomics and metabolomics [2,3], environment [4,5] or petroleum industry [6]. If for volatile compounds, comprehensive gas chromatography [7] is a natural choice, for non volatile compounds, multidimensional liquid chromatography is the only option despite its lower degree of maturity.

The increase in resolution obtained in bidimensional liquid chromatography (2D-LC) depends on the degree of orthogonality

* Corresponding author. Tel.: +33 1 40 79 47 79; fax: +33 1 40 79 47 76.
E-mail address: jerome.vial@espci.fr (J. Vial).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.03.031

of the coupled systems, *i.e.* on the significance of the difference between the separation mechanisms they involve [8]. Dissimilar separation mechanisms are obtained when the retention of solutes results from different interactions between the solutes, the stationary phase and the mobile phase, *i.e.* the organic modifier and the pH could also have a dramatic effect as illustrated by studies on column characterization [9–11]. For instance, in reverse phase liquid chromatography (RPLC), acetonitrile and methanol exhibit significantly different selectivity [12,13]. Each organic modifier has a different influence on the solute–solvent interaction due to the difference in dipole moment, polarizability, hydrogen bond basicity and acidity. In RPLC, for charged molecules, the pH of the mobile phase also greatly influences the retention. The effect of pH and the fraction of the organic modifier have been widely studied [14–19], with a single organic modifier, methanol [20–33] or acetonitrile [34]. Moreover, the choice of the probe solutes has also a striking impact on the evaluation of the orthogonality of couples of chromatographic systems [35], but it has not been studied extensively so far.

Up to now, several approaches have been proposed to evaluate the degree of orthogonality of two chromatographic systems. Liu et al. [36] used the retention times to establish a correlation matrix, from which a peak spreading angle matrix was calculated using a geometric approach to factor analysis. In this paper, the authors defined the orthogonality using the effective area of the 2D separation space covered by the eluting peaks. Slonecker et al. [37,38] developed criteria for describing the independence of separation modes using information theory, *i.e.* the informational similarity and the synentropy percentage for the description of data scatter plot in 2D-separation space. The authors also used additional descriptors, such as peak spreading angle and practical peak capacity (N_P) introduced earlier by Liu et al. [36]. However, both mathematical approaches have some limitations. First, multiple descriptors are used to define the orthogonality. Second, the proposed methods may not satisfactorily describe the orthogonality for the situations where the analytes are not distributed diagonally along the 2D separation space but form several distinct clusters not intersecting the diagonal. More recently, Gilar et al. [39,40] developed a geometric characterization of data orthogonality. The 2D separation space was divided in rectangular bins, and an orthogonality percentage was defined as the difference between the number of occupied bins and the number of bins on the diagonal, divided by the number of bins expected to be occupied in the case of an ideally orthogonal distribution. In [39], Gilar also showed that having different pH values of the mobile phase was a very powerful method for separating charged solutes in RPLC. The correlation coefficient is a frequently used parameter to evaluate the orthogonality of the two dimensions [35,41–48]. Van Gyseghem et al. [45,47] used Pearson's correlation coefficient to evaluate the orthogonality for eight silica-based stationary phases that were applied in conjunction with four mobile phases at different pH values to determine the impurity profile of a drug. Similarly, Forlay-Frick et al. [49] attempted at comparing the three classical coefficients, Pearson's, Spearman's and Kendall's together with a generalization of the pair-correlation method (PCM) combined to different statistical tests [50]. Recently, another approach was applied to select the orthogonal columns for cationic drug solutes by using Snyder–Dolan (S–D) hydrophobic subtraction method of column classification [51]. The advantage of this model is that a single parameter called the "column selectivity function, F_s" can be used to quantitatively compare the overall selectivity of any two columns. This approach assumes that the column behavior is the same whatever the conditions (organic modifier fraction and type, temperature, solvent type, etc.). In [52], this model was also applied to non ionized solutes.

The present paper aims at comparing the criteria we consider as most relevant for orthogonality evaluation in RPLC × RPLC, at establishing which one(s) is (are) most appropriate, and at determining quantitatively the factors having the largest influence on orthogonality. To this end, a set of 63 test compounds, covering a wide distribution of physico-chemical properties, was built in order to probe orthogonality between couples of RP chromatographic systems in a generic gradient mode. This set includes neutral, acidic and basic compounds differing by their p*Ka* values (between 0.6 and 14.0), their molecular mass (between 76.12 and 1485.71 g/mol), their hydrophobicity (log *P*-values are evenly distributed from −1.08 to 7.72) and the presence of heteroatoms. To ensure a perfect accessibility to the testing procedure, test compounds had to be easily available, meaning cheap and not forensic products with sufficient stability. The retention times of these compounds were measured with every combination of the eight different columns (*i.e.* stationary phases), the two different organic modifiers (methanol and acetonitrile) and the two different pH values (2.5 and 7.0), *i.e.* with $8 \times 2 \times 2 = 32$ distinct chromatographic systems. The orthogonality of the 496 system couples was evaluated and ranked with the eight criteria we considered most relevant: the three classical correlation coefficients (Pearson, Spearman and Kendall), two geometric criteria characterizing the coverage of the 2D separation space, Slonecker's information similarity and two chi-square statistics of independence. Since we establish the equivalence between the rankings with the likelihood ratio statistics and Slonecker's information similarity, see Section 3.2, there are in fact only seven distinct criteria. Each of them was evaluated according to its capacity to reveal the influence of the factors acting upon orthogonality using ANOVA. Finally, the most orthogonal chromatographic systems among the ones evaluated are presented.

2. Experimental

2.1. Instrumentation

Gradient separations were carried out using a Dionex HPLC system (Ultimate™ 3000 Nano HPLC) equipped with a UV detector (Ultimate™ 3000 variable wavelength) operated at 3 detection wavelengths: 220, 230, and 250 nm depending on the solute (rate of data acquisition was 2.5 Hz, time constant was 0.40 s, conventional 2.5 μl cell with 7.5 mm path length), two pumps (Ultimate 3000), a degasser (LPG-3000), a thermostatic automated autosampler (Ultimate™ 3000 series Nano/Cap) and a column oven (Ultimate 3000 column compartment).

2.2. Chemicals and reagents

Acetonitrile (HPLC ultra gradient grade) and methanol (HPLC ultra gradient grade) originated from Carlo Erba Reactifs (Val de Reuil-France). Ultrapure water for HPLC mobile phases was produced by a Milli-Q Plus purification system (Millipore, Molsheim, France). Phosphoric acid (85%) and potassium phosphate were obtained from PROLABO, whereas hydrochloric acid was from Carlo Erba. Tris base [tris(hydroxymethyl)amino-methane] was supplied by Sigma.

2.3. Preparation of samples solutions

Stock solutions of the 63 test compounds were prepared at the concentration of ≈1000 μg/ml in pure methanol. The injection solutions for the chromatographic runs were diluted from the stock solutions in methanol/water 50/50 (v/v) in order to provide an UV absorbance around 200 mAU (milli-absorbance unit), (the range of concentrations was 10–500 ppm). Most of

Table 1
Characteristics of the compounds of the test set.

No	Compound	Manufacturer	Molecular structure	Molar- mass	p*K*a	Log*P*
1.	Phenanthrene	Jansen	$C_{14}H_{10}$	178.23		4.46[54]
2.	Naphthalene	Aldrich	$C_{10}H_8$	128.17		3.30[54]
3.	Anthracene	Prolabo	$C_{14}H_{10}$	178.23		4.45[54]
4.	Triphenylene	Fluka	$C_{18}H_{12}$	228.3		5.49[9,11,58]
5.	Salicylic acid	Prolabo	$C_7H_6O_3$	138.12	2.97[54]	2.26[54]
6.	4-Hydroxybenzoic acid	Aldrich	$HOC_6H_4CO_2H$	138.12	4.54[54]	1.58[54]
7.	4-Hydroxy-3-methylbenzoic acid	Aldrich	$HOC_6H_3(CH_3)CO_2H$	152.15	4.26/9.78[53]	1.84[53]
8.	Benzoic acid	Merck	C_6H_5COOH	122.12	4.19[54]	1.87[54]
9.	Mandelic acid	Touzart & Matignon	$C_6H_5CH(OH)CO_2H$	152.15	3.41[54]	0.62[54]
10.	(S)-(+)-Ibuprofen	Aldrich	$(CH_3)_2CHCH_2C_6H_4CH(CH_3)CO_2H$	206.28	4.51[55]	3.6[56]
11.	Phenylacetic acid	Ega	$C_6H_5CH_2CO_2H$	136.15	4.31[54]	1.41[54]
12.	p-Toluamide	Lancaster	$CH_3C_6H_4CONH_2$	135.16		1.18[54]
13.	Loperamide hydrochloride	Sigma	$C_{29}H_{33}ClN_2O_2\cdots HCl$	513.5	N,F	5.15[9,11,58]
14.	Benzamide	Sigma	$C_6H_5CONH_2$	121.14	1.82[54]	0.64[54]
15.	Phenol	Acros	C_6H_5OH	94.11	9.99[9,11,58]	1.46[9,11,58]
16.	1-Naphthol	Fluka	$C_{10}H_7OH$	144.17	9.34[54]	2.85[54]
17.	4-Isopropylphenol	Aldrich	$(CH_3)_2CHC_6H_4OH$	136.19	10.2[54]	2.9[54]
18.	4-Dodecylresorcinol	Aldrich	$C_{18}H_{30}O_2$	278.43	11.61/9.2[53]	6.77[53]
19.	Piperonal	Aldrich	$C_8H_6O_3$	150.13		0.64[54]
20.	3-Hydroxybenzaldehyde	Aldrich	HOC_6H_4CHO	122.12	8.98[54]	1.29[54]
21.	3,4-Dichloroaniline	Aldrich	$Cl_2C_6H_3NH_2$	162.02	2.97[54]	2.69[54]
22.	2,4,6-Trichloroaniline	Aldrich	$Cl_3C_6H_2NH_2$	196.46		3.52[54]
23.	Bromacil	Dr Ehrenstorfer	$C_9H_{13}BrN_2O_2$	261.12	9.30[54]	2.11[54]
24.	Napropamid	Dr Ehrenstorfer	$C_{10}H_7OCH(CH_3)CON(CH_2CH_3)_2$	271.35		3.36[54]
25.	Vinclozolin	Dr Ehrenstorfer/Fluka	$C_{12}H_9Cl_2NO_3$	286.11		3.1[54]
26.	Carbaryl	Fluka	$C_{10}H_7OCONHCH_3$	201.22		2.36[54]
27.	Diuron	Dr Ehrenstorfer	$C_9H_{10}Cl_2N_2O$	233.09		2.68[54]
28.	Monuron	Dr Ehrenstorfer	$ClC_6H_4NHCON(CH_3)_2$	198.65		1.94[54]
29.	Linuron	Fluka	$C_9H_{10}Cl_2N_2O_2$	249.09		3.20[54]
30.	Atrazine-desisopropyl	Dr Ehrenstorfer	$C_5H_8ClN_5$	173.6		1.15[54]
31.	Prometryn	Dr Ehrenstorfer	$C_{10}H_{19}N_5S$	241.36	4.1[54]	3.51[54]
32.	Atraton	Dr Ehrenstorfer	$C_9H_{17}N_5O$	211.26		2.69[54]
33.	Toluene	Aldrich	$C_6H_5CH_3$	92.14		2.73[9,11,58]
34.	Ethylbenzene	Aldrich	$C_6H_5C_2H_5$	106.17		3.15[9,11,58]
35.	Propylbenzene	Fluka	$C_6H_5CH_2CH_2CH_3$	120.19		3.69[54]
36.	Butylbenzene	Aldrich	$C_6H_5(CH_2)_3CH_3$	134.22		4.38[54]
37.	Pentylbenzene	Aldrich	$C_6H_5(CH_2)_4CH_3$	148.24		4.9[9,11,58]
38.	Imipramine hydrochloride	Sigma	$C_{19}H_{24}N_2\cdots HCl$	316.87	9.4[9,11,58]	4.8[9,11,58]
39.	Caffeine	Fluka	$C_8H_{10}N_4O_2$	194.19	0.6/14.0[9,11,58]	−0.07[9,11,58]
40.	Phenothiazine	Aldrich	$C_{12}H_9NS$	199.27	2.52[54]	4.15[54]
41.	Carbazole	Aldrich	$C_{12}H9N$	167.21		3.72[54]
42.	Umbelliferone	Sigma	$C_9H_6O_3$	162.14		1.03[54]
43.	Nicotine	Sigma	$C_{10}H_{14}N_2$	162.23	3.10[54]/8.02[57]	1.17[54]
44.	1,2-Phenylenediamine	Merck	$C_6H_4(NH_2)_2$	108.14	4.47[54]	0.15[54]
45.	Methyl 4-hydroxybenzoate	Merck	$HOC_6H_4CO_2CH_3$	152.15		1.96[54]
46.	Propyl 4-hydroxybenzoate	Merck	$HOC_6H_4CO_2CH_2CH_2CH_3$	180.21	7.91[54]	3.04[54]
47.	Bis(2,4,6-trichlorophenyl) oxalate	Fluka	$C_{14}H_4Cl_6O_4$	448.9		7.72[53]
48.	Estrone	Sigma	$C_{18}H_{22}O_2$	270.37		3.13[54]
49.	Cortisone	Merck	$C_{21}H_{28}O_5$	360.44		1.47[54]
50.	Estriol	Sigma	$C_{18}H_{24}O_3$	288.38		2.45[54]
51.	Benzylamine	Fluka	$C_6H_5CH_2NH_2$	107.15	9.33[9,11,58]	1.09[9,11,58]
52.	Clofazimine	Sigma	$C_{27}H_{22}Cl_2N_4$	473.4	7.57[9,11,58]	7.66[9,11,58]
53.	Strychnine hemisulfate salt	Sigma	$C_{21}H_{22}N_2O_2\cdots 1/2H_2SO_4$	383.45	8.26[9,11,58]	1.93[9,11,58]
54.	o-Terphenyl	Fluka	$C_6H_5C_6H_4C_6H_5$	230.3		5.52[9,11,58]
55.	Digitoxin	Sigma	$C_{41}H_{64}O_{13}$	764.94		1.85[9,11,58]
56.	Thiourea	Aldrich	NH_2CSNH_2	76.12	2.03[54]	−1.08[54]
57.	Ampicillin	Sigma	$C_{16}H_{18}N_3NaO_4S$	371.39	3.7/7.3[9,11,58]	1.35[9,11,58]
58.	Vancomycin	Sigma	$C_{66}H_{75}Cl_2N_9O_{24}$	1485.71	3.6/8.2/9/9.2/10.3/10.8[9,11,58]	N,F
59.	Amiodarone hydrochloride	Sigma	$C_{25}H_{29}I_2NO_3\cdots HCl$	681.77	8.73[9,11,58]	7.57[9,11,58]
60.	(+)-Tubocurarine chloride-hydrate	Sigma	$C_{37}H_{42}Cl_2N_2O_6$	681.65	8.1/9.1[9,11,58]	N,F
61.	Atropine	Sigma	$C_{17}H_{23}NO_3$	289.37	9.43[9,11,58]	1.83[9,11,58]
62.	Phloroglucinol	Merck	$C_6H_6O_3$	126.11	8.45[54]	0.16[54]
63.	Cyanocobalamin	Sigma	$C_{63}H_{88}CoN_{14}O_{14}P$	1355.37	7.64[9,11,58]	3.57[9,11,58]

the solutes were detected at 220 nm except: phenanthrene, anthracene, triphenylene, naphthalene, toluene, ethylbenzene, propylbenzene, butylbenzene, pentylbenzene, phenothiazine, carbazole and o-terphenyl, which were detected at 250 nm. Between injections, the samples were stored at 4 °C or less to avoid degradation. The characteristics of the solutes are gathered in Table 1 where p*K*a and log *P* values were obtained from Refs. [9,11,53–58]).

2.4. Buffer preparation

The mobile phases were buffered [59–61] using 5 mM of KH_2PO_4 for pH 2.5 and 5 mM of Tris base for pH 7.0. The choice of these buffers has been guided by their buffer capacity at the chosen pH rather than by their volatility. In fact, coupling with mass spectrometry was not the aim of the present study. They were prepared by dissolving the accurate quantity of each salt in pure water sepa-

Table 2
Characteristics of the 8 stationary phases used in this study.

Stationary phase	Endcapping	Manufacturer	pH range	%C	Surface area (m^2/g)	Pore size (A°)	Grafting
XBridge shield RP18	Yes	Waters, Ireland	2–11	17	185	135	Polar embedded-Octadecyl
Kromasil C18	Yes	Macherey-Nagel, Germany	1–10	20	330	110	Octadecyl
Zorbax SB-CN	No	Agilent, USA	1.8–8	4	180	80	Cyanopropyl
Luna C8(2)	Yes	Phenomenex, USA	1.5–10	13.5	400	100	Octylsilane
Luna Phenyl-Hexyl	Yes	Phenomenex, USA	1.5–10	17.5	400	100	Phenyl-Hexyl
Discovery HS PEG	No	Supelco, USA	2–8	12	300	120	Polyethyleneglycol
Discovery HS F5	Yes	Supelco, USA	2–8	12	300	120	Pentafluorophenylpropyl
Capcell Pak SG C18		Shiseido, Japan	2–9	14	300	120	Octadecyl

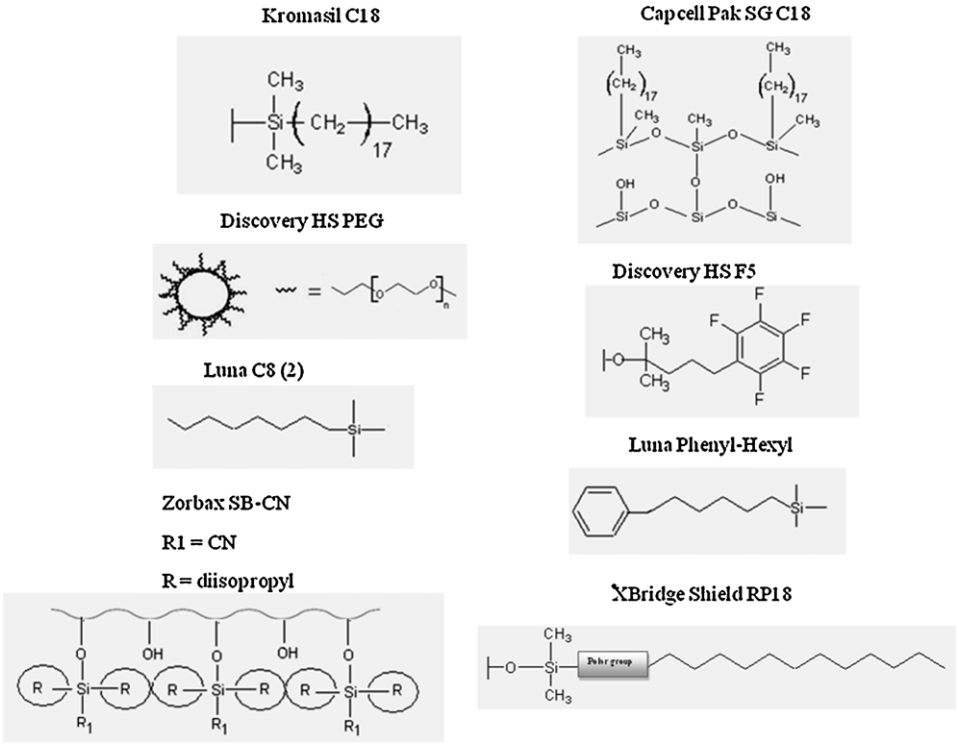

Fig. 1. Structures of the different stationary phases.

rately. After making up to 2 L in a volumetric flask with pure water, pH value of 2.5 was adjusted with phosphoric acid, while for the value of pH 7.0, hydrochloric acid was used. All buffers were filtered through 0.45 μm HA type filters (Millipore, Moleshiem, France) before mixing with acetonitrile (MeCN) or methanol (MeOH) in the desired volume ratio. Then the mixtures were degassed by ultrasonication for 20 min immediately before use.

2.5. Stationary phases

The testing procedure has been applied to 8 stationary phases, the physical and chemical properties of which can be found in Table 2. The set of columns consisted of different stationary phases that are commonly used for reversed phase liquid chromatography. In all cases, column dimensions were 150 mm × 4.6 mm I.D. with 5 μm particles diameter, except for Zorbax SB-CN filled with 3.5 μm particles. The chosen RP columns differed from each other in the grafting and protection against residual silanol groups. They were chosen because they are structurally very different from each other, as illustrated in Fig. 1.

2.6. Running conditions

The composition of the mobile phases used is given in Table 3. A linear generic gradient was systematically used. Mobile phase

Table 3
pH and compositions of the mobile phases.

pH	Mobile phase A	Mobile phase B
2.5	Potassium phosphate in water–MeCN 90:10% Potassium phosphate in water–MeOH 90:10%	Potassium phosphate in water–MeCN 10:90% Potassium phosphate in water–MeOH 10:90%
7.0	Tris-base in water–MeCN 90:10% Tris-base in water–MeOH 90:10%	Tris-base in water–MeCN 10:90% Tris-base in water–MeOH 10:90%

Table 4
2*2 contingency table of the retention time coordinates when separation space discretization is necessary.

Y / X	b_1	b_2	…	b_q	total
a_1	n_{11}	n_{12}	…	n_{1q}	n_{1+}
a_2	n_{21}	n_{22}	…	n_{2q}	n_{2+}
…	…	…	…	…	…
a_p	n_{p1}	n_{p2}	…	n_{pq}	n_{p+}
total	n_{+1}	n_{+2}	…	n_{+q}	n

B increased from 0 to 100% in 30 min, then 100% of B was maintained for 10 min. Some compounds with certain phases were not eluted using this gradient. In that case, the plateau at 100% B was maintained until 200 min. After the end of each gradient run, the composition of the mobile phase was gradually set back to the starting conditions and 15 column volumes were pumped for equilibration before starting the next analysis. Mobile phases were freshly prepared just before use to avoid any degradation. To ensure a stable baseline, at least 1 h of equilibration was performed for each mobile phase before the injection of 1.0 μl of adequate mixtures of tested compounds. Two consecutive repeated injections were done, and the mean of the retention times was registered, because they could not be considered as true independent repetition of the whole analytical process. In addition, each solute was injected individually for identification purposes. Three wavelengths 220, 230 and 250 nm were used depending on the compound. The column hold-up volumes were measured as the elution volumes of Thiourea since it is not retained.

System back pressure without column was 30 bars with MeCN/buffer, and it was 37 bars for MeOH/buffer. Column temperature was kept constant at 35 °C during the overall tests in the oven. All runs were operated at a flow rate of 1.0 ml/min, checked by using a burette and a stopwatch. The dwell volume of the system was 0.70 ml.

2.7. Software

All chromatographic data acquisition and processing were conducted using Chromeleon (6.8 chromatography data system) software. The statistical analysis was performed with MATLAB (Mathworks, Natick, MA, USA) software v. 7.7.0.471 (2008b).

3. Statistical analysis

In statistics, two random variables X and Y are orthogonal if the mathematical expectation of their product is zero, *i.e.* $E(X\,Y) = 0$. Since $\mathrm{cov}(X,Y) = E(X\,Y) - E(X)\,E(Y)$, this is often the consequence of their independence or only absence of correlation, provided one of them has zero mean. In 2D-LC, the mean of the retention times playing no role in the selectivity, what is searched for is indeed the maximal independence of the separation mechanisms. Quantifying the so-called orthogonality of a couple of chromatographic systems amounts hence to estimate the degree of independence or only uncorrelation of the retention time coordinates in the 2D separation space, using a representative set of probe-molecules.

Thus, it seems natural to resort to classical independence χ^2 statistics as potential orthogonality criteria, though to our knowledge, this has never been done before. Not only is it natural, but we will show that the maximum likelihood ratio χ^2 statistic is equivalent for ranking systems according to their orthogonality to the information similarity proposed by Slonecker et al. [37].

One can also have recourse to classical measures of non-correlation (Pearson's, Spearman's and Kendall's correlation coefficients), as already proposed in Forlay-Frick et al. [49]. However, in the latter study, the correlation coefficients were not directly used to rank the couples of systems, but the correlation matrix was thresholded and summed over rows in order to obtain an orthogonality measure for each system, not for each couple, as will be detailed in Section 3.3. Here, we will use the three correlation coefficients for ranking the couples of chromatographic systems.

Finally, geometric criteria characterizing the coverage of the separation space can be derived as in Liu et al. [36] or Gilar et al. [39] specifically for chromatographic applications, or more generally. We derive two criteria, the area of a confidence ellipse for supposedly Gaussian retention times in the 2D separation space, and a percentage of coverage of the 2D separation space that makes no distributional assumption.

Note that distances (e.g. Euclidian or Manhattan) could also be considered, but they clearly favor the detection of points along the diagonal, and do not so much characterize independence or uncorrelation. Since the latter is what we are searching for, they were not included in the tested criteria. Along the same line, we did not include generalized PCM in the criteria extensively used in [49]. As a matter of fact, PCM aims at choosing between two independent variables X_1 and X_2 that are positively and almost equally correlated with a dependent response variable Y, the one that best explains Y. But for our problem, no variable (no chromatographic system retention time) has the status of response variable.

In the following, the set of the retention time coordinates of the n molecules is denoted by $\{(x^1, y^1), \ldots, (x^n, y^n)\}$, the $\{x^i\}$ and $\{y^i\}$ being considered as realizations of two random variables X and Y. We work on the retention times scaled with respect to the least and most retained molecules in the data set, the $\{x^i\}$ and $\{i\}$ being hence systematically in the interval [0,1].

Some of the considered criteria require the discretization of the 2D separation space into $p * q$ rectangular bins, *i.e.* the definition of the contingency table of the observed counts, as shown in Table 4. The event a_i is realized if X belongs to the interval $[(i-1)/p; i/p]$, and in the same way, b_i is realized if Y belongs to the interval $[(j-1)/q; j/q]$.

3.1. Classical chi-square statistics of independence

There are two well-known χ^2 statistics of independence [62]. Both are measures of the dissimilarity between the observed counts $\{n_{ij}\}$ and the expected counts under the independence assumption, *i.e.* $(n_{i+}\,n_{+j})/n$. Pearson's χ^2 statistic is derived by considering the multinomial distribution and the central limit theorem:

$$d^2 = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{(n_{ij} - n_{i+}n_{+j}/n)^2}{(n_{i+}n_{+j}/n)}$$

If independence holds, d^2 is asymptotically χ^2 distributed with $(p-1)(q-1)$ degrees of freedom (d.o.f.). Maximum likelihood estimation leads to the likelihood-ratio statistic:

$$g^2 = 2\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} \log\left(\frac{n_{ij}n}{n_{i+}n_{+j}}\right)$$

$$= 2n\log(n) + 2\sum_{i}\sum_{j} n_{ij} \log\left(\frac{n_{ij}}{n_{i+}n_{+j}}\right)$$

Under the same assumptions, g^2 is also χ^2 distributed with $(p-1)(q-1)$ d.o.f. Both statistics can be normalized: $\max(d^2) = n$

min $(p-1, q-1)$, max $(g^2) = 2n\log(\min(p, q))$. However, since our interest in this paper is mainly to rank couples of systems according to their increasing orthogonality, there is no need here for an absolute measure. This is also the reason why we do not attempt to use the p-values associated to the statistics given their asymptotic distributions. The disadvantage of these statistics is to require the discretization of the separation space, *i.e.* a suitable choice of p and q.

3.2. Information theory quantities

In Shannon's information theory, event uncertainty is called "information" or "informational entropy": the less likely the observed value x of X, the more informative it is. The information represented by the observation of x is measured in bits by $-\log_2(P(X=x))$. In the following, we will equivalently use natural logarithms. If X can take the values $a_1, \ldots, a_p$, with $P(X=a_i)=p_i$, the entropy of X is defined as:

$$H(X) = -\sum_{i=1}^{p} p_i \log p_i$$

and similarly for Y. Let $p_{ij} = P((X=a_i)\cap(Y=b_j))$. The joint entropy of X and Y is defined as the quantity:

$$H(X, Y) = -\sum_{i=1}^{p}\sum_{j=1}^{q} p_{ij} \log p_{ij}$$

If X and Y are independent: $p_{ij} = p_i\ p_j$, and $H(X, Y) = H(X) + H(Y)$. The mutual entropy of X and Y is defined as the quantity:

$$I(X;\ Y) = H(X) + H(Y) - H(X,\ Y)$$

We have $I(X; Y) \geq 0$, and $I(X; Y) = 0$ only if X and Y are independent. Thus, $I(X;\ Y)$ measures the degree of independence of X and Y. In [37], a normalized version of $I(X;\ Y)$ is proposed as orthogonality criterion.

Let us show the link with the likelihood-ratio χ^2 statistic. The entropies are estimated using the data set by replacing the theoretical probabilities $\{p_i\}$, $\{p_j\}$ and $\{p_{ij}\}$ with the corresponding observed frequencies:

$$\hat{H}(X) = -\sum_i \frac{n_{i+}}{n}\log\left(\frac{n_{i+}}{n}\right), \quad \hat{H}(Y) = -\sum_j \frac{n_{+j}}{n}\log\left(\frac{n_{+j}}{n}\right),$$

$$\hat{H}(X, Y) = -\sum_i\sum_j \frac{nij}{n}\log\left(\frac{n_{ij}}{n}\right)$$

The mutual entropy is hence estimated with:

$$\hat{I}(X; Y) = \log(n) + \frac{1}{n}\sum_i\sum_j n_{ij}\log\left(\frac{n_{ij}}{n_{i+}n_{+j}}\right) = \frac{g^2}{2n}$$

Thus, mutual information (measured in natural logarithms) and the likelihood-ratio χ^2 statistic differ only by a constant factor (twice the number of points). They are hence equivalent for the ranking of couples of chromatographic systems according to their orthogonality.

In [37], a second criterion was also developed, that is the percentage of mutual information accounted for by the diagonal, called "percentage of synentropy". Since it was designed to measure the degree of clustering along the diagonal rather than orthogonality, we did not include it in the tested criteria.

3.3. Correlation coefficients

Pearson's, Spearman's and Kendall's coefficients are summarized with Daniel's formula:

$$r_D = \frac{\sum_{ij}\alpha^{ij}\beta^{ij}}{\sqrt{\sum_{ij}(\alpha^{ij})^2\sum_{ij}(\beta^{ij})^2}}$$

where

$$\text{Pearson}: \begin{cases}\alpha^{ij} = x^i - x^j\\ \beta^{ij} = y^i - y^j\end{cases}, \quad \text{Spearman}: \begin{cases}\alpha^{ij} = r^i - r^j\\ \beta^{ij} = s^i - s^j\end{cases},$$

$$\text{Kendall}: \begin{cases}\alpha^{ij} = \text{sign}(x^i - x^j) = \dfrac{x^i - x^j}{|x^i - x^j|}\\ \beta^{ij} = \text{sign}(y^i - y^j)\end{cases}$$

and the $\{r^i\}$ and $\{s^i\}$ denote the ranks of the $\{x^i\}$ and $\{y^i\}$, respectively [63,64]. Again, since our aim is only to rank, we do not try to use the p-values associated to the tests of uncorrelation.

All three coefficients have the advantage not to require the discretization of the separation space. However, Pearson's zero correlation is equivalent to independence only for Gaussian data, it does not measure a nonlinear association satisfactorily, and it is sensitive to extreme values. On the contrary, being less and less parametric, Spearman's and Kendall's coefficients are good measures of any nonlinear monotonous association, and are insensitive to extreme values, Kendall especially. As a matter of fact, r_K is simply equal to the difference between the numbers of concordant and discordant pairs, divided by the total number of pairs (note that it is equivalent here to Goodman and Kruskal's γ since there are no ties). Spearman and Kendall's coefficients are hence good candidates for orthogonality evaluation in the general case of a wide spectrum of solutes and chromatographic systems. In the case of a smaller and more homogenous set of solutes (typically of neutral solutes only) and when testing columns with quasi-normal distributions (like C18 stationary phases for example), one can expect Pearson's coefficient to be more sensitive to small changes of correlation. However it cannot be a good strategy to restrict the possibilities in this manner if orthogonal conditions are looked for.

The three coefficients were previously evaluated for the selection of orthogonal chromatographic systems in Forlay-Frick et al. [49], with the conclusion that Spearman's and Kendall's coefficients are "not sensitive enough to the orthogonality of these chromatographic systems". Unfortunately, this conclusion is based on an error. In [49], the three correlation matrices were computed for a set of $n_s = 38$ chromatographic systems. The aim being to characterize each column (not each couple of columns), the focus was made on each row of the correlation matrices, which were discretized in the following way: a correlation coefficient is replaced by +1 if it can be decided non zero with a 5% risk, by −1 if it cannot. For each row, the orthogonality ratio of the corresponding system is defined as the number of −1, divided by $n_s - 1$. It is hence close to 100% for a column that is systematically uncorrelated with the others. However, an error was made on the 5% limit value for Pearson's coefficient. The number of d.o.f. that must be considered for the significance of Pearson's coefficient is equal to n (the number of molecules) − 2, in their case 68 − 2 = 66, and the corresponding limit value at 5% is 0.239. Instead, [49] considered n_s (the number of chromatographic systems) − 2 d.o.f., *i.e.* 38 − 2 = 36 d.o.f., wrongly leading to a much higher limit value of 0.320. Since Spearman's and Kendall's limiting values were (almost) correctly taken at 0.237 and 0.166, respectively (*i.e.* the values corresponding to 68 − 2 = 66 d.o.f.), the orthogonality ratios derived from Pearson's coefficient were systematically much larger than the other two coefficients, and it was wrongly concluded that the other two coefficients "showed less dis-

crimination power for the different systems" [49]. The same error was made again in Van Gyseghem et al. [65] ($n_s - 2 = 46 - 2 = 44$ d.o.f. wrongly lead to a limit value of 0.291 instead of the correct value 0.239).

3.4. Geometric criteria

Some geometric criteria of orthogonality were developed as early as in [36], where an indicator of the coverage of the separation space, and hence of a "practical peak capacity" (N_p), is defined. This indicator being a monotonous function of Pearson's correlation coefficient, it would lead to the same ranking of chromatographic system couples than Pearson's coefficient itself.

More recently, [39] developed another geometric characterization of orthogonality. The 2D separation space is divided in rectangular bins, and an orthogonality percentage is defined as the difference between the number of occupied bins and the number of bins on the diagonal, divided by the number of bins expected to be occupied in the case of an ideally orthogonal distribution, *i.e.* two independent uniform distributions along each dimension. However, we think that it is not a good idea to favor the diagonal: as a matter of fact, the retention time coordinates of two strongly correlated systems can lie far away from the diagonal due, for example, to an extreme value together with the scaling of the retention times in the interval [0; 1]. Thus, we propose to test two simple geometric criteria of the coverage of the separation space that do not suffer from this drawback.

The first criterion does not necessitate space discretization, but it makes a distributional assumption. It is defined as the area covered by a given fraction of the data, assuming a Gaussian distribution, or confidence ellipse. A simple calculation shows that its area equals:

$$a(\alpha) = -2\pi \ln(\alpha)\sqrt{\lambda_x \lambda_y}$$

where $1 - \alpha$ is the chosen fraction, *i.e.* the confidence level, and λ_x and λ_y are the eigenvalues of the maximum likelihood estimate of the covariance matrix of the data: the largest $a(\alpha)$, the more orthogonal the two systems. The choice of α does not matter for ranking the couples of systems.

The second criterion, inspired from [39], does not make any assumption about the data, but it requires an adequate discretization of the separation space. It is simply defined as the percentage of occupied bins (the percentage of the area corresponding to the diagonal is not subtracted): the largest this percentage, the more orthogonal the two systems. Let $o_{ij} = 1$ if bin ij is occupied by retention time coordinates, and $o_{ij} = 0$ if not, then the separation space coverage percentage equals:

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{o_{ij}}{pq}$$

3.5. Summary of the tested criteria

Table 5 illustrates the criteria whose ranking of the chromatographic system couples are going to be evaluated. The criteria requiring the discretization of the separation space were tested with two resolutions: $p = q = 5$ and $p = q = 10$. Again, no attempt was made to normalize the corresponding computed quantities, but absolute values or a minus sign were applied to some of them so that the smaller the computed quantity, the more orthogonal the two systems.

3.6. Quantitative analysis of orthogonality through analysis of variance

Since we have $n_s = 32$ different chromatographic systems, we are going to rank the $n_s(n_s - 1)/2 = 496$ couples of systems by decreasing order of the computed quantities associated to our criteria. In order not to be dependent on the number of couples, instead of the ranks, we will equivalently consider scores in the interval [0, 1], the couple with a score of 0 (*i.e.* with rank 1) being the least orthogonal among all couples, and the couple with a score of 1 (*i.e.* with rank $n_s(n_s - 1)/2$) l the most orthogonal one.

These $n_s = 32$ systems result from all the possible combinations of:

- 8 stationary phases,
- 2 organic modifiers (MeOH or MeCN),
- 2 pH values of the mobile phase (2.5 or 7.0).

Since there are no true repetitions (the mean of the duplicate retention times was taken), the factors we will consider as candidates for contributing to orthogonality, (*i.e.* to high scores of couples) will be: the difference in stationary phase ($s\varphi =$ or $s\varphi \neq$), the difference in organic modifier (om= or om $\neq$) and the difference in pH value (pH= or pH $\neq$). According to these three factors, the 496 couples distributed among the 8 possible categories shown in Table 6.

We have chosen to evaluate the effects of the three factors with a classical three-way analysis of variance (ANOVA). Of course, many assumptions cannot be made, nor the Gaussian assumption (the scores are uniformly distributed in [0, 1], nor the independence assumption (many couples of systems have one system in common). However, we can nevertheless consider the ANOVA *F*-statistics and the associated *p*-values as normalized criteria of the significance of the factors. We computed the *p*-values associated to the main effects and two-by-two interactions (the interaction of the three factors cannot be tested due to the absence of couples with all factors equal).

Finally, in order to account for the finite character of the probe set (of $n = 63$ solutes), we have performed leave-one out (LOO) cross-validation for the whole procedure, *i.e.* the ANOVA was performed on $n = 63$ different sets of solutes, the *i*th set being obtained by removing the *i*th solute from the whole probe set. The means and standard deviations of the effects will be given as well as the frequencies of rejection of the associated null hypotheses (*i.e.* of the significance of the effects) with a type I error risk of $\alpha = 5\%$.

4. Results and discussion

4.1. Orthogonality evaluation

The scores of the 496 couples of chromatographic systems were computed using the seven criteria (twice for those requiring discretization, with 5×5 and 10×10 grids), and ANOVA was performed. For all criteria, there was never a significant interaction between the difference in pH or in stationary phase and the difference in organic modifier.

Thus, we performed the ANOVA without the two non-significant interactions with the organic modifier difference; hence, we had only five parameters. We did so, but with a parameterization which is more suited to our problem than the classic (centered) one. As shown in Table 7, it allows to characterize the mean score increase, *i.e.* the mean orthogonality increase, due to the factor modalities of interest, *i.e.* the differences in pH, stationary phase or organic modifier, as opposed to identity of the latter.

The meaning of the five parameters is the following:

Table 5
List of the tested criteria.

Criterion description	Computed quantity	Short name	Discretization
Pearson's correlation coefficient	$\|r_P\|$	r_P	No
Spearman's correlation coefficient	$\|r_S\|$	r_S	No
Kendall's correlation coefficient	$\|r_K\|$	r_K	No
Confidence ellipse	$-\lambda_x \cdot \lambda_y$	conf	No
Separation space coverage percentage	$1 - \sum\sum o_{ij}/pq$	%cov	Yes
Pearson's χ^2 statistics	d^2	d^2	Yes
Likelihood ratio χ^2 statistics	g^2	g^2	Yes

Table 6
Distribution of the 496 couples of chromatographic systems.

	om=		om ≠	
	$s\varphi=$	$s\varphi \neq$	$s\varphi=$	$s\varphi \neq$
pH=	0	112	16	112
pH ≠	16	112	16	112

Table 7
ANOVA reparametrization.

	om=		om ≠	
	$s\varphi=$	$s\varphi \neq$	$s\varphi=$	$s\varphi \neq$
pH=	m	$m+b=$	$m+c$	$m+b=+c$
pH ≠	$m+a$	$m+a+b \neq$	$m+a+c$	$m+a+b \neq +c$

- m represents the mean score when all three parameters are identical,
- a represents the mean score increase due to different pHs (whatever the organic modifiers),
- $b=$ represents the mean score increase due to different stationary phases when the pH values are equal (whatever the organic modifiers),
- $b \neq$ represents the mean score increase due to different stationary phases when the pHs are different (whatever the organic modifiers),
- c represents the mean score increase due to different organic modifiers (whatever the pHs and the stationary phases).

The values of a, $b=$, $b \neq$ and c (the means and standard deviations estimated with LOO) obtained with the different criteria, as well as the associated frequencies of significance (with $\alpha = 5\%$) are given in Table 8. The sum $a + b \neq +c$ is also given: it represents the mean score increase between identical chromatographic systems and systems that differ with respect to all the three factors (pH, stationary phase and organic modifier).

Concerning orthogonality in general, Table 8 shows that whatever the orthogonality criterion, any difference in pH, stationary phase or organic modifier leads to an increase of orthogonality (a, $b=$, $b \neq$ and c are all positive). Furthermore, for all of them, $a > b \neq > c$, *i.e.* orthogonality benefits mostly from a difference in pH, then from different stationary phases and last from different organic modifiers.

Among the criteria not requiring space discretization, while Pearson's and Spearman's coefficients are not sensitive to the difference in organic modifier, Kendall's coefficient shows the greatest sensitivity to all factors, and the largest score difference between couples of identical systems, and couples differing in everything (0.70). This is in accordance with its ability to detect nonlinear behaviors and its smaller sensitivity to extreme values.

All criteria requiring space discretization are more sensitive to all factors with a 10×10 discretization (the total score increase is larger, and the sensitivity to the difference in organic modifier also). We did not attempt to find the best discretization step (which is not necessarily the same for all of them), this only shows that the necessity to discretize could be a drawback of these criteria. With the best resolution, they all consider the effect of the difference of organic modifier significant, and Pearson's χ^2 statistics d^2 is almost as good as that of Kendall's coefficient. The fact that Pearson's χ^2 statistics performs better than the likelihood-ratio χ^2 statistics is not surprising: it is well known that the convergence to χ^2 is quicker for d^2 than g^2, especially when some expected counts are below 5 [62].

Table 8
ANOVA parameter values and associated frequencies of significance with a type I error risk $\alpha = 5\%$. The values a, $b=$, $b \neq$ and c are obtained as means over the $n = 63$ LOO values, the standard deviations being given in parenthesis. The frequencies of significance are the frequencies of rejection of the null hypotheses obtained on the $n = 63$ LOO sets.

	Criterion	f_a (%)	$f_{b=}$ (%)	$f_{b\neq}$ (%)	f_c (%)	a	$b=$	$b \neq$	c	$a+b \neq +c$
	r_P	100.0	100.0	98.4	0.0	0.45 (0.005)	0.37 (0.005)	0.12 (0.004)	0.03 (0.001)	0.60 (0.006)
	r_S	100.0	100.0	38.1	0.0	0.54 (0.007)	0.34 (0.007)	0.09 (0.006)	0.03 (0.002)	0.66 (0.006)
	r_K	100.0	100.0	100.0	100.0	0.52 (0.006)	0.32 (0.006)	0.13 (0.006)	0.06 (0.003)	0.70 (0.005)
	conf	100.0	100.0	96.8	1.6	0.51 (0.007)	0.35 (0.006)	0.11 (0.009)	0.04 (0.002)	0.66 (0.005)
	%cov	100.0	100.0	3.2	0.0	0.40 (0.009)	0.32 (0.006)	0.09 (0.006)	0.03 (0.002)	0.52 (0.007)
5×5	d^2	100.0	100.0	100.0	100.0	0.39 (0.009)	0.35 (0.005)	0.12 (0.006)	0.06 (0.003)	0.57 (0.007)
	g^2	100.0	100.0	95.2	31.7	0.31 (0.007)	0.28 (0.004)	0.11 (0.004)	0.05 (0.002)	0.47 (0.006)
	%cov	100.0	100.0	100.0	100.0	0.40 (0.009)	0.28 (0.005)	0.11 (0.005)	0.07 (0.003)	0.58 (0.007)
10×10	d^2	100.0	100.0	100.0	100.0	0.35 (0.008)	0.35 (0.007)	0.22 (0.009)	0.07 (0.003)	0.65 (0.011)
	g^2	100.0	100.0	100.0	100.0	0.33 (0.006)	0.27 (0.003)	0.14 (0.004)	0.06 (0.002)	0.54 (0.006)

Table 9
ANOVA parameter values and associated frequencies of significance for the two χ^2 statistics, the choice of the bin shape being optimized.

	Criterion	f_a	$f_{b=}$	$f_{b\neq}$	f_c	a	$b=$	$b\neq$	c	$a+b\neq+c$
5×5	d^2	100.0	100.0	100.0	100.0	0.46 (0.007)	0.31 (0.005)	0.18 (0.006)	0.06 (0.003)	0.70 (0.005)
	g^2	100.0	100.0	100.0	100.0	0.47 (0.007)	0.31 (0.005)	0.17 (0.006)	0.06 (0.003)	0.70 (0.005)
10×10	d^2	100.0	100.0	100.0	100.0	0.43 (0.009)	0.33 (0.005)	0.19 (0.006)	0.10 (0.003)	0.72 (0.006)
	g^2	100.0	100.0	100.0	100.0	0.45 (0.011)	0.33 (0.005)	0.18 (0.008)	0.10 (0.004)	0.73 (0.006)

This gave us the idea to choose the rectangular bins so as to maximize the minimal expected count numbers, (*i.e.* by having all the n_{i+} and the n_{+j} approximately equal). The results are shown in Table 9.

As expected, the results obtained for d^2 and g^2 are now almost identical. Furthermore, whatever the resolution, they are now the most sensitive criteria with respect to all differences. Thus, to conclude about the quality of the criteria for evaluating orthogonality, the χ^2 independence statistics perform very well and are not too sensitive to the number of bins provided their number and limits are chosen so as to maximize the expected counts. However, their disadvantage with respect to Kendall's correlation coefficient is to require not only discretization, but an optimized one.

4.2. Ranking the factors according to their effect on orthogonality

As demonstrated above, the difference in pH of the mobile phase and the difference of the type of stationary phase are the two parameters with the largest impact on the selectivity in RPLC gradient elution and are the primary parameters to act upon in order to design a RP × RP system. These results are in good agreement with previously published results [46,47,66]. In addition, although its impact is less significant, the difference of organic modifier type was also shown to have an influence.

The impact of the three parameters is illustrated in Fig. 2. For simplicity, we use Kendall's correlation coefficient for orthogonality quantification. We could have taken one of the χ^2 measures as well, but Kendall's coefficient values do not need space discretization and are of more straightforward interpretation: the closer to zero, the more orthogonal the corresponding couple of systems.

To illustrate these results, an example is taken where at the beginning the three factors influencing the orthogonality are the same, *i.e.* $r = 1$ (theoretically). Then by changing the values of the factors, we obtain different values of r_K (from r_{Ka} to r_{Kf}) as illustrated in Fig. 2(a–f), where:

(a) Illustrates the gain in orthogonality due solely to the use of different pH values (r_K decreases from 1 to 0.62).
(b) Reflects the gain in orthogonality due solely to the use of different stationary phases (r_K decreases from 1 to 0.77): it is less than the previous one.
(c) Reflects the gain due to the use of different organic modifiers only (r_K decreases from 1 to 0.87): it is even less than the previous one.
(d) Reflects the gain due to the use of different pH and stationary phases (r_K decreases to 0.54): orthogonality is improved compared with the only change pH.
(e) Demonstrates the gain due to the use of different pH, organic modifiers and stationary phases (r_K decreases to 0.51).
(f) Reflects the gain due to the use of different pH and organic modifiers (0.6): we verify that orthogonality is higher than when only changing pH ($r_{Ka} = 0.62$), but lower than when different pH and stationary phases are used ($r_{Kd} = 0.54$).

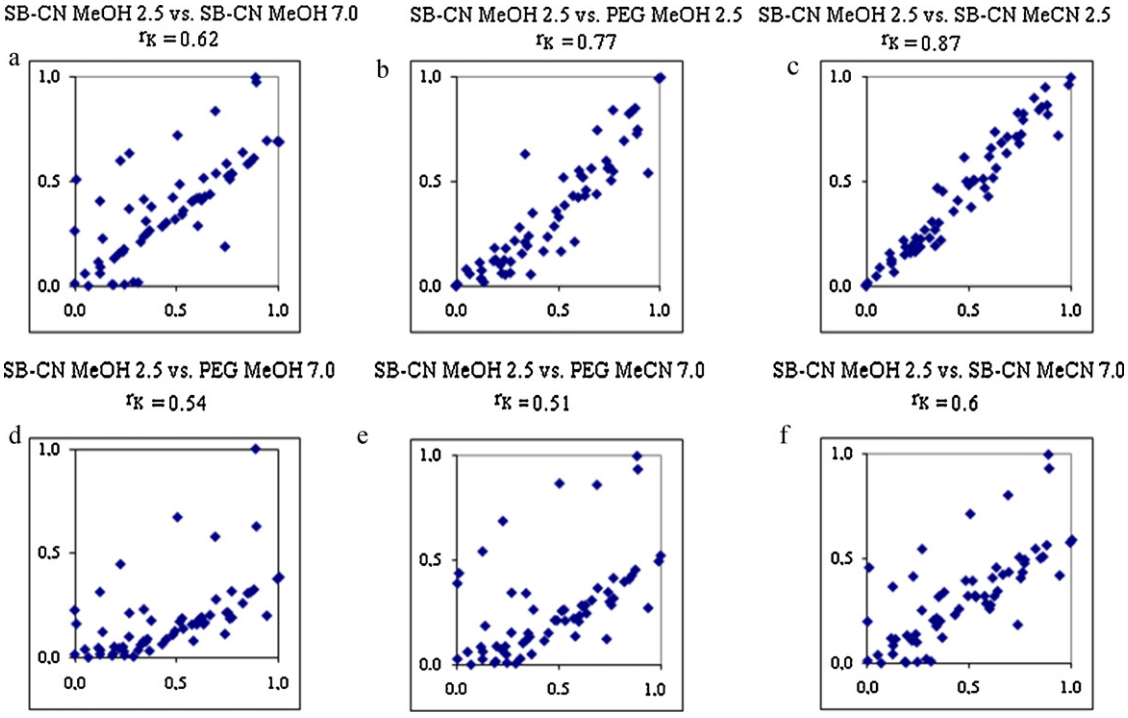

Fig. 2. 2D plots illustrate the effect of changing the stationary phase, the mobile phase pH value and the organic modifier on the orthogonality. The *x*-axis represents the scaled retention time of the solutes on SB-CN phase with MeOH at pH = 2.5, and the *y*-axis represents the scaled retention time of the solutes in a second chromatograhic system.

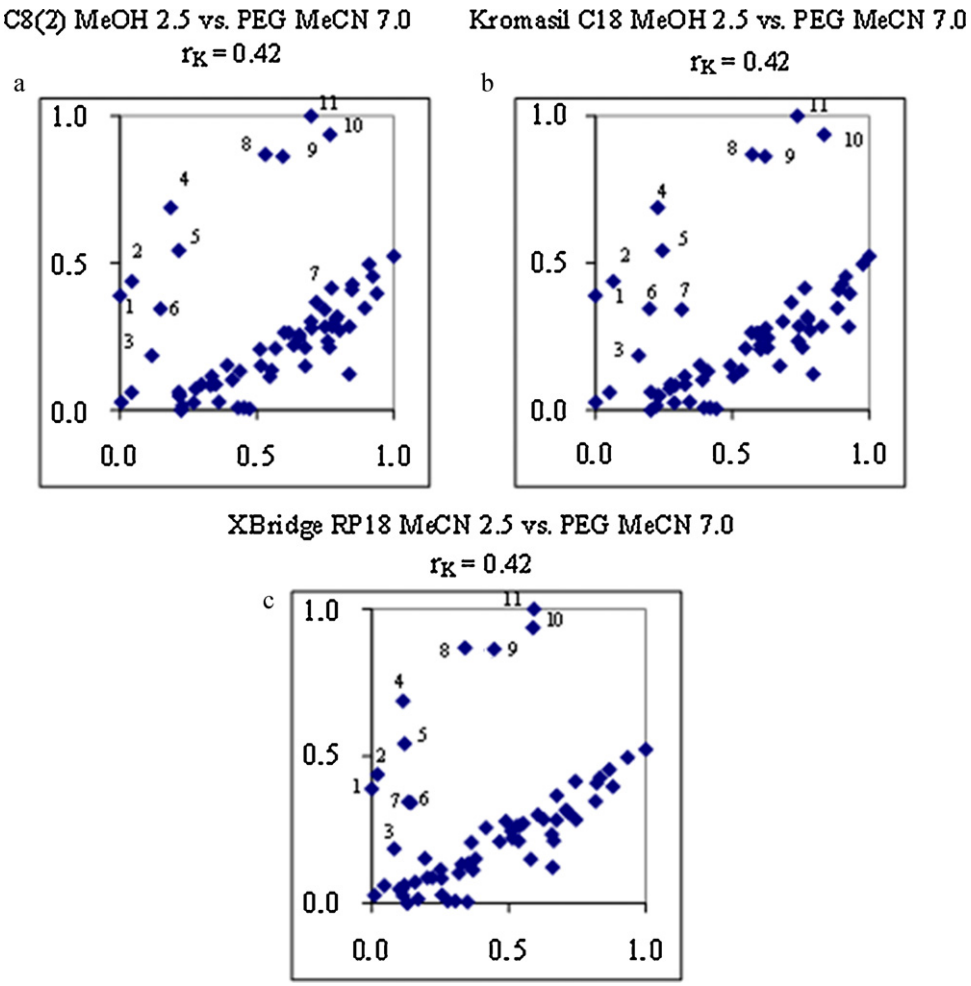

Fig. 3. 2D plots for the most orthogonal couples (r_K = 0.42) among the 496 couples. The indexed points are: (1) nicotine, (2) benzylamine, (3) vancomycin, (4) strychnine, (5) atropine, (6) (+)-tubocurarine, (7) 1,2-phenylendiamine, (8) imipramine, (9) loperamide, (10) amiodarone and (11) clofazimine.

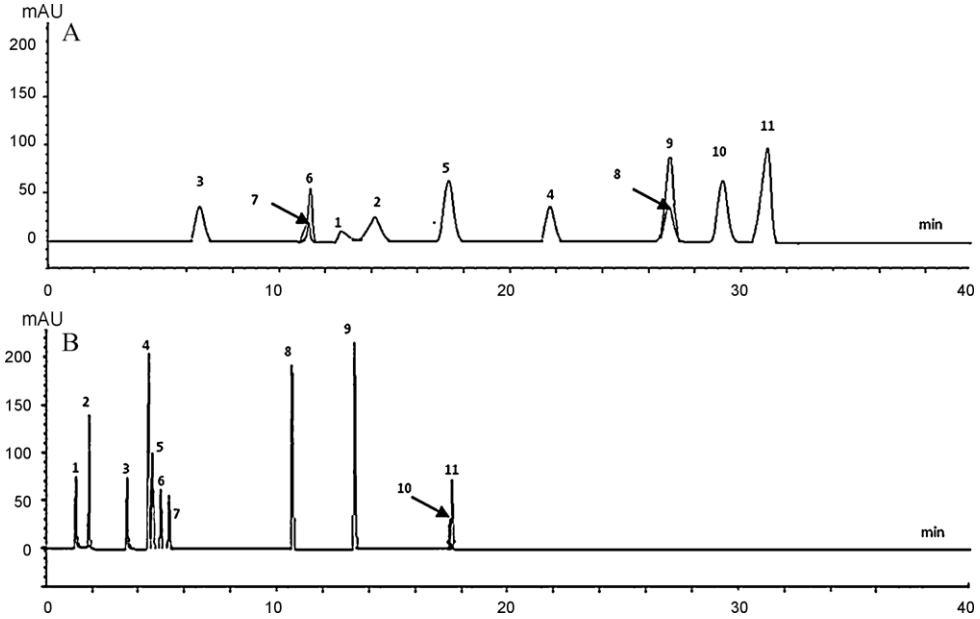

Fig. 4. Chromatograms for 11 amino basic compounds: (1) nicotine, (2) benzylamine, (3) vancomycin, (4) strychnine, (5) atropine, (6) (+)-tubocurarine, (7) 1,2-phenylendiamine, (8) imipramine, (9) loperamide, (10) amiodarone and (11) clofazimine, for the most orthogonal couples of systems, (A) Polyethyleneglycol (PEG) with MeCN at pH 7.0 and (B) XBridge RP18 with MeCN at pH 2.5. Remark: (HS PEG phase in this case showed a decrease in the efficiency due to the large number of injections), the apparent efficiency for Toluene at the beginning of use was equal to 11,069 theoretical plates, while it was only 2171 at the end.

Table 10
Illustration of the difference of selectivity for basic compounds between the most orthogonal couples of systems: Polyethyleneglycol (PEG) with MeCN at pH 7.0 versus: [XBridge RP 18 with MeCN at pH 2.5], [C8(2) with MeOH at pH 2.5], and [Kromasil C18 with MeOH at pH 2.5].

XBridge, pH 2.5, MeCN

No.	Compound	Retention-time/min
1.	Nicotine	1.3
2.	Benzylamine	1.9
3.	Vancomycin	3.6
4.	Strychnine	4.5
5.	Atropine	4.6
6.	(+)-Tubocurarine	5.0
7.	1,2- Phenyldiamine	5.3
8.	Imipramine	10.7
9.	Loperamide	13.4
10.	Amiodarone	17.5
11.	Clofazimine	17.6

PEG, pH 7.0, MeCN

No.	Compound	Retention-time/min
3.	Vancomycin	6.8
7.	1,2- Phenyldiamine	11.5
6.	(+)-Tubocurarine	11.6
1.	Nicotine	12.9
2.	Benzylamine	14.4
5.	Atropine	17.5
4.	Strychnine	21.8
9.	Loperamide	27.0
8.	Imipramine	27.2
10.	Amiodarone	29.3
11.	Clofazimine	31.2

C8(2), pH 2.5, MeOH

No.	Compound	Retention-time/min
1.	Nicotine	1.6
2.	Benzylamine	2.9
3.	Vancomycin	5.2
6.	(+)-Tubocurarine	6.1
4.	Strychnine	7.2
5.	Atropine	8.2
8.	Imipramine	17.8
9.	Loperamide	19.7
7.	1,2- Phenyldiamine	24.4
11.	Clofazimine	22.9
10.	Amiodarone	25.0

PEG, pH 7.0, MeCN

No.	Compound	Retention-time/min
3	Vancomycin	6.8
7.	1,2- Phenyldiamine	11.5
6.	(+)-Tubocurarine	11.6
1.	Nicotine	12.9
2.	Benzylamine	14.4
5.	Atropine	17.5
4	Strychnine	21.8
9.	Loperamide	27.0
8.	Imipramine	27.2
10.	Amiodarone	29.3
11.	Clofazimine	31.2

Kromasil C18, pH 2.5, MeOH

No.	Compound	Retention-time/min
1.	Nicotine	1.6
2.	Benzylamine	3.8
3.	Vancomycin	6.8
6.	(+)-Tubocurarine	8.1
4.	Strychnine	9.1
5.	Atropine	9.6
7.	1,2- Phenyldiamine	11.9
8.	Imipramine	20.3
9.	Loperamide	21.9
11.	Clofazimine	25.7
10.	Amiodarone	28.9

PEG, pH 7.0, MeCN

No.	Compound	Retention-time/min
3.	Vancomycin	6.8
7.	1,2- Phenyldiamine	11,5
6.	(+)-Tubocurarine	11,6
1.	Nicotine	12,9
2.	Benzylamine	14,4
5.	Atropine	17,5
4.	Strychnine	21,8
9.	Loperamide	27,0
8.	Imipramine	27,2
10.	Amiodarone	29,3
11.	Clofazimine	31,2

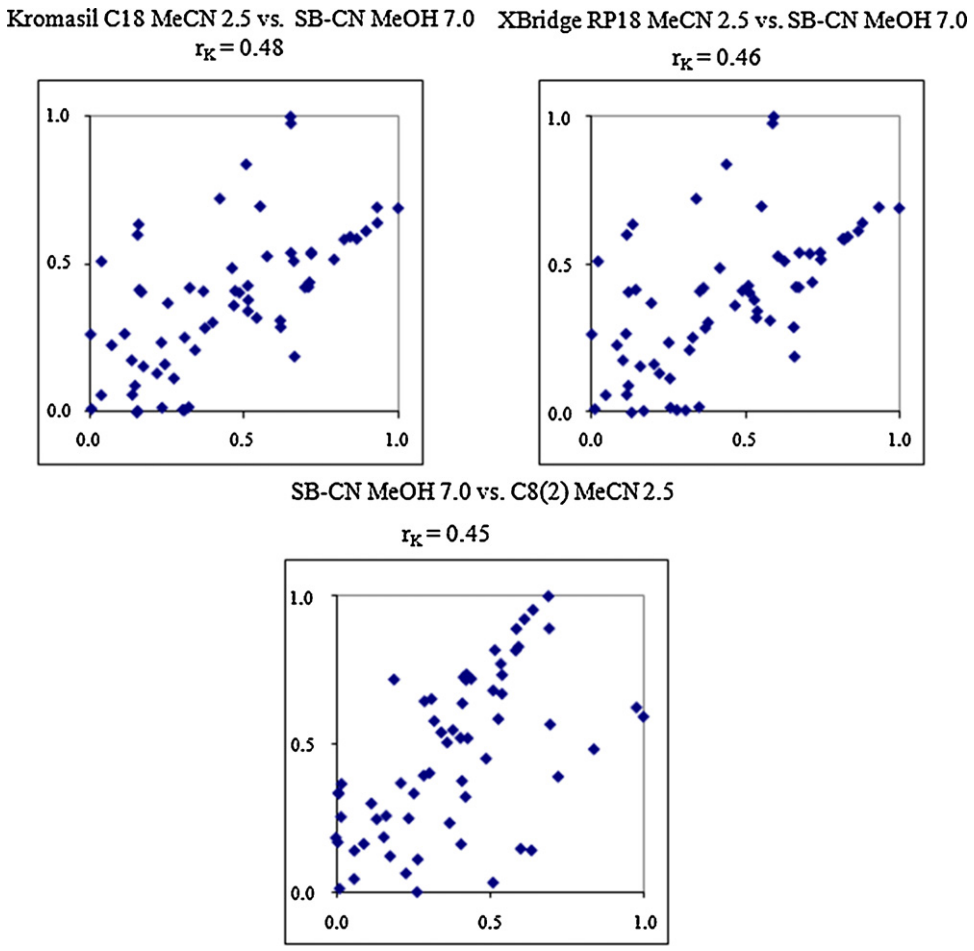

Fig. 5. 2D plots of other very orthogonal couples among the 496 couples: Zorbax SB-CN versus Kromasil C18, XBridge RP18 and Luna C8(2) phases.

4.3. Most orthogonal systems

In the present study, among the 496 couples of systems, couples involving PEG stationary phase always show a high degree of orthogonality, especially together with C18 and C8 stationary phases and when using different pH values as illustrated in Fig. 3.

As illustrated in Table 10 and Fig. 3(a–b), the highest degree of orthogonality obtained in this study was for two systems having both different stationary phases, different pH values of the mobile phase and different organic modifiers. Meanwhile, Fig. 3(c) shows one of the highest dissimilar systems when changing only pH and stationary phase (see Fig. 4 for the chromatograms).

In this study, the PEG column appeared to be a good candidate to build a reverse phase orthogonal system in combination with C8 and C18 phases. It provides separations very different from the C18 and C8 as illustrated in Figs. 3 and 4. This difference in behavior can be related to the specific structure of this phase as given in Fig. 1, it has ether groups that can develop original polar interactions with the analytes. In addition, PEG presents a lower hydrophobicity than C18, which explains a weaker retention [67] for compounds mainly hydrophobic. These reasons describe the effectiveness to use PEG as a stationary phase coupled with C8 and C18 phases to provide a high degree of orthogonality.

Our results about PEG are in good agreement with other mentioned in the literature [68]; they showed that PEG phase provides complementary selectivity to silica based C18. Jandera et al. [69] described the effectiveness to use PEG for separation phenolic and flavons antioxidants.

PEG shows a performance similar to the CN column, but with higher differences among the less polar compounds, which makes CN phase useful for the analysis of some groups of compounds. Therefore, regarding the 496 couples of systems, we found that Zorbax SB-CN is the second stationary phase, after PEG, that can be used to provide a high degree of orthogonality when coupled with C18 and C8 phases as illustrated in Fig. 5.

In agreement with our results, pronounced differences in retention and selectivity have been reported for cyano versus alkyl-silica columns [70,71].

5. Conclusion

The orthogonality of 496 couples of systems was evaluated and ranked based on different criteria: the three classical correlation coefficients (Pearson's, Spearman's and Kendall's), two geometric criteria characterizing the coverage of the 2D separation space, Slonecker's information similarity and two χ^2 statistics of independence. Kendall's coefficient showed the greatest sensitivity to all factors, and the largest score difference between couples of identical systems, and couples differing in all factors. So did the two chi-square statistics (and hence Slonecker's informational similarity which was shown to be equivalent to the likelihood ratio χ^2 statistics), but only if the discretization of the separation space was performed appropriately. Therefore, Kendall's correlation coefficient appeared as both appropriate and user friendly to measure the degree of orthogonality between chromatographic systems. Regarding the factors influencing orthogonality, we concluded that,

for the set of molecular probes, stationary and mobile phases used:

- Unsurprisingly, the greatest orthogonality changes were obtained by changing all the parameters together; different column types, different pH values and different organic modifiers.
- The strongest orthogonality changes were caused by pH changes.
- Strong orthogonality differences were observed when changing the stationary phases but less than when changing the pH values.
- Although the nature of the organic modifier has a small impact, it is not negligible.

However, it is expected that, for particular sets of probes (for example restricted to neutral compounds only or conversely to ionic ones), the hierarchy of the factors influencing orthogonality may be different. It might also have an impact on the sensitivity to orthogonality of some of the criteria. Nevertheless, the proposed strategy, *i.e.* the quantitative analysis through ANOVA of the orthogonality of couples of chromatographic systems offers precisely the means to perform an efficient and extensive study of the effect of the probe set nature.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2011.03.031.

References

[1] S. Eeltink, S. Dolman, R. Swart, M. Ursem, P.J. Schoenmakers, J. Chromatogr. A 1216 (2009) 7368.
[2] C. Song, M. Ye, G. Han, X. Jiang, F. Wang, Z. Yu, R. Chen, H. Zou, Anal. Chem. 82 (2009) 53.
[3] I. François, D. Cabooter, K. Sandra, F. Lynen, G. Desmet, P. Sandra, J. Sep. Sci. 32 (2009) 1137.
[4] D. Mao, H. Van De Weghe, L. Diels, N. De Brucker, R. Lookman, G. Vanermen, J. Chromatogr. A 1179 (2008) 33.
[5] L. Vogt, T. Gröger, R. Zimmermann, J. Chromatogr. A 1150 (2007) 2.
[6] T. Dutriez, M. Courtiade, D. Thiébaut, H. Dulot, F. Bertoncini, J. Vial, M.-C. Hennion, J. Chromatogr. A 1216 (2009) 2905.
[7] A.E. Sinha, K.J. Johnson, B.J. Prazen, S.V. Lucas, C.G. Fraga, R.E. Synovec, J. Chromatogr. A 983 (2003) 195.
[8] R.D. Patel, S.T. Balke, J. Polym. Sci. B Polym. Lett. 18 (1980) 453.
[9] K. Le Mapihan, J. Vial, A. Jardy, Chromatographia 57 (2003) S163.
[10] K. Le Mapihan, J. Vial, A. Jardy, J. Chromatogr. A 1030 (2004) 135.
[11] K. Le Mapihan, J. Vial, A. Jardy, J. Chromatogr. A 1088 (2005) 16.
[12] N. Tanaka, H. Goodell, B.L. Karger, J. Chromatogr. 158 (1978) 233.
[13] T.H. Dzido, T.E. Kossowski, D. Matosiuk, J. Chromatogr. A 947 (2002) 167.
[14] G.F.J. Barbosa, I. Marques, V. Sanz-Nebot, I. Toro, Anal. Chim. Acta 351 (1997) 397.
[15] I. Marqués, G. Fonrodona, S. Butí, J. Barbosa, Trends Anal. Chem. 18 (1999) 472.
[16] J. Barbosa, I. Toro, R. Bergés, V. Sanz-Nebot, J. Chromatogr. A 915 (2001) 85.
[17] J. Barbosa, G. Fonrodona, I. Marqués, S. Butí, I. Toro, Trends Anal. Chem. 16 (1997) 104.
[18] J. Barbosa, D. Barrón, J.L. Beltrán, S. Butí, Talanta 45 (1998) 817.
[19] U.D. Neue, C.H. Phoebe, K. Tran, Y.-F. Cheng, Z. Lu, J. Chromatogr. A 925 (2001) 49.
[20] U.D. Neue, B.A. Alden, T.H. Walter, J. Chromatogr. A 849 (1999) 101.
[21] U.D. Neue, E. Serowik, P. Iraneta, B.A. Alden, T.H. Walter, J. Chromatogr. A 849 (1999) 87.
[22] R.J.M. Vervoort, E. Ruyter, A.J.J. Debets, H.A. Claessens, C.A. Cramers, G.J. de Jong, J. Chromatogr. A 931 (2001) 67.
[23] R.J.M. Vervoort, E. Ruyter, A.J.J. Debets, H.A. Claessens, C.A. Cramers, G.J. de Jong, J. Chromatogr. A 964 (2002) 67.
[24] H. Engelhardt, H. Löw, W. Götzinger, J. Chromatogr. 544 (1991) 371.
[25] M. Jungheim, H. Engelhardt, Chromatographia 29 (1990) 59.
[26] H. Engelhardt, T. Lobert, Anal. Chem. 71 (1999) 1885.
[27] H.A. Claessens, Trends Anal. Chem. 20 (2001) 563.
[28] H.A. Claessens, M.A. van Straten, C.A. Cramers, M. Jezierska, B. Buszewski, J. Chromatogr. A 826 (1998) 135.
[29] M. Turowski, T. Morimoto, K. Kimata, H. Monde, T. Ikegami, K. Hosoya, N. Tanaka, J. Chromatogr. A 911 (2001) 177.
[30] N. Tanaka, K. Kimata, K. Hosoya, H. Miyanishi, T. Araki, J. Chromatogr. A 656 (1993) 265.
[31] M.R. Euerby, P. Petersson, LC–GC Eur. 13 (2000) 665.
[32] W.E. May, N.M. Trahey, Certificate of Analysis, Standard Reference Material 870, NIST, Gaitherburgs, MD, 2000.
[33] L.C. Sander, S.A. Wise, Certificate of Analysis, Standard Reference Material 869a, NIST, Gaitherburgs, MD, 1998.
[34] M.J. Walters, J. Assoc. Off. Anal. Chem. 70 (1987) 465.
[35] M.J. Gray, G.R. Dennis, P.J. Slonecker, R.A. Shalliker, J. Chromatogr. A 1015 (2003) 89.
[36] Z. Liu, D.G. Patterson, M.L. Lee, Anal. Chem. 67 (1995) 3840.
[37] P.J. Slonecker, X. Li, T.H. Ridgway, J.G. Dorsey, Anal. Chem. 68 (1996) 682.
[38] M. Gray, G.R. Dennis, P. Wormell, R. Andrew Shalliker, P. Slonecker, J. Chromatogr. A 975 (2002) 285.
[39] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Anal. Chem. 77 (2005) 6426.
[40] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, J. Sep. Sci. 28 (2005) 1694.
[41] F. Cacciola, P. Jandera, E. Blahová, L. Mondello, J. Sep. Sci. 29 (2006) 2500.
[42] P.J. Boersema, N. Divecha, A.J.R. Heck, S. Mohammed, J. Proteome Res. 6 (2007) 937.
[43] T.J. Whelan, M.J. Gray, P.J. Slonecker, R.A. Shalliker, M.A. Wilson, J. Chromatogr. A 1097 (2005) 148.
[44] E.K. Jeong, H.J. Cha, Y.S.Y.W. Ha, I.J.H. Kim, Y.C. Na, J. Chromatogr. A 1217 (2010) 4375.
[45] E. Van Gyseghem, M. Jimidar, R. Sneyers, D. Redlich, E. Verhoeven, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1074 (2005) 117.
[46] E. Van Gyseghem, S. Van Hemelryck, M. Daszykowski, F. Questier, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 988 (2003) 77.
[47] E. Van Gyseghem, I. Crosiers, S. Gourvénec, D.L. Massart, Y. Vander Heyden, J. Chromatogr. A 1026 (2004) 117.
[48] S. Zhu, J. Chromatogr. A 1216 (2009) 3312.
[49] P. Forlay-Frick, E. Van Gyseghem, K. Héberger, Y. Vander Heyden, Anal. Chim. Acta 539 (2005) 1.
[50] R. Rajko, K. Héberger, Chemometr. Intell. Lab. 57 (2001) 1.
[51] W. Fan, Y. Zhang, P.W. Carr, S.C. Rutan, M. Dumarey, A.P. Schellinger, W. Pritts, J. Chromatogr. A 1216 (2009) 6587.
[52] J.W. Dolan, L.R. Snyder, J. Chromatogr. A 1216 (2009) 3467.
[53] http://www.chemaxon.com/marvin/sketch/index.jsp.
[54] http://www.syrres.com/what-we-do/databaseforms.aspx?id=386.
[55] E. Beetge, J. Plessis, D.J. Müller, C. Goosen, F.J.v. Rensburg, Int. J. Pharm. 193 (2000) 261.
[56] H. Wan, A. Holmén, M. Någård, W. Lindberg, J. Chromatogr. A 979 (2002) 369.
[57] L.A. Ciolino, J.A. Turner, H.A. McCauley, A.W. Smallwood, T.Y. Yi, J. Chromatogr. A 852 (1999) 451.
[58] K. Le Mapihan, J. Vial, A. Jardy, J. Chromatogr. A 1061 (2004) 149.
[59] D.V. McCalley, J. Chromatogr. A 738 (1996) 169.
[60] D.V. McCalley, J. Chromatogr. A 769 (1997) 169.
[61] D.V. McCalley, J. Chromatogr. A 793 (1998) 31.
[62] A. Agresti, Categorical Data Analysis, Wiley, 2002.
[63] H.E. Daniels, Biometrika 33 (1944) 129.
[64] G. Saporta, Probabilités: analyse des données et statistiques. Technip, 2ème éd., 2006.
[65] E. Van Gyseghem, B. Dejaegher, R. Put, P. Forlay-Frick, A. Elkihel, M. Daszykowski, K. Héberger, D.L. Massart, Y.V. Heyden, J. Pharm. Biomed. Anal. 41 (2006) 141.
[66] W. Steuer, I. Grant, F. Erni, J. Chromatogr. 507 (1990) 125.
[67] P. Jandera, P. Cesla, T. Hájek, G. Vohralík, K. Vynuchalová, J. Fischer, J. Chromatogr. A 1189 (2008) 207.
[68] E. Blahová, P. Jandera, F. Cacciola, L. Mondello, J. Sep. Sci. 29 (2006) 555.
[69] F. Cacciola, P. Jandera, Z. Hajdú, P. Cesla, L. Mondello, J. Chromatogr. A 1149 (2007) 73.
[70] X. Chen, L. Kong, X. Su, H. Fu, J. Ni, R. Zhao, H. Zou, J. Chromatogr. A 1040 (2004) 169.
[71] F. Bedani, W. Kok, H.-G. Janssen, Anal. Chim. Acta 654 (2009) 77.