Talanta 119 (2014) 524–532

Contents lists available at ScienceDirect

Talanta

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

Fuzzy clustering evaluation of the discrimination power of UV–Vis and (±) ESI-MS detection system in individual or coupled RPLC for characterization of *Ginkgo Biloba* standardized extracts

Andrei Medvedovici [a,b], Florin Albu [a,b], Rodica Domnica Naşcu-Briciu [c], Costel Sârbu [c,*]

[a] University of Bucharest, Faculty of Chemistry, Department of Analytical Chemistry, Panduri Ave., No. 90, Bucharest 050663, Romania
[b] Bioanalytical Laboratory, SC Labormed Pharma SA, No. 44B, Th. Pallady Blvd., Bucharest 032266, Romania
[c] Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Department of Chemistry, Arany Janos Street, No. 11, Cluj-Napoca 400028, Romania

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form
6 November 2013
Accepted 9 November 2013
Available online 27 November 2013

Keywords:
Ginkgo Biloba standardized extracts
Fuzzy hierarchical clustering
Principal components analysis
Cluster analysis
HPLC/UV/(±)ESI-MS

ABSTRACT

Aim: Discrimination power evaluation of UV–Vis and (±) electrospray ionization/mass spectrometric techniques, (ESI-MS) individually considered or coupled as detectors to reversed phase liquid chromatography (RPLC) in the characterization of *Ginkgo Biloba* standardized extracts, is used in herbal medicines and/or dietary supplements with the help of Fuzzy hierarchical clustering (FHC).
Experimental: Seventeen batches of *Ginkgo Biloba* commercially available standardized extracts from seven manufacturers were measured during experiments. All extracts were within the criteria of the official monograph dedicated to dried refined and quantified *Ginkgo* extracts, in the European Pharmacopoeia. UV–Vis and (±) ESI-MS spectra of the bulk standardized extracts in methanol were acquired. Additionally, an RPLC separation based on a simple gradient elution profile was applied to the standardized extracts. Detection was made through monitoring UV absorption at 220 nm wavelength or the total ion current (TIC) produced through (±) ESI-MS analysis. FHC was applied to raw, centered and scaled data sets, for evaluating the discrimination power of the method with respect to the origins of the extracts and to the batch to batch variability.
Results: The discrimination power increases with the increase of the intrinsic selectivity of the spectral technique being used: UV–Vis < MS(−) < MS(+), but it is strongly sensitive to chemometric transformation of data. Comparison with cluster analysis (CA) and principal components analysis (PCA) indicates that the FHC algorithm produces better classification. However, PCA and CA may be successfully applied to discriminate between the manufacturing sources of the standardized extracts, and at some extent, to monitor the inter-batch variability. Although the chromatographic dimension sensibly contributes to the discrimination power, spectral MS data may be used as the lone powerful holistic alternative in characterization of standardized *Ginkgo Biloba* extracts.

1. Introduction

Nowadays fingerprinting and pattern recognition algorithms represent valuable tools for the characterization of the complex chemical mixtures of natural origins, with promising results in various application fields such as food [1,2], beverages [3,4], agriculture [5,6], chemotaxonomy [7,8], herbal medicines [9], dietary supplements [10,11], metabolic profiling [12–14], environmental [15,16] and standardization [17]. The European Medicines Agency [18] recommends that the appropriate fingerprinting procedures should be based on chromatographic techniques. However, other techniques, such as the spectral ones, may lead to interesting and useful results [19] as the US Food and Drug Administration [20] recommends.

Evolvement of fingerprinting procedures during the last decade has been supported by the application of powerful and advanced chemometric methods, among which principal component analysis (PCA), partial least squares (PLS), cluster analysis (CA), linear discriminant analysis (LDA) and artificial neuronal networks (ANN)

Abbreviations: ANN, artificial neuronal network; a.m.u., atomic mass unit; CA, cluster analysis; DAD, diode array detection; (±)ESI, positive/negative electro spray ion source; FHC, Fuzzy hierarchical clustering; HPLC, high pressure liquid chromatography; IBJ, Backer–Jain Index; LDA, linear discriminant analysis; mAU, milli-absorption unit; MS, mass spectrometry; PCA, principal components analysis; PCN, normalized partition coefficient; PEN, normalized partition entropy; PLC, partial least squares; RAM, random access memory; RPLC, reversed phase liquid chromatography; TIC, total ion current; TOC, total organic carbon; UV-Vis, ultraviolet and visible

* Corresponding author. Tel.: +40 264 593833; fax: +40 264 590818.
E-mail address: csarbu@chem.ubbcluj.ro (C. Sârbu).

0039-9140/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2013.11.035

should be mentioned as the most frequently used [21–25]. These methods generally lead to very efficient classifications, but are highly sensitive to outliers, missing data, and/or inadequate linear correlation between variables due to their poor distribution, as major sources for erroneous conclusions. These disadvantages may be eliminated by using robust techniques, such as Fuzzy hierarchical clustering (FHC). Fuzzy logic is a form of many-valued logic, which deals with reasoning that is approximate rather than fixed and exact. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true (1) and completely false (0) [26]. In classical CA each object must be assigned to exactly one cluster and this is leading to ambiguity and error in cases of outliers or overlapping clusters and affords information loss, while FHC is leading to classes and subclasses representing a collection of ill-defined and not-distinct objects with undefined boundaries in which the transition from membership to non-membership in a subclass of a reference set is gradual rather than abrupt.

The Fuzzy theory is basically a theory of graded concepts. It is an extreme generalization of ordinary set theory and is basically designed to handle the concept of partial truth or fuzziness. It provides an adequate conceptual framework as well as a mathematical tool to model the real world problems which are often vagueness and indistinct [27–29]. Most FHC algorithms are based on objective functions and determine an optimal classification by minimizing them. In objective function based clustering each cluster is usually represented by a cluster prototype, which consists in a cluster center (whose name already indicates its meaning) and maybe some additional information about the size and the shape of the cluster. The affiliation to a specific cluster is given by the degrees of membership, computed as the distance from the data point to the cluster center defined by the fuzzy means. The closer a data point lies to the center of a cluster, the higher is its degree of membership to this cluster. Hence, the problem to divide a data set into clusters can be stated as the task to minimize the distances of the data points to the cluster centers, since, of course, the principal target represents the maximization of membership degrees [30,31].

The leading concepts of the Fuzzy theory may successfully support not only concrete fingerprinting procedures of complex natural mixtures, but also a holistic characterization of complex mixtures, with reduced risks relating to the misinterpretation of the primary data.

Ginkgo Biloba is a medicinal plant frequently used as dried vegetal material or refined extracts in herbal medicines or dietary supplements. The composition of *Ginkgo Biloba* derived materials is complex [32], usually requiring powerful analytical tools for a comprehensive characterization [33–36].

The herein presented approach aims to deliver responses to the following main questions: (i) Do the spectrometric techniques (UV/Vis and $\pm$ MS) have, taken separately, enough potential to produce experimental data with the required discriminating power to assess the origins and batch to batch variability of complex natural samples? (ii) Is the chromatographic dimension essential for the correct assessment of the origins and reproducibility in the production stages of standardized natural extracts? (iii) Which of the chemometric methods (HFC, PCA and CA) is the most suitable one for informational discrimination assessment of the analytical experimental data? (iv) Is chemometric data treatment affecting the discrimination power?

Our starting decision was to focus on standardized extracts and not on dried vegetal materials, considering this approach as a "worst" case application (discrimination is tested on materials meeting strictly specified quantitative criteria).

The opportunity of our approach is based on the increasing success of dietary supplements from natural source, on the medical markets, which are characterized by affordable prices.

This market is, however, less strictly regulated than the one of classic medicines (drugs) and the variability of the raw active materials is naturally higher than for synthetic products. An evident contradiction appears between the need of efficient production with the lowest expenditure and the objective necessity of deep analytical characterization of the complex active raw materials of natural origins and their inherent quality control, involving expensive resources. Holistic approaches as those investigated in the present manuscript may represent straightforward and relatively inexpensive alternatives to the assessment of the composition variability with respect to origins and batch to batch variability of complex mixtures from natural sources. The information produced through application of chemometric methods to spectral or chromatographic data should not be considered as a substitute for the quantitative assays involved in the characterization of standardized materials according to official monographs, but only an easier and rapid way of controlling the source (manufacturer) and batch to batch reproducibility.

2. Materials and methods

2.1. Chemicals

Acetonitrile and methanol were HPLC gradient grade from Merck (Darmstadt, Germany). Formic acid (extra pure grade) from Merck was also used during experiments. Water for chromatography (resistivity of minimum 18.2 MΩ and residual total organic carbon content – TOC – of maximum 30 ng mL^{-1}) was produced within the laboratory by means of a TKA Lab HP 6UV/UF instrument (TKA Instruments as part of Thermo Fischer Scientific, Niederelbert, Germany).

2.2. Samples

Seventeen *Ginkgo Biloba* standardized extracts from six different manufacturers (A–F) were used during experiments. Samples 1–3 are batches produced at 1 year distance by manufacturer A. Samples 4–11 are eight different batches from manufacturer B, produced over 2 years interval (samples 7–11 are consecutive batches). Samples 12 and 13 are batches from manufacturer C, while samples 14 and 15 are consecutive batches from manufacturer D. Samples 16 and 17 are from suppliers E and F. All analyzed batches were placed within their declared shelf life period at the moment of the analysis. All standardized extracts are declared by manufacturers to comply with requirements of the official monograph of European Pharmacopoeia for *Ginkgo Biloba* dry extract, refined and quantified [37].

2.3. Equipment

Experiments were performed with an Agilent 1200 SL series LC/MSD (Agilent Technologies) system consisting of the following modules: degasser (G1379B), binary pump (G1312B), automated injector (G1367C and its corresponding thermostat G1330B), column thermostat (G1316B), diode array detector (G1315C) fitted with a semi-micro 5 μL flow cell (G1314-60011), ESI standard source (G1948B), and triple quadrupole mass spectrometric detector (G2571A). System control, data acquisition and interpretation were made with the Agilent Mass Hunter software version B 04.01 (B4114 Patch 1) incorporating both qualitative and quantitative packages.

2.4. MS parameters

The parameters controlling the ESI ion source were as following: drying gas (N_2); temperature (350 °C); drying gas flow (13 L min^{-1}); pressure of the nebulizing gas (60 psi); capillary voltage (4000 V).

The fragmentor potential was set at 135 V. EMV was 600 V. The resolution of the mass analyzer was set to 0.1 a.m.u. A scan time of 500 ms was used. The MS detection was carried out in the MS2 scan mode. MS spectra were acquired in the 50–1200 a.m.u. interval.

2.5. Sample preparation

Standardized *Ginkgo Biloba* extracts were dissolved in methanol, at a nominal concentration of 1 mg mL^{-1}. No further processing of the resulting solutions was applied.

2.6. Chromatographic separation

A Zorbax Eclipse XDB-C18, 150 mm length, 4.6 mm internal diameter and 3.5 μm particle size (Agilent Technologies, cat. no. 963967-902) fitted with a Guard Cartridge C18, 4 mm × 2 mm (Phenomenex, prod. no. AJO-4286) was used and thermostated at 25 °C. The components of the mobile phase were 0.1% (v/v) formic acid solutions in acetonitrile (solvent A) and water (solvent B). A gradient elution with the following profile was applied (time: min/solvent A: %): 0/10→5/10→65/100→90/100→90.01/10→95/10. The last two stages in the gradient profile correspond to column re-equilibration. The flow rate was 0.8 mL min^{-1}. Injected sample volume was 2.5 μL.

2.7. Experimental design

For the acquisition of UV-Vis and/or (±)MS spectra, 2.5 μL from each standardized extract were directly injected in a stainless steel (S.S.) tube 2 m length × 0.12 mm i.d., also thermostated at 25 °C. The carrier was a mixture 1/1 (v/v) between solvents A and B, at 0.8 mL/min flow rate. Consecutive injections were made at 4 min interval. UV–Vis spectra were acquired in the 190–800 nm spectral range, with a spectral resolution of 1 nm.

Chromatographic experiments consisted in the application of the separation method described under Section 2.6. to the standardized extracts' solutions. The monitoring was at 220 nm for DAD detection with a data acquisition rate of one reading each 0.0066 min. For MS detection, spectra were recorded in the *m*/*z* range 50–1200 a.m.u. with 0.1 a.m.u. spectral resolution, either in (+) or (−) ES ionization. The data acquisition speed used in the total ion current (TIC) working mode was one reading each 0.0097 min.

2.8. Raw data processing

UV-Vis or (±) MS spectra were taken at the apex of the peak that was registered after the transport of the sample through the S. S. tube. Spectra or chromatograms were exported to Excel sheets, as two functionally related numeric strings: wavelength (nm)/absorbance (mAU) for UV-Vis spectra; *m*/*z* value (a.m.u)/intensity (cps) for MS spectra; time (min)/absorbance at 220 nm (mAU) for UV monitored chromatograms; time (min)/relative intensity (%) for (±) MS, TIC monitored chromatograms. UV-Vis spectra were represented by 610 values for each of the numeric strings. The acquisition interval of the MS spectra (from 50 to 1200 a.m.u.) was reduced to the interval 100–1100 a.m.u. Each 10 consecutive data from the mass spectra were averaged, in order to produce 1 a.m.u. spectral resolution. Consequently, numeric strings contain 1000 values. UV monitored chromatograms were considered only in the 5–30 min retention time interval. Five successive acquired data were averaged, in order to produce one reading for a 0.033 min period. Corresponding numeric strings contain 751 values each. MS monitored chromatograms were considered in the 0–40 min retention time interval. Five consecutive data values were averaged in order to produce one reading for a 0.047 min period. Corresponding numeric strings contain 842 values. This additional processing of raw acquired data was necessary in order to limit the length of the strings up to 1000 values. This was truly necessary as long as processing of numeric strings larger than 1000 values through the related chemometric software required very long periods and often determined PC blocking. Parts from the chromatograms after min 30 (for UV detection) and 40 (for MS detection) were discarded as long as the signal to noise ratio (S/N) for peaks appearing in the respective intervals were below 3.

2.9. Data analysis

Matrices used for chemometric analysis were differed according to the analytical system used. Thus, the matrices corresponding to HPLC-MS (±) system were consisted of 17 samples × 842 variables, while for HPLC-UV-Vis the matrix was formed by 17 samples × 751 variables. Moreover, in the case of MS (±) spectra the matrix was formed by 17 samples × 1000 variables, while for UV-Vis spectra the matrix was formed by 17 samples × 610 variables. The chemometric analysis tried to elucidate if FHC is more informative than PCA and CA. Moreover, it tried to indicate which of the analytical systems used is more selective, and with a higher discrimination capacity, as well as to underline how the data transformation is influencing the final results.

The FHC analysis was performed on raw, centered and scaled data. The FHC partitions were obtained through application of a hierarchical algorithm, which is minimizing the distances of the data points to the cluster centers. For each case the quality of partition was evaluated along with the cluster validity indices referred in Table 1. The entire FHC analysis was performed with Sadic 8.3., a personal software package. On the raw matrices presented above, the PCA and CA were performed using Statistica 8.0 software. Software were run on a personal computer (laptop) Sony Vaio equipped with an Intel® core™ – i5 3210 M 2.5 GHz, 8 GB RAM. The computational processes took between 8 and 24 h.

3. Results and discussion

3.1. Premises

In Supplementary material: Part A, some illustrative chromatograms and spectra representing the source of the raw numerical

Table 1
The clustering validity indices values corresponding to FC algorithms.

Index[a]	Raw data						Centered data						Scaled data					
	HPLC-MS (−)	HPLC-MS (+)	HPLC-UV	MS (−)	MS (+)	UV	HPLC-MS (−)	HPLC-MS (+)	HPLC-UV	MS (−)	MS (+)	UV	HPLC-MS (−)	HPLC-MS (+)	HPLC-UV	MS (−)	MS (+)	UV
PCN	0.361	0.392	0.572	0.414	0.461	0.474	0.361	0.392	0.572	0.414	0.461	0.479	0.251	0.208	0.562	0.113	0.291	0.381
PEN	0.555	0.549	0.356	0.490	0.462	0.474	0.555	0.548	0.356	0.489	0.462	0.474	0.706	0.812	0.358	0.917	0.744	0.581
IBJ	0.766	0.772	0.887	0.825	0.819	0.797	0.766	0.772	0.887	0.825	0.819	0.797	0.616	0.457	0.889	0.321	0.515	0.729

[a] PCN, normalized partition coefficient; PEN, normalized partition entropy; IBJ, Backer–Jain Index.

data sets used in FHC, PCA and CA evaluation are provided. Intentionally, samples from different manufacturers and batch numbers have been chosen, to force visual comparison. Although some differences may be noted through naked eye observation, a classification of samples based on the simple visual inspection seems an unrealistic task from the beginning.

It is to note that the gradient profile of the RPLC separation was intentionally kept low (1.5%/min) to increase the separation resolution. A rough comparison between the chromatograms resulting after the use of different detection systems indicates that the number of peaks does not differ substantially (i.e. 48 peaks in (−) MS monitored chromatograms, 49 peaks in (+) MS monitored ones and 56 peaks in UV −220 nm, if considering sample 1). Both positive and negative mass spectra contained 100 peaks. Although positive ionization was favored by the chemistry of analytes in the acidic mobile phase being used, negative ionization also appears, probably due to occurrence of phenolic groups and glycosidated forms of separated analytes. UV detection provides chromatograms exhibiting the highest number of separated compounds. However, due to the intrinsic poor sensitivity of the detection technique, most of the peaks have low intensity. Compounds exhibiting extreme retention characteristics (both reduced and increased retention) were better observed under (+) ESI monitoring. Signals in (+) MS spectra are about one order of magnitude higher than those existing in the (−) MS ones. However, negative ionization allows a better observation of peaks in the 700–1000 m/z interval.

3.2. Fuzzy hierarchical clustering assessment

By performing FHC on the raw, centered and scaled data, the samples were distributed on clusters of different sizes. To assess the clustering performance of FHC algorithms, three validity indices were calculated [38–40]. The values of the validity indices obtained for the fuzzy algorithm applied in this study are presented in Table 1.

By comparing the values in Table 1 and the variation of validity indices shown in Supplementary material: Part B, one can conclude that the raw and centered data are leading to identical results, thereby only raw and scaled data will be further investigated. Moreover, the validity indices are suggesting that the best separation of the investigated samples is offered by the raw data.

The assignment of a sample to a particular cluster is evaluated by means of the membership degree, which varies between 0, when a sample does not belong to a particular cluster, and 1, situation when the sample is considered to be a member of a selected cluster. In other words, each sample belongs to one and only one cluster when a membership degree is 1. In our case, sample distribution within the obtained clusters was more or less in good agreement with samples origin (Supplementary material: Part C).

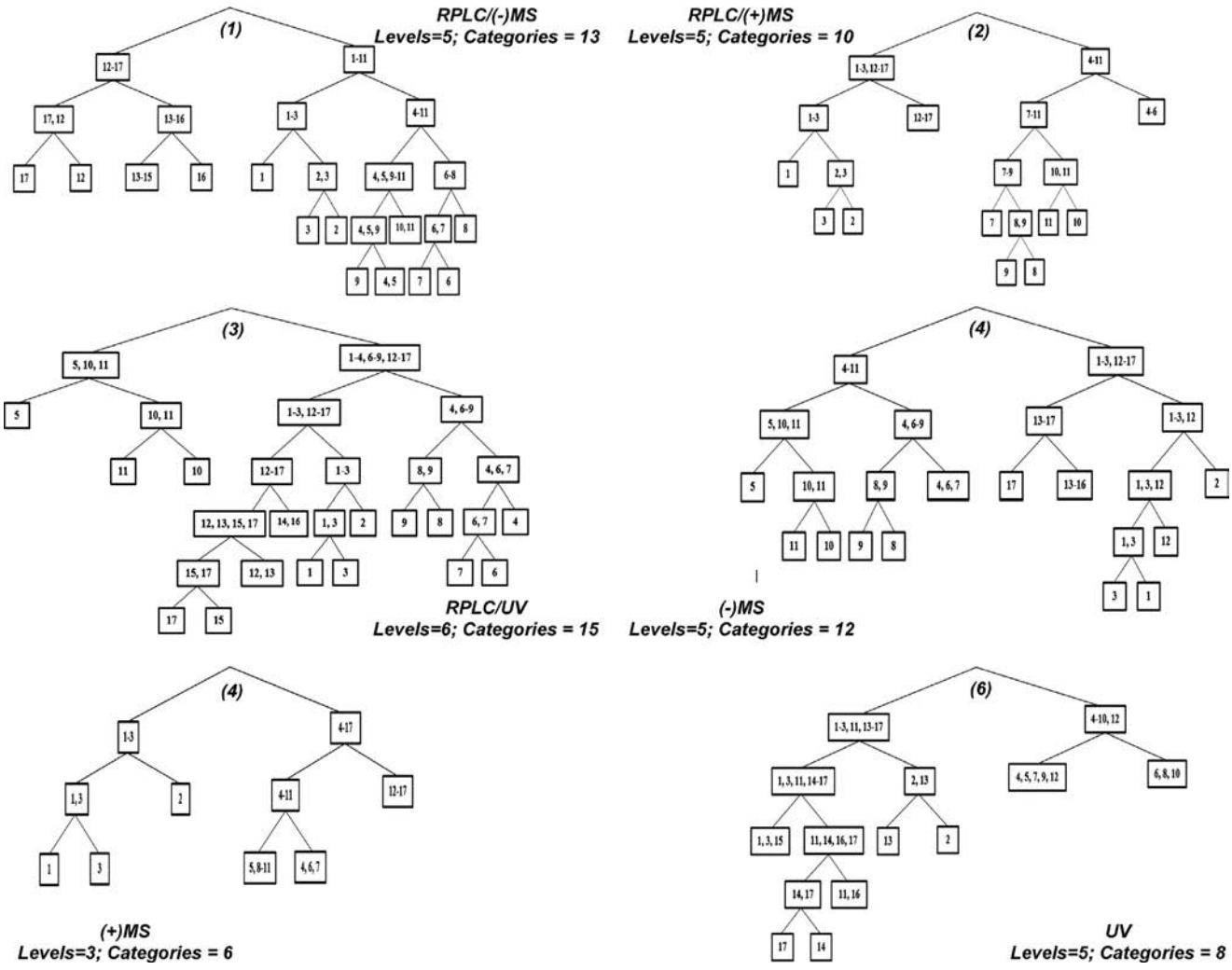

Fig. 1. The hard partition tree corresponding to the fuzzy hierarchical clustering on normal data, for the *Ginkgo Biloba* samples: (1) RPLC/(−)MS, (2) RPLC/(+)MS, (3) RPLC/UV, (4) (−)MS, (5) (+)MS, and (6) UV.

The correlations between various samples may be evaluated by hard partition corresponding to the fuzzy successive partition of the samples, produced by using fuzzy divisive hierarchical clustering. The hard partition is obtained by defuzzification of the fuzzy partition. The samples are assigned to the cluster with the highest membership degree. Figs. 1 and 2 illustrate the assignment of the samples in agreement with the highest membership degree, for raw and scaled data sets.

From the point of view of a non-specialist end-user, the FHC results should discriminate against six different manufacturers

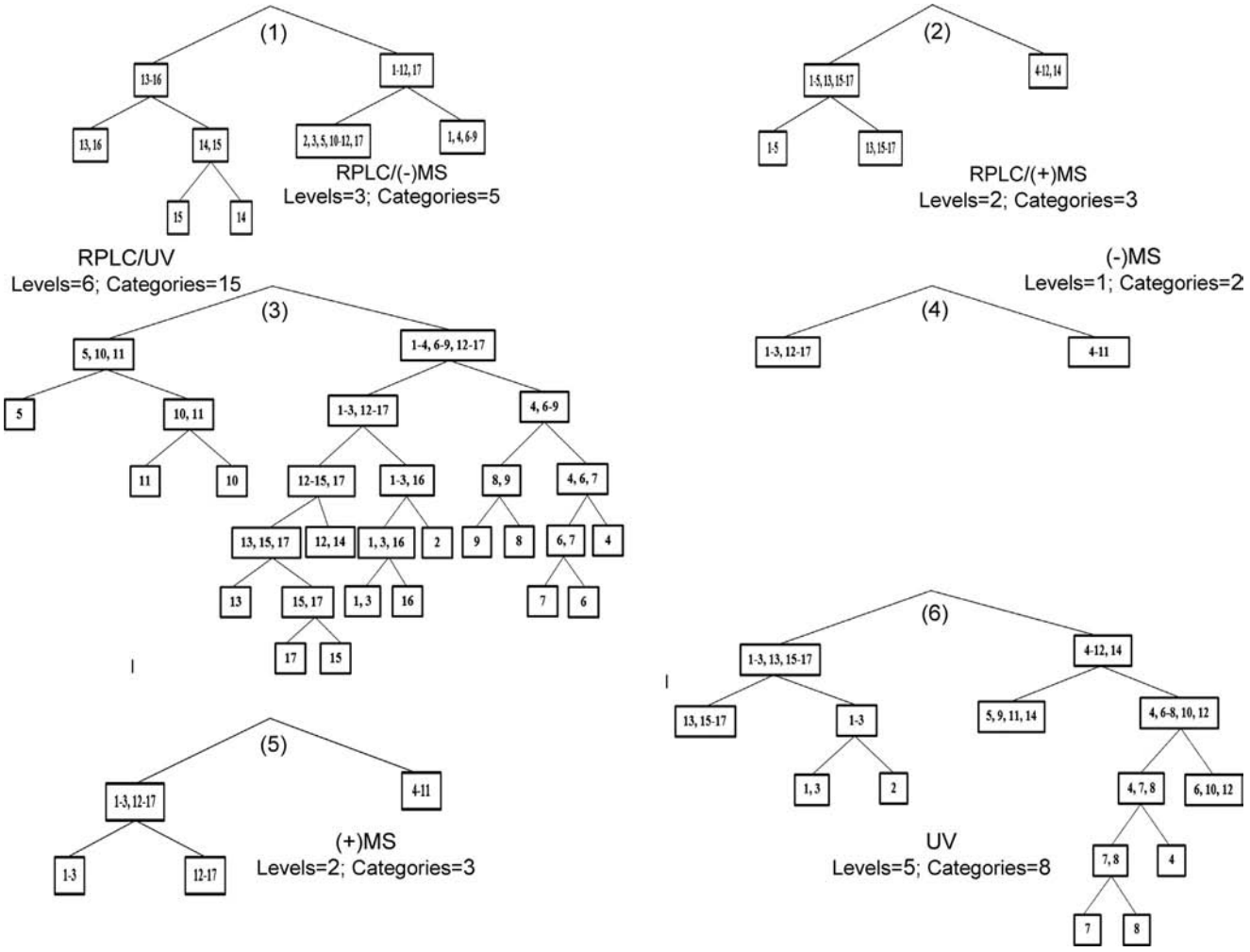

Fig. 2. The hard partition tree corresponding to the fuzzy hierarchical clustering on scaled data, for the *Ginkgo Biloba* samples: (1) RHPLC/(−)MS, (2) RPLC/(+)MS, (3) RPLC/UV, (4) (−)MS, (5) (+)MS, and (6) UV.

Table 2
Hierarchical levels and categories resulting through application of the FC algorithms on raw and scaled data sets, when considering samples or manufacturers. Tentative ranking of the techniques with respect to their ability of classifying according to origin (manufacturer) and batch succession.

Technique	Data set	Fuzzy hierarchical clustering on samples		Fuzzy hierarchical clustering on manufacturers		R1[a]	R2[b]	*F*[c]	Rank	Technique
		Levels	Categories	Levels	Categories					
RPLC/(−)MS	Raw	5	13	3	4	0.500	1.67	0.833	1	RPLC/(+)MS (R)
	Scaled	3	5	2	1	0.125	1.50	0.188	2	RPLC/(−)MS (R)
RPLC/(+)MS	Raw	5	10	2	4	0.500	2.50	1.250	3	(−)MS (R)
	Scaled	2	3	2	0	0.000	1.00	0.000	4	RPLC/UV (R)
RPLC/UV	Raw	6	15	6	3	0.375	1.00	0.375	4	(+)MS (R)
	Scaled	6	15	6	2	0.250	1.00	0.250	5	UV (S)
(−)MS	Raw	5	12	3	2	0.250	1.67	0.417	6	RPLC/UV (S)
	Scaled	1	2	1	1	0.125	1.00	0.125	6	(+)MS (S)
(+)MS	Raw	3	6	2	2	0.250	1.50	0.375	7	RPLC/(−)MS (S)
	Scaled	2	3	2	2	0.250	1.00	0.250	8	(−)MS (S)
UV	Raw	5	8	5	1	0.125	1.00	0.125	8	UV (R)
	Scaled	5	8	2	1	0.125	2.50	0.313	9	RPLC/(+)MS (S)

[a] *R1* = (identified manufacturers + batch succession discrimination)/categories to be determined (8).
[b] *R2* = identified hierarchical levels based on samples/hierarchical levels classifying according to manufacturers.
[c] *F* is the ranking score.

(A–F) and two categories of batches (non-consecutive or consecutive ones, applying in case of manufacturer B, more precisely batch series 4–6 and 7–11, respectively, further denoted B1 and B2). Such an achievement is based theoretically on six hierarchical levels and isolation of eight different categories (the worst case situation was considered, corresponding to the isolation of manufacturer B in the fifths hierarchical level, followed by discrimination according to batch succession in the sixth one). By shifting from sample numbers to manufacturer's names in the hard partition trees resulting from application of the FHC algorithm on normal or scaled data, one can easily evaluate the discrimination power of each experimental investigation technique. The discussion is illustrated in Table 2. The ratio between the numbers of positively identified manufacturers (including batch succession in case of manufacturer B) and total categories to be determined, as well as the ratio between hierarchical levels needed to discriminate against samples and manufacturers, respectively, were tentatively used for ranking the classifying ability of FHC.

The fuzzy hierarchical clustering applied to raw data has increased discrimination power against manufacturers compared to the algorithm applied on scaled data. From the resulting partitions, mainly focusing on raw data processing, one can conclude that: (i) manufacturer A (samples 1–3) was clearly isolated against the others by all processed data sets derived from chromatographic separations coupled to the spectrometric detection systems, but only by individually taken (+) MS data; (ii) manufacturer B (samples 4–11) was clearly isolated against the others by processing raw data from the chromatographic separations coupled to MS detection (either under positive or negative ionization) as well as data from the mass spectra taken individually; RPLC/UV and UV data sets failed to classify manufacturer B with respect to the others; (iii) the only technique which was sensitive to batch succession under manufacturer B, grouping separately samples 4–6 and 7–11, is RPLC/(+)MS; (iv) manufacturers C–F are clearly classified against manufacturers A and B by RPLC/(−)MS, RPLC/(+)MS, RPLC/UV and (+) MS data sets; (v) manufacturer C is clearly isolated against the others when RPLC/UV data was processed; (vi) manufacturer D was isolated with respect to the other manufacturers when FHC is performed on scaled RPLC/(−)MS data; (vii) manufacturers E and F were classified against the others when RPLC/(−)MS raw data sets were processed; (viii) manufacturer E was isolated against the others only by processing RPLC/(−)MS raw data and RPLC/UV scaled data sets; (ix) manufacturer F was isolated against the others through the processing of RPLC/(−)MS, RPLC/UV data and (−)MS and UV data sets; (x) none of the data sets (raw or scaled) produced classification of manufacturers C and D against manufacturers E and F. As expected, the coupling of the spectrometric detectors to RPLC separation provides data leading through higher classification power in FHC processing.

According to the validity indices presented in Table 1 and Supplementary material: Part B, the highest discrimination was reached when applying FHC to RPLC-UV data sets. However, previous interpretations lead to the conclusion that RPLC/(−)MS is producing a more realistic discrimination of samples, which is in fair agreement with samples' origin (manufacturer). Although RPLC/UV (identifying the largest number of peaks) is leading to

Table 3
PCA eivenvalues for the first two principal components.

Analytical technique	Proportion (%)		
	PC1	PC2	Cumulative
RPLC/(−)MS	55.67	25.84	81.51
RPLC/(+)MS	45.86	30.13	75.99
RPLC/UV (220 nm)	98.94	0.63	99.57
(−)MS	85.68	8.42	94.10
(+)MS	90.16	3.66	93.81
UV	85.27	6.73	92.00

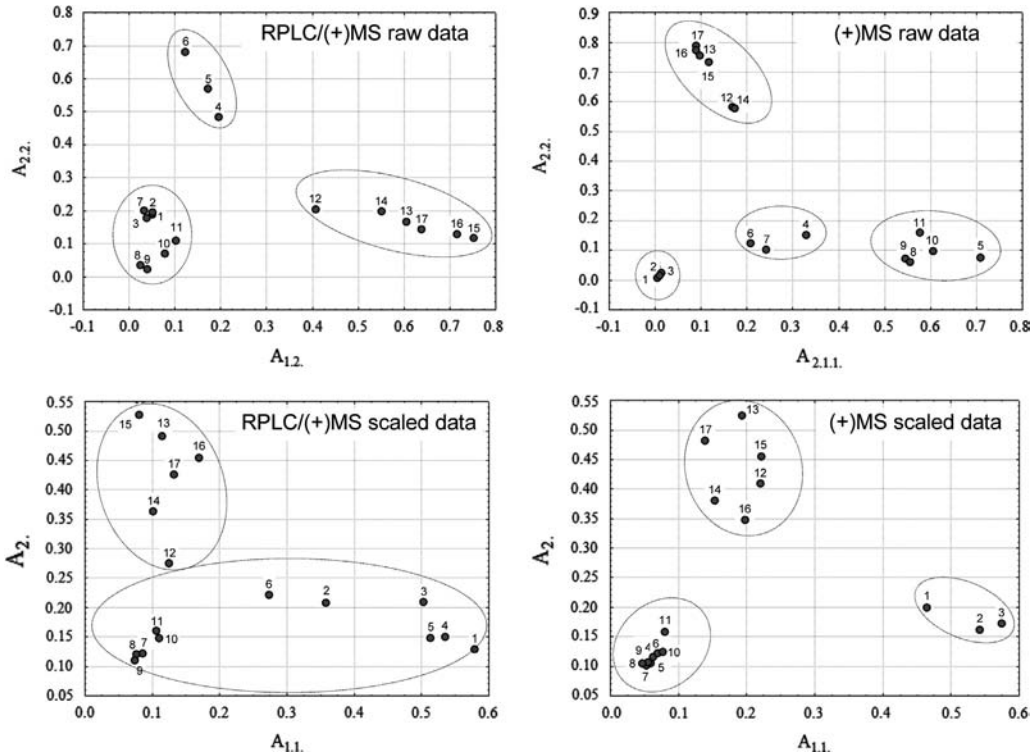

Fig. 3. Projection of all samples in the space defined by the membership degrees to various clusters obtained from RPLC/(+)MS raw data, (+)MS raw data, RPLC/(+)MS scaled data and (+)MS scaled data.

the highest number of clusters, its discrimination seems chaotic. Batch succession sensitivity was brought only by the RPLC/(+)MS approach. The classification power seems related to the intrinsic selectivity of the detection system response, and follows the (−) MS > (+) MS > UV (220 nm) order. Spectrometric techniques taken individually also obey to the selectivity rule. (+) MS relatively conserved the ability to classify against batch succession, indicating that for similar selectivity characteristics, the intrinsic sensitivity acts favorably on discrimination.

An alternative approach for evaluating the discrimination power of the processing algorithm in relation to manufacturer and batch succession may be offered by the graphical representation of the membership degrees. As a first approach, the evaluation of all possible combination of clusters has been made, to obtain the best classification with respect to the samples origins. Comparing the results provided by the processing of raw data against those from scaled data sets, one can conclude that raw data improved the grouping of samples in relation to manufacturers. This behavior may be a direct effect of the scaling process, which tends to gather the data and deteriorates the discrimination.

In another turn, as Fig. 3 illustrates, data sets obtained by the uncoupled (+) MS technique produced better results when

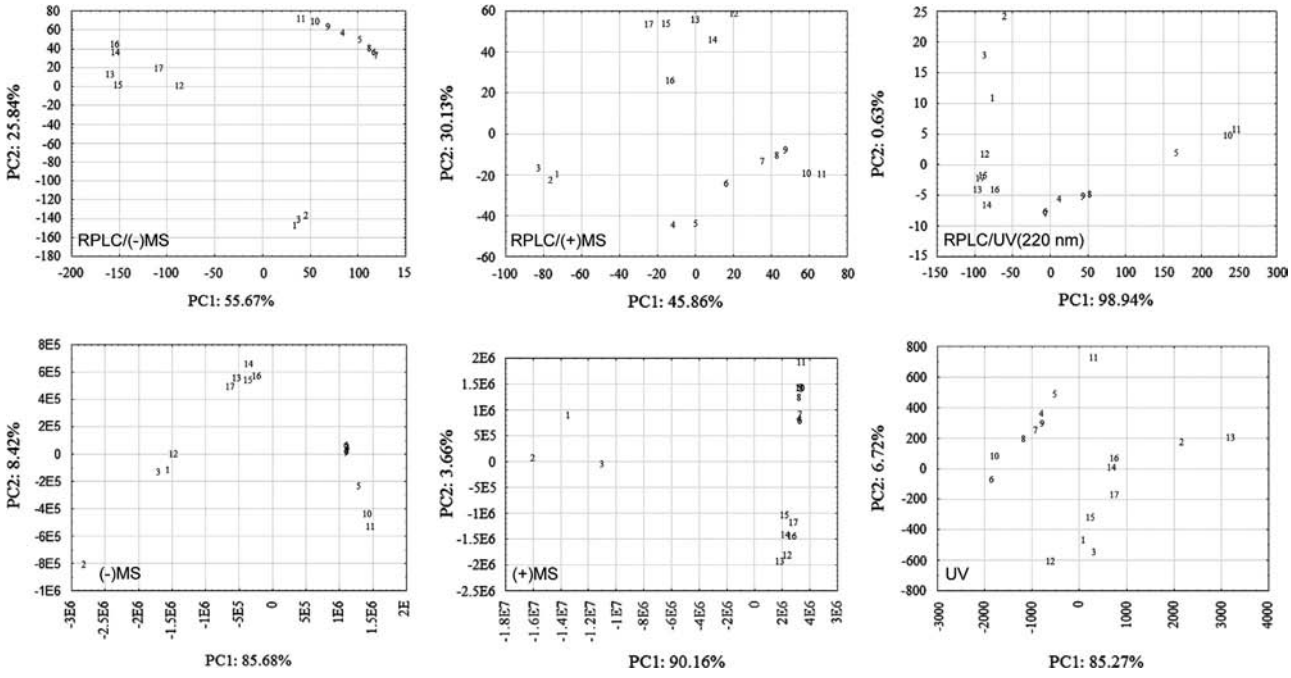

Fig. 4. PC1–PC2 score plots for RPLC/(−)MS, RPLC/(+)MS, RPLC/UV (220 nm), (−)MS, (+)MS, and UV.

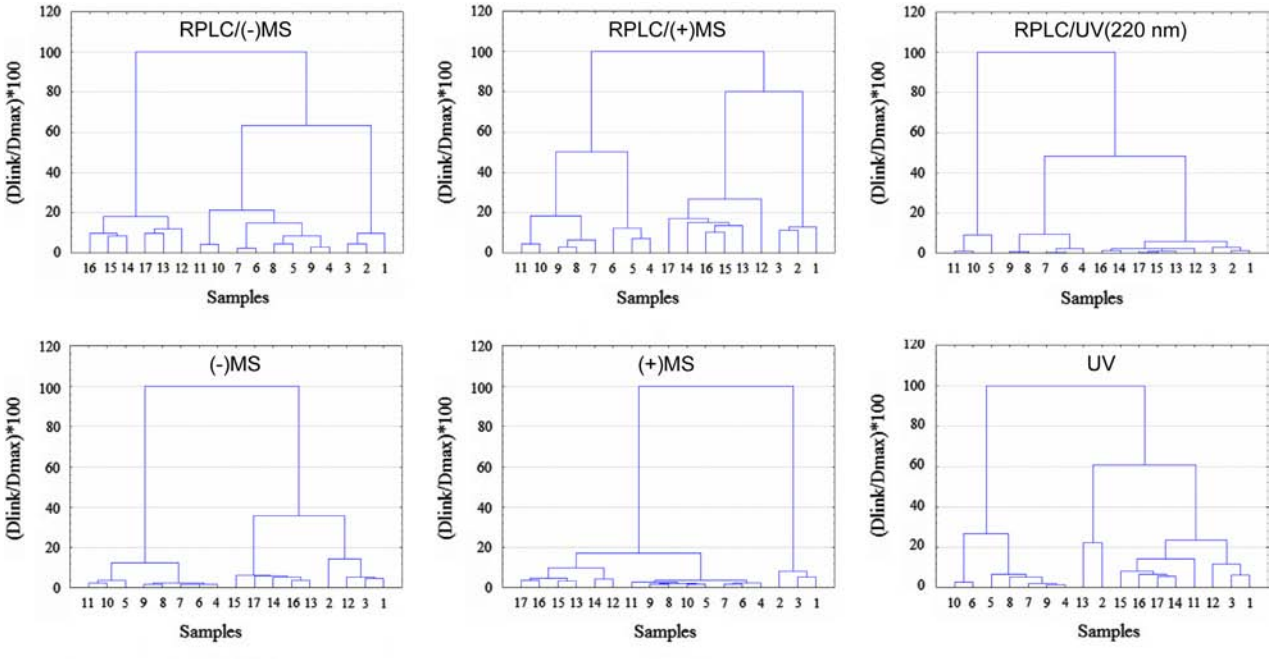

Fig. 5. Hierarchical clustering of the *Ginkgo Biloba* samples using: for RPLC/(−)MS, RPLC/(+)MS, RPLC/UV (220 nm), (−) MS, (+) MS, and UV.

comparing to RPLC/(+)MS data. This remains valid for both raw and scaled values.

For informational purpose, projection of all samples in the space defined by the membership degrees to various clusters obtained from raw and scaled data sets is presented in the Supplementary material: Parts D and E, respectively. According to the technique of investigation, classification power varies for raw data in the order UV < RPLC/(−)MS < RPLC/UV ≈ (−)MS < RPLC/(+)MS ≈ (+)MS, while for the scaled data the succession differs slightly: UV ≈ RPLC/(−)MS ≈ RPLC/(+)MS < RPLC/UV < (−)MS < (+)MS.

3.3. Comparison of FHC to PCA and CA results

The discrimination power provided by FHC may be also evaluated through comparison to results produced by the common classification methods, more specifically PCA and CA. The PCA eigenvalues enlisted in Table 3 confirmed the FHC results, meaning that the hyphenation of MS to RPLC substantially favors discrimination power between samples.

Both processing algorithms behave similarly on the given data sets. Figs. 4 and 5 contain classifications produced through application of PCA and CA, respectively.

Both PCA and CA produced isolation of manufacturer A against the others when considering all data sets, except those originating from (−) MS and UV techniques. Exactly the same situation could be observed on grouping manufacturers C–F against A and B. Manufacturer B was identified based on data provided by all techniques when processed by PCA. Under CA processing algorithm, RPLC/UV and UV taken individually failed to discriminate between manufacturer B and the other ones. Succession within manufacturer's B batches was obtained only with PCA and CA, based on RPLC/(+)MS data set. Exclusion of the RPLC dimension made processing procedures insensitive to the batch production order.

PCA and CA did not allow identification of manufacturers E and F, as individual batches. PCA and CA failed also to classify manufacturers C and D, each containing two batches. Under PCA processing conditions the classification power varied in the order UV < (−)MS < (+)MS ≈ RPLC/UV < RPLC/(−)MS < RPLC/(+)MS. The order is slightly modified when data were treated through the CA algorithm: UV < (−)MS < RPLC/UV < (+) MS < RPLC/(−)MS < RPLC/(+)MS.

4. Conclusions

It appeared from the experimental approach that batch to batch discrimination may be achieved only through RPLC/(+)MS and FHC or CA analysis of data. Manufacturers A and B were easily identified by either HPLC/MS or MS approaches. With respect to the ability of discrimination between manufacturers C–F, the FHC approach is more powerful compared to PCA and CA. Data interpretation through PCA allowed the lower classification power against the used chemometric treatments. Within the FHC treatment, processing of raw and centered data produced similar classifications, while scaling data induced a negative effect on the discrimination capacity that derived from their gathering tendency.

As expected, the raw intrinsic information produced by the analytical investigation technique of samples plays an important role in tuning the discrimination power. The chromatographic dimension was necessary to observe the batch to batch variability. However, the chromatographic separation does not add significant insights with respect to manufacturer's identification. The holistic comparison of the classification ability of the analytical techniques used during the present work results in the following ranking order: UV < RPLC/UV < (−)MS < RPLC/(−)MS < (+)MS < RPLC/(+)MS. This hierarchy seems difficult to interpret at a first sight, as long as the number of peaks in chromatograms is relatively the same irrespective to the detection system that was used, while signals in the (+)/(−) mass spectra are identical. The last position occupied by the UV spectrometry can be easily explained, since differences between samples are based on relatively small shifts of the three or four existing absorption bands. The ability of the MS spectral investigation modes to overcome UV is also obvious. The selectivity induced by the RPLC separation is balanced by the ionization characteristics of the ESI and the intrinsic sensitivity of the MS detection. Compounds in samples simultaneously ionize and the mild ionization technique seriously limits the fragmentation of the produced molecular ions. This results in a limited overlapping of signals in the resulting spectra, while conserving the ability to consider most of the components from the complex mixture, including minor ones. Thus, the mass spectra become perfect carriers of subtle differences and the discrimination ability increases. The superiority of (+) MS over (−) MS in detection is far more difficult to explain since negative ionization generally occurs more selectively compared to the positive one. However, one should consider that the composition of the mobile phase was optimized for positive ionization. Consequently, (−) MS spectra exhibit signals having at least one order of magnitude lower intensity compared to (+) MS ones, and the ability of observing minor components decreases. The increased sensitivity in the positive ionization mode accounts for a specific pattern added particularly by these minor compounds in the samples and largely contributes to the increase of the differentiation ability.

Acknowledgments

This work was possible with the financial support offered by the Romanian Ministry of Education, Research, Youth and Sport through research Grants PN-II-ID-PCE-2011-3-152 and PN-II-ID-PCE-201-3-0366.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2013.11.035.

References

[1] L.A. Beruetta, R.M. Alonso-Salces, K. Héberger, J. Chromatogr. A 1158 (2007) 196–214.
[2] M. Peris, L. Escuder-Gilabert, Anal. Chim. Acta 638 (2009) 1–15.
[3] P. Ciosek, W. Wróblewski, Talanta 71 (2007) 738–746.
[4] A. de Villiers, P. Alberts, A.C.J. Tredoux, H.H. Nieuwoudt, Anal. Chim. Acta 730 (2012) 2–23.
[5] Y. Huang, Y. Lan, S.J. Thomson, A. Fang, W.C. Hoffmann, R.E. Lacey, Comput. Electron. Agr. 71 (2010) 107–127.
[6] K. Villez, K. Steppe, D.J.W. De Pauw, Biosyst. Eng. 103 (2009) 23–34.
[7] J.T. Fischedick, A. Hazekamp, T. Erkelens, Y.H. Choi, R. Verpoorte, Phytochemistry 71 (2010) 2058–2073.
[8] S. Dussert, A. Laffargue, A. de Kochko, T. Joët, Phytochemistry 69 (2008) 2950–2960.
[9] C. Tistaert, B. Dejaegher, Y.V. Heyden, Anal. Chim. Acta 690 (2011) 148–161.
[10] A. Lanzarotta, J.B. Crowe, M. Witcowski, B.M. Gamble, J. Pharm. Biomed. Anal. 67 and 68 (2012) 22–27.
[11] A.B. Champagne, K.V. Emmel, Vib. Spectrosc. 55 (2011) 216–223.
[12] D.S. Wishart, Trends Food Sci. Technol. 19 (2008) 482–493.
[13] K.H. Liland, TrACs Trends Anal. Chem. 30 (2011) 827–841.
[14] I. García-Pérez, M. Vallejo, A. García, C. Legido-Quigley, C. Barbas, J. Chromatogr. A 1204 (2008) 130–139.
[15] A. Hildebrandt, M. Guillamón, S. Lacorte, R. Tauler, D. Barceló, Water Res. 42 (2008) 3315–3326.
[16] R.L. Olsen, R.W. Chappell, J.C. Loftis, Water Res. 46 (2012) 3110–3122.
[17] J. Sarembaud, R. Pinto, D.N. Rutledge, M. Feinberg, Anal. Chim. Acta 603 (2007) 147–154.
[18] European Medicines Agency. Note for guidance on quality of herbal medicinal products, London, 2001, p. 6. ⟨http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003370.pdf⟩ (accessed 27.02.2013 US).

[19] C. Sârbu, R.D. Naşcu-Briciu, A. Kot-Wasik, S. Gorinstein, A. Wasik, J. Namieśnik, Food Chem. 130 (2012) 994–1002.
[20] Food and Drug Administration. FDA guidance for industry – botanical drug products (draft guidance), Rockville, 2000, p. 4. ⟨http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070491.pdf⟩ (accesed 27.02.2013).
[21] H. Mark, J. Workman, Chemometrics in Spectroscopy, Academic Press-Elsevier, 2007.
[22] J.J. Harynuk, A.P. de la Mata, N.A. Sinkov, in: K. Varmuza (Ed.), Chemometrics in Practical Applications, InTech Publishing, 2012, pp. 305–326.
[23] A.C. Duarte, S. Capelo, J. Liq., Chromatogr. Relat. Technol. 29 (2006) 1143–1176.
[24] A.C. Moţ, F. Soponar, C. Sârbu, Talanta 81 (2010) 1010–1015.
[25] K.-H. Ott, N. Aranibar, B. Singh, G.W. Stockton, Phytochemistry 62 (2003) 971–985.
[26] L.A. Zadeh, Inf. Control 8 (1965) 338–353.
[27] T.R. Cundari, J. Deng, H.F. Pop, C. Sârbu, J. Chem. Inf. Comput. Sci. 40 (2000) 1052–1061.
[28] C. Sârbu, K. Zehl, J.W. Einax, Chemom. Int. Lab. Syst. 86 (2007) 121–129.
[29] C. Sârbu, J.W. Einax, Anal. Bioanal. Chem. 390 (2008) 1293–1301.
[30] J.C. Bezdek, R. Ehrlich, W. Full, Comput. Geosci. 10 (1984) 191–203.
[31] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorihms, Plenum, New York, 1987.
[32] S. Ding, E. Dudley, S. Plummer, J. Tang, R.P. Newton, A.G. Bretnon, Phytochemistry 69 (2008) 1555–1564.
[33] T.A. van Beek, J. Chromatogr. A 967 (2002) 21–55.
[34] T.A. van Beek, P. Montoro, J. Chromatogr. A 1216 (2009) 2002–2032.
[35] D.I. Sora, V. Stefanescu, V. David, A. Medvedovici, J. Pharm. Biomed. Anal. 50 (2009) 459–468.
[36] S. Udrescu, I.D. Sora, V. David, A. Medvedovici, J. Liq., Chromatogr. Rel. Technol. 33 (2010) 133–149.
[37] European Pharmacopoeia, seventh ed., Council of Europe, Strasbourg, France, Monograph 04/2008:1827, pp. 1233–1236.
[38] M. Halkidi, Y. Batistakis, M. Vazirgiannis, J. Intell. Inf. Syst. 17 (2001) 107–145.
[39] W. Wang, Y. Zhang, Fuzzy Sets Syst. 158 (2007) 2095–2117.
[40] K.L. Wu, M.S. Yang, Pattern Recognit. Lett. 26 (2005) 1275–1291.