Talanta 84 (2011) 471–479

Contents lists available at ScienceDirect

Talanta

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

Direct coupling of a gas–liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools

R. Garrido-Delgado[a], L. Arce[a], A.V. Guamán[b,c], A. Pardo[b], S. Marco[b,c], M. Valcárcel[a,*]

[a] *Department of Analytical Chemistry, University of Córdoba, Annex C3 Building, Campus of Rabanales, E-14071 Córdoba, Spain*
[b] *Departament d'Electrònica, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*
[c] *Artificial Olfaction Lab, Institute for BioEngineering of Catalonia, Baldiri i Rexach 4-6, 08028 Barcelona, Spain*

ARTICLE INFO

Article history:
Received 21 September 2010
Received in revised form
23 December 2010
Accepted 16 January 2011
Available online 26 January 2011

Keywords:
Classification
White wines
Ultraviolet-ion mobility spectrometry
Gas phase separate
Vanguard method
Continuous flow system
Chemometric analysis

ABSTRACT

The potential of a vanguard technique as is the ion mobility spectrometry with ultraviolet ionization (UV–IMS) coupled to a continuous flow system (CFS) have been demonstrated in this work using a gas phase separator (GPS). This vanguard system (CFS–GPS–UV–IMS) has been used for the analysis of different types of white wines to obtain a characteristic profile for each type of wine and their posterior classification using different chemometric tools. Precision of the method was 3.1% expressed as relative standard deviation. A deep chemometric study was carried out for the classification of the four types of wines selected. The best classification performance was obtained by first reducing the data dimensionality by principal component analysis (PCA) followed by linear discriminant analysis (LDA) and finally using a k-nearest neighbour (kNN) classifier. The classification rate in an independent validation set was 92.0% classification rate value with confidence interval [89.0%, 95.0%] at 95% confidence level.

The same white wines analyzed using CFS–GPS–UV–IMS were analyzed using gas chromatography with a flame detector (GC–FID) as conventional technique. The chromatographic method used for the determination of superior alcohols in wine samples shown in the Regulation CEE 1238/1992 was selected to carry out the analysis of the same samples set and later the classification using appropriate chemometrics tools. In this case, strategies PCA-LDA and kNN classifier were also used for the correct classification of the wine samples. This combination showed similar results to the ones obtained with the proposed method.

1. Introduction

One trend of analytical chemistry today is the development of new vanguard systems to address the high demand of quality information society requires. Briefly, we can summarize that vanguard systems provide fast and cheap results; also, systems require little or no sample preparation and can give qualitative rather than quantitative results. In them, productivity related properties are of great relevance at the expense of the level of basic properties [1]. Under this definition, ion mobility spectrometry (IMS) could be included into the vanguard analytical systems because it can provide analytical information of a high number of samples in a simple and rapid way.

Wine classification is an important topic, among other reasons, to detect possible frauds and to determine their authenticity with a view to assuring their quality to an expanding international market. The main purpose of Regulatory Councils is preventing fraud by guaranteeing the origin of theirs wines. The composition of wine samples is influenced by many and diverse factors corresponding to the specific area of production, such as grape variety, soil and climate, culture, yeast, winemaking practices, transport and storage [2]. Generally, the sensory analysis based on the trained experts panel test is useful in the wine classification task, but it is not always feasible because of high-cost and time consuming and sometimes without any objective estimation [3]. Therefore, the development and application of vanguard analytical systems suitable for the varietal authentication of wines are matters of interest in order to avoid fraud.

To date, the most common way to classify varietal wines is by monitoring the content of volatile aroma compounds mainly by employing gas chromatography (GC) [2,4–7]. Although recently, a new method based on an untargeted ultra high performance liquid chromatography coupled to ultra high resolution mass spectrometry approach has been used to classify wine samples from the metabolomic profiling [8]. But, although these conventional techniques are precise, selective, sensitive and the results that it provides should have a high level of metrological [1], it normally needs a process of sample preparation before chromatographic

---

* Corresponding author.
E-mail address: qa1meobj@uco.es (M. Valcárcel).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.01.044

analysis [9–13]. Therefore in recent years, the use of chemical sensors combined with multivariate statistical analysis techniques was an interesting alternative to classify wines [14–20]. The uses of sensors as electronic noses or artificial olfactory systems try to imitate a simplified human sense of smell but removing the subjective component implicit in it. These systems have the advantage of high portability for in situ and on-line measurements with lower cost [3]. Also they have high sensitivity, fast response, ease operation and capability to recognize different volatile compounds profiles. In this sense, IMS technique could be an alternative and appropriate sensor for the characterization and classification of different types of wines.

Moreover IMS equipment is very versatile and therefore different sample introduction systems (SIS) are an essential supplement for introducing volatile analytes in an efficient way into the IMS equipment. Different SIS coupled to IMS are described in the bibliography [21]. Recently, various SIS have been proposed by our research group, all of them are characterized by their simplicity and low cost to match with the advantage of the IMS technique [22,23]. In this work, an appropriate gas phase separator (GPS) [24] has been proposed for the in-line introduction of volatile analytes present in liquid samples in an IMS equipment. This system allows coupling a continuous flow system (CFS) to the IMS equipment. The CFS is characterized by their simplicity, ease of use and automatization. In addition they do not require special sample handling and consume little sample and reagent volume. Despite the potential advantages of CFS coupled with IMS equipment, this coupling has rarely been shown to be compatible [25–30]. In this case, it has been used for the extraction and injection of volatile analytes presents in wine samples and the classification of different types of white wines using their volatile analytes profiles.

Until this point, only the advantages of the IMS have been commented and this technique presents the problem that each analysis provides a huge amount of data. Therefore, an appropriate treatment of data is required to obtain high quality results. The multivariate statistical techniques provide the possibility to extract and use information based on the overall properties of the sample and perform a classification without the need for additional compositional data [31].

In summary, the CFS–GPS–UV–IMS proposed for the first time in this work has been used for the in-line extraction of volatile compounds present in wine samples to therefore obtain a characteristic fingerprint from each type of wine. Wine samples were classified taking into account the different origins using their profile of volatile compounds. Monovarietal white wine samples of four Certified Brands of Origin (CBOs) in southern Spain, with different alcohol content and different winemaking were used here to demonstrate the objective of this work. In order to confirm and validate the classification obtained with the CFS–GPS–UV–IMS (vanguard method) the results were compared with the classification obtained by GC coupled with flame ionization detector (FID) as rearguard method.

2. Materials and methods

2.1. Reagents and samples

A total of 54 white wine samples of four different origins from south of Spain were bought from local markets or were obtained from different CBOs. These wines were selected since their similar organoleptic characteristics. These wines could be potential products for frauds in the international market. For this reason, it is necessary to develop easy and reliable analytical tools to control the characteristic of each wine and to confirm the information of their labels. The CBOs used in this work were: *Montilla-Moriles* ($n = 15$), *Jerez* ($n = 13$), *Valdepeñas* ($n = 14$) and *Condado de Huelva* ($n = 12$). The wine samples were stored in their original container at ambient temperature during the realization of this work.

The studied wines were of two types as regards alcohol content. *Jerez* and *Montilla-Moriles* wines contained 15–23% ethanol while *Condado de Huelva* and *Valdepeñas* wines contained only 9–14.5%. Also each group of wine samples were elaborated with a type of different grape. Therefore, *Montilla-Moriles* wines made with *Pedro Ximenez* grape; *Jerez* wines made with *Palomino Fino* grape, *Condado de Huelva* wines made with *Zalema* grape and *Valdepeñas* wines made with *Airén* grape.

The compounds used to calibrate the chromatographic method were: acetaldehyde, ethyl acetate, methyl acetate, 2-methyl-1-butanol, 3-methyl-1-butanol, methanol, isobutanol, 1-butanol, 1-propanol, 2-butanol and acetoin. All were supplied by Sigma (St. Louis, MO, USA.)

2.2. Ion mobility spectrometer

Measurements were made with a portable UV–IMS instrument from Gesellschaft für Analytische Sensorsysteme (GAS, Dortmund, Germany). The UV–IMS system was 350 × 350 × 150 mm in size, had a tube length 12 cm, weighed 5 kg and used a constant electric field of $333\,\mathrm{V\,cm^{-1}}$. The ionization source was a 10.6 eV UV lamp. The instrument was operated at ambient pressure and temperature.

GASpector software was used to record spectra, which were acquired in the positive ion mode. A total of 50 spectra were continuously recorded for about 2 min per analysis, each spectrum being the average of 32 scans. The spectrum length was 1024 points, the grid pulse width 500 μs, the repetition rate 50 ms and the sampling frequency 30 kHz.

2.3. Gas phase separator as sample introduction system

The manifold employed for the in-line vapor phase generation and mobility measurements is depicted in Fig. 1. The SIS selected in this work was previously used by Armenta et al. [24] for seafood freshness determination using Fourier transform infrared (FTIR) spectroscopy. The SIS consisted in a peristaltic pump of four channels, furnished with Teflon tubes, which was used to introduce the wine samples in the flow system. The connections and reaction coils were made from Teflon tubing. The gaseous phase was separated from the liquid stream in a home-made Pirex-glass of 4.7 mL volume GPS. The manifold incorporates an oven to improve the removal of volatile compounds from wine samples.

Wine samples were continuously aspirated inside the manifold shown in Fig. 1 at $0.7\,\mathrm{mL\,min^{-1}}$ flow. Wine samples were passed through a 100 cm length coil heated at 80 °C and introduced in the GPS from which the generated volatile compounds were transported to the UV–IMS equipment after 15 min of generation of volatile compounds using a N_2 carrier flow of $20\,\mathrm{mL\,min^{-1}}$. During these 15 min, the stream of highly pure nitrogen (6.0) from Abelló Linde (Barcelona, Spain) was passed through the UV–IMS system for cleaning and stabilization while drift gas (N_2) pass in crosscurrent to a flow of $70\,\mathrm{mL\,min^{-1}}$.

2.4. Gas chromatographic method

Acetaldehyde, ethyl acetate, methyl acetate, 2-methyl-1-butanol, 3-methyl-1-butanol, methanol, isobutanol, 1-butanol, 1-propanol, 2-butanol, and acetoin were determined in wine samples using Gas Chromatograph HP 6890 coupled to FID and an autosampler HP 7683. A CPWAX-57-CB (25 m × 0.25 mm) capillary column was used. The injection was made in split mode (20:1) using a temperature of 250 °C and 0.5 μL of sample volume were injected

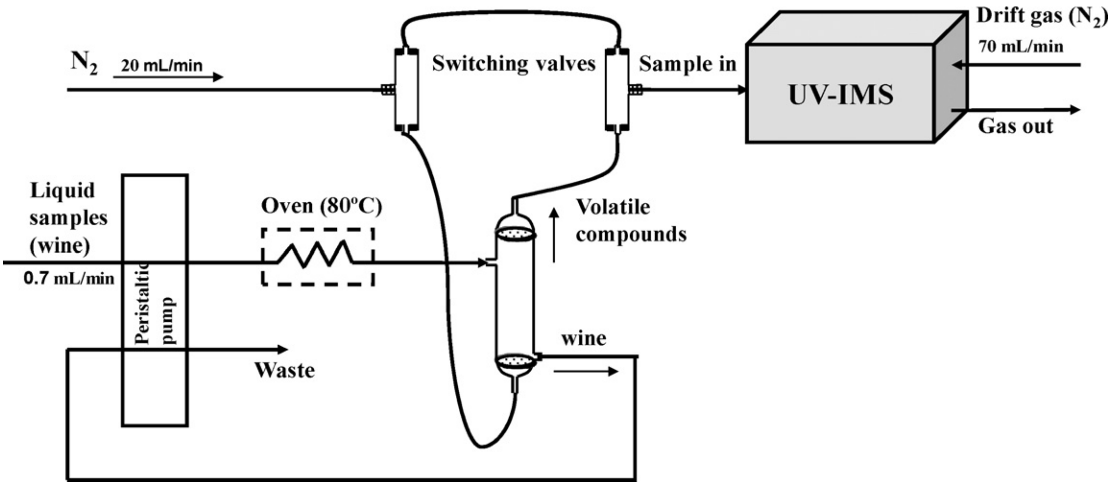

Fig. 1. Manifold employed for the in-line vapor phase generation from wine samples and analysis of the volatile compounds by UV–IMS.

to a pressure of 9.96 psi. Oven temperature program started at 40 °C and immediately a gradient of 1 °C min^{-1} ramp was applied up to 60 °C. Finally a 30 °C min^{-1} ramp up to 180 °C held for 1 min was applied. Helium (purity 99.999%) was used as carrier gas at a flow of 1 mL min^{-1}. The detector temperature was set at 250 °C, with air flow of 400 mL min^{-1}, hydrogen flow of 40 mL min^{-1} and nitrogen make up flow of 29 mL min^{-1}. 4-Methyl-2-pentanol was used as internal standard. The chromatographic method following in this work, used for the determination of superior alcohols in wine samples, was shown in the Regulation CEE 1238/1992.

2.5. Data analysis from IMS and GC

From the 50 IMS spectra continuously recorded for each sample, a data set with only spectra in which signal appear (from 15 to 49 spectrum) has been built and used for the statistical and pattern recognition analysis (see Fig. 2). The drift time range used for chemometric analysis was from 15.4 to 27 ms (351 variables in all). A smoothing procedure using a Savitzky–Golay filter, a baseline correction and a peak alignment procedure were performed on all spectra. Additional details about the specific implementation of these pre-processing techniques will be given later.

Two data sets from GC analysis have been built using the peak areas of the chromatograms. The first data set includes the eleven peak areas identified by the chromatographic method and the second data set includes the thirty-six areas of the peaks that appear throughout the chromatogram.

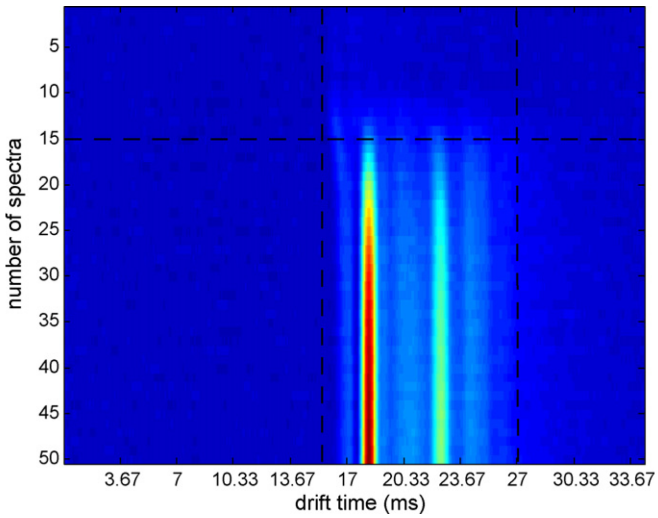

Fig. 2. Topographic plot of an ion mobility measurement of a white wine sample.

The multiway chemometric data processing were performed using the MATLAB (The Mathworks Inc., Natick, MA, USA, 2007) and PLS Toolbox 5.2.2 (Eigenvector Research, Inc., Manson, WA, USA). The techniques used for the classification of the wine samples were principal component analysis (PCA) and linear discriminant analysis (LDA) for dimensionality reduction and the k-nearest neighbour (kNN) for classification. For the validation of the results, hold out validation and Bootstrap validation techniques have been implemented. Moreover significant differences in each compound identified by GC in the wines analyzed were determined by one-way analysis of variance (ANOVA) using in this case the package software SPSS 12.0 (SPSS Inc., Chicago, IL).

PCA [32] is a signal processing technique that generates projections along the directions of maximum variance of the analyzed data. It has also been used as an unsupervised technique to beat the course of dimensionality generating projections of the original data in a reduced dimension where the maximum variance is preserved.

The term LDA [32,33] is a signal processing technique that generate discriminant functions that are linear in the input variables. The resulting linear combination may be used as a linear classifier or for dimensionality reduction before classification. In a specific sense, the technique maximizes class separability, generating projections where the examples of each class from compact clusters and the different clusters are far from each other. LDA has been used as supervised linear projection technique to find directions of maximum separation from a set of samples for which class membership is known in order to be able to predict the class membership of unknown samples.

The kNN [32] method is a technique used to generate non-linear classifications, finding the closest *k* examples in the dataset to the unknown class, and selecting the predominant class for it. kNN has been used for the classification of the wine samples in their correct group. In this work, value of *k* equal to 3 was used.

The hold out validation technique splits the initial data set in two subsets: the training subset, where the model is built; and the validation subset, where the performance of the model is measured. Normally, less than a third of the initial sample is used for validation data. However, in small data sets, the hold out validation technique could provide misleading values of performance. In order to overcome this limitation, the Bootstrap validation procedure allows to resample the data set with replacement, to generate sets of observations that may be used for the estimation of statistic magnitudes as the performance of the model [34].

3. Results and discussion

A full optimisation of the CFS–GPS–UV–IMS was firstly carried out to extract the volatile compounds present in wine samples and in-line analysis by UV–IMS equipment. Subsequently, a signal and data processing study was needed to classify wines according to their origin from the IMS recorded spectra. Later, the analysis of the same set of wine samples by GC–FID were carried out so to compare the obtained results by IMS to demonstrate the potential of this technique.

3.1. Optimization of the sample introduction system

Although the SIS used in this work has been used in a previous work coupled to a FTIR [24], in this application it has been reoptimized because of different features between UV–IMS and FTIR. The main objective of the optimization of the SIS was to observe characteristic bands in the ion mobility spectra of wine samples which differ for the different types of wines analyzed.

The experimental variables optimized were: (i) the temperature of the oven to produce volatile compounds from wine samples; (ii) the time that the wine was passing through the GPS to accumulate volatile compounds in this device; and (iii) sample gas and drift gas flows. The optimization was carried out using a representative wine sample of each type.

The first variable optimized was the temperature of the oven. Values between ambient temperature and 100 °C were used. This parameter was important to favour the extraction of volatile compounds from the wine matrix. 80 °C was considered as the optimum value since at this temperature the bands of the ion mobility spectra of the wine samples were better defined compared with other temperatures tested.

The second variable optimized and one of the most important was the time that the wine passes across the GPS. For it, times between 0 and 30 min were tested. 15 min was considered as optimum although more characteristics bands begin to appear after 15 min. These bands were more intense after 20 min. But the objective of our work was not to get the maximum sensibility. Therefore, the wine analyzed passed across the SIS during 15 min to reach a compromising situation between maximum signal and time of analysis.

Finally, the last variable optimized was the N_2 flow rate. The SIS proposed presents the problem that it cannot be used with high sample gas flows because the gas could sweep away to the liquid sample to the IMS equipment which could be dangerous for the equipment. With this limitation 20 mL min^{-1} was used. Also the drift gas greatly affected the ion mobility profile of the wine samples. Drift gas was optimized between 20 and 90 mL min^{-1}. At low values of drift gas no characteristics bands were observed in the spectrum. These bands began to appear from 70 mL min^{-1}. Therefore this valour of drift gas was selected as optimum.

3.2. Ion mobility spectra data processing and pattern recognition of wine samples

As it has been mentioned before, a data set has been formed with the IMS spectra of the different wines. Figs. 3 and 4 show the mean spectra for every wine and their standard deviation. As can be observed in the standard deviation (dotted line) from both figures, the variability of the volatile compounds profile of the samples within each group is small except for the *Condado de Huelva* group. As can be seen in Fig. 2 each IMS measurement comprised 50 spectra but only 34 spectra (15–49) were used for data evaluation. Due to the sampling procedure, the first 14 spectra and the last one do not include relevant information. The strategy of using 34 spectra has been employed rather than the more usual single scan at the maximum, in order to ensure adequate representativeness of each analysis. The relevant information in all cases was included only in the spectral region between 15.4 and 27 ms (351 variables). The dimension of the data matrix is 54 wine samples × 351 variables and, before the outlier detection procedure, each wine sample includes 34 spectra.

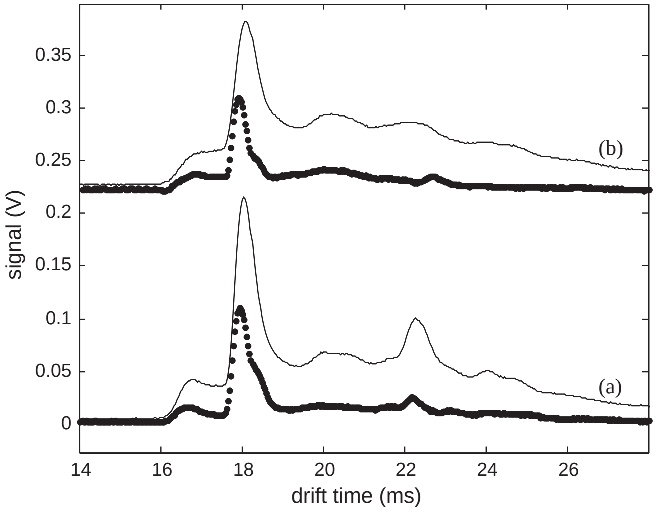

Fig. 3. Average spectra (solid lines) and standard deviation (dotted lines) for wines from group I. (a) *Montilla-Moriles* wines and (b) *Jerez* wines.

Data must be carefully pre-processed, since any inaccuracy introduced at this stage can cause significant errors in the statistical analysis. First a smoothing Savitzky–Golay filter of order 3 was used to improve the signal to noise ratio of all spectra of wine samples. Later, the baseline from each spectrum was corrected subtracting the mean value of an empty area of peaks (between 0 and 15 ms), common to all the original spectra of wine. Additionally, all spectra were aligned with a shift in *x*-axis based on a polynomial function fitted to a reference peak [35]. New positions of the peaks are maximally close among the different spectra.

Finally, the precision of the method proposed was assessed by analysing the same wine sample on the same or three different days

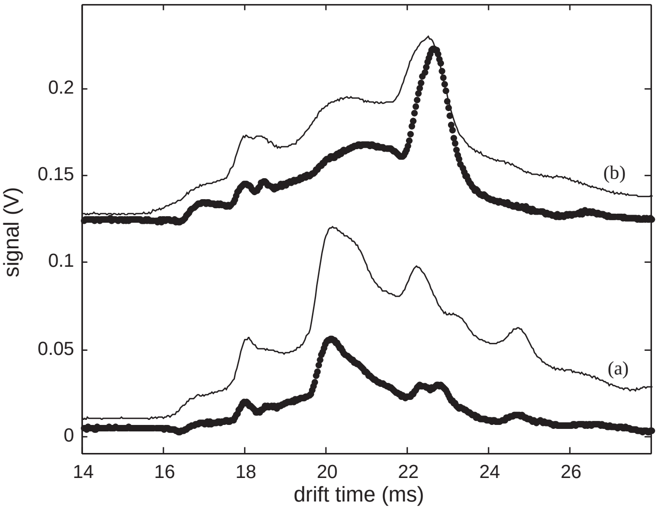

Fig. 4. Average spectra (solid lines) and standard deviation (dotted lines) for wines from group II. (a) *Valdepeñas* wines and (b) *Condado de Huelva* wines.

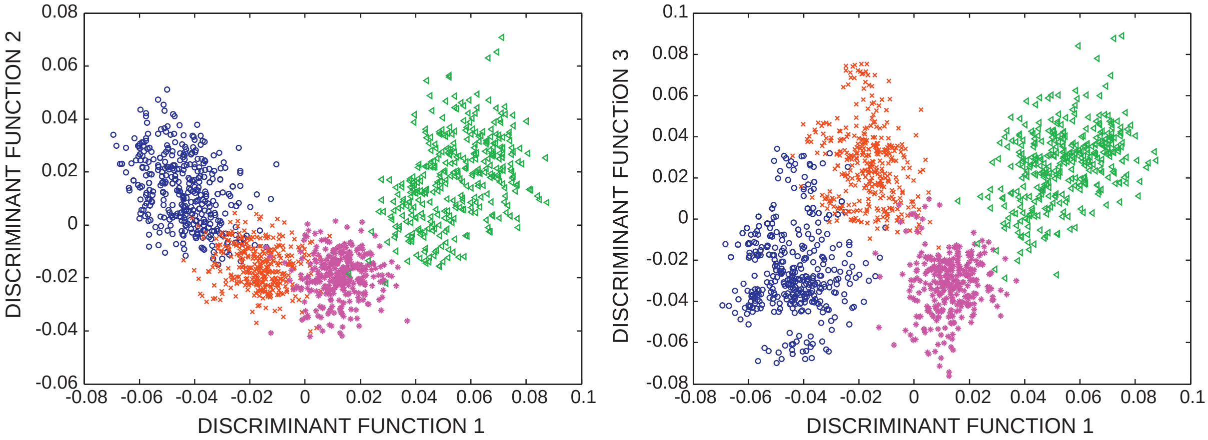

Fig. 5. Scatter plot for the LDA obtained using 16 PCs on training set from IMS data. Samples: *Montilla-Moriles* wines:◦; *Jerez* wines: ×; *Valdepeñas* wines: *; *Condado de Huelva* wines: Δ.

under identical testing conditions. The within-day precision was obtained in eleven replicates on the same day and the between-day precision was obtained in three replicates within three consecutive days. The within-day and between-day precision values were obtained using all data of the range selected in each sample (mean value of 34 spectra per sample from 15.4 to 27 ms). The within-day and between-day precision values obtained were 2.2% and 3.1% respectively calculated as relative standard deviation (RSD).

In order to find possible disturbing outliers for the pattern recognition analysis, a Hotelling's *T* square statistic and *Q* statistic tests has been implemented. For both, a confidence interval of 95% has been used. An IMS spectrum has been considered as an outlier if either values of the *T* square test or the *Q* test exceeded the confidence limit. A wine sample has been considered as an outlier if the majority of its spectra (more than 17) were outliers. As a result of these tests, three samples were discarded from *Montilla-Moriles* wines, two samples were discarded from *Jerez* wines, three samples were discarded from *Condado de Huelva* wines and three samples were discarded from *Valdepeñas* wines. Therefore, after the discarded samples, the final data set for the analysis is composed of 43 samples, 12 samples from *Montilla* wine, 11 samples from *Jerez* wine, 9 samples from *Huelva* wine and 11 samples from *Valdepeñas* wine. Due to the removal of some within sample spectra, it must be highlighted that, now, each wine sample does not contain the same number of spectra. The dimension of the new data matrix is 43 × 351 and the total amount of spectra is 1424. Table 1 summarizes the data and the distribution of spectra per sample per each type of wine sample analyzed.

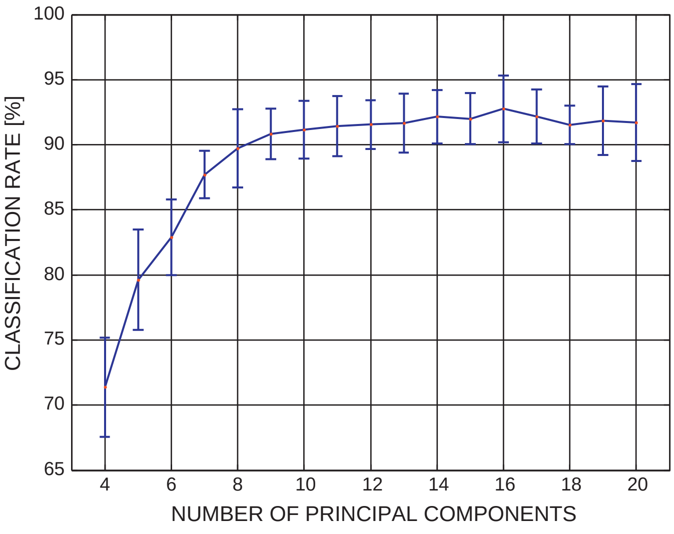

Fig. 6. Scanning plot of the PCA-LDA strategy from 4 until 20 PCs.

For the first pattern recognition analysis, the data matrix has been splitted in two subsets (Table 1), one to estimate the calibration model (training set), containing 31 samples (approximately 75% of the total amount) and the other to validate the model (validation set), with 12 samples (approximately 25% of the total amount). As a next step in the pattern recognition model estimation, PCA was applied on the training data set. Data were mean centered and the criterion used for selecting the number of principal components (PCs) was to collect more than 90% of the total variance. The first 3 PCs were selected, because jointly explained 95.7% of the variance. However, by this approach, the different wine samples could only be separated by their alcohol content which it was not the objective of this work.

In order to improve the wine clustering beyond their alcohol contents, an additional LDA signal processing technique has been applied. A strategy of dimensionality reduction through PCA has been used previously to perform LDA in order to overcome the known tendency of this algorithm to overfitting in small-sample-size problems, where the dimensionality is higher than the number of vectors in the training set [36]. Hold out validation has not been used due to the small number of samples in the validation set; a Bootstrap technique [34] has been implemented instead. Under the Bootstrap validation procedure, the training set is randomly selected (with replacement) over the total number of data and the remaining samples that were not selected for training are used for the validation. This procedure is repeated for a specific number of folds ($B = 100$). It must be again highlighted that selection has been done over the samples, not over the spectra, i.e. to select a wine sample means to select all its corresponding spectra. For every step in the procedure, PCA and LDA combination is built using the information of the training set. Then the validation set is projected over the model and a kNN classifier ($k = 3$) is used for estimate the classification rate of the model. A sample is assigned to a wine class through a majority vote procedure, i.e. if the majority of its spectra belong to that wine class. Thus the percentage of prediction is made for samples and not by spectra. Applying the PCA–LDA combination strategy using 16 PCs (99.9% of the variance) the four groups of wines are projected in separate clusters (Fig. 5). Table 2A represents the confusion matrix for this solution. The classification rate on the validation set using 3NN classifier is 93.2% with confidence interval [90.5%, 95.5%] at $P = 0.05$ confidence level.

Table 1
Summary of data set from IMS used for the chemometric treatment.

	Samples	*Montilla-Moriles*	*Jerez*	*Condado de Huelva*	*Valdepeñas*
		Number of spectra per sample			
	1	34	32	–	33
	2	34	–	34	34
	3	34	34	34	–
	4	29	34	34	–
	5	34	34	34	34
	6	–	31	34	34
	7	–	32	34	34
	8	34	34	34	29
	9	34	32	34	33
	10	34	34	–	34
	11	33	34	34	31
	12	28	31	–	34
	13	34	–		30
	14	34			–
	15	–			
		12 Samples 396 spectra	11 Samples 362 spectra	9 Samples 306 spectra	11 Samples 360 spectra
Data set	Training	9 Samples	8 Samples	6 Samples	8 Samples
	Validation	3 Samples	3 Samples	3 Samples	3 Samples

A scanning from 4 until 20 PCA dimensions has been done in order to test the performance of the PCA–LDA strategy. Fig. 6 shows the evolution of the overall classification rate with error bars representing the confidence interval at 95% confidence level. It can be seen that the best classification rate is achieved with 16 PCs but on the other hand, taking into account the statistical significance, the classification rate of 8 PCs is comparable to the best solution, with a 92.0% classification rate value with confidence interval [89.0%, 95.0%] at $P = 0.05$ confidence level. Table 2B shows the confusion matrix for the 8 PCs solution.

3.3. *GC data processing and pattern recognition of wine samples*

In order to compare the performance of the IMS analysis, the same set of wine samples were measured in duplicate by GC, following the methodology described in section 2.3. One way ANOVA using the Tukey's B test was carried out according to the type of wine for their correct classification with the data of concentration from the wines analyzed by GC. The results obtained are summarized in Table 3. As it can be seen in Table 3, compounds as acetaldehyde, methyl acetate, 1-propanol, isobutanol, 1-butanol, 2-methyl-butanol and acetoin classify to the wines in two groups. *Montilla-Moriles* and *Jerez* (higher alcohol content) have been clustered in one group and *Condado de Huelva* and *Valdepeñas* (lower alcohol content) have been clustered in other group. Only 3-methyl-butanol classifies the wines in four groups. There is a high concentration of this compound in the wines analyzed. Satisfactory results were not obtained by applying ANOVA to the other compounds (methyl acetate, methanol and 2-butanol).

The use of concentration values obviously requires prior quantification, which is often complex and time consuming. Therefore, the areas of the peaks identified by GC (11 peaks) and of all the other peaks from chromatogram (36 peaks) were used to carry out the chemometric treatment.

Once the outliers have been eliminated, as in the case of the IMS data set analysis, the final data set was conformed by a matrix with 100 samples and 11 variables. The matrix data has been splitted in two subsets, one for the estimation of the pattern recognition model (72% of the total of samples) and the other for the validation (28% of the total of samples) of the estimated model. As in the case of the IMS presented above, the next step has been to apply PCA on the training subset. In this case, data were autoscaled and the criterion used for selecting the number of PCs was also collecting more than 90% of the total variance. Following this criterion, 7 PCs have been

Table 2
Confusion matrix obtained for the 16 PCs solution (A) and 8 PCS (C). V1: Montilla-Moriles wines; V2: Jerez wines; V3: Condado de Huelva wines; V4: Valdepeñas wines.

Confusion matrix A (16 PCs)		Predicted			
		V1	V2	V3	V4
Real	V1	92%	7%	1%	0%
	V2	4%	93%	3%	0%
	V3	1%	7%	89%	3%
	V4	0%	0%	2%	98%
Conf. inter. ($P = 0.05$)		89–94%	91–95%	85–92%	96–99%
Overall percentage			93.2% (90.5–95.5%)		

Confusion matrix B (8 PCs)		Predicted			
		V1	V2	V3	V4
Real	V1	93%	7%	0%	0%
	V2	10%	87%	3%	0%
	V3	3%	6%	87%	4%
	V4	0%	0%	1%	99%
Conf. inter. ($P = 0.05$)		90–95%	83–90%	83–91%	97–99.5%
Overall percentage			92.0% (89.0–95.0%)		

Table 3
Average values (mg L^{-1}) and standard deviation (mg L^{-1}) of all wine samples analyzed by GC–FID method for each wine group.

Compounds	*Montilla-Moriles*		*Jerez*		*Condado de Huelva*		*Valdepeñas*	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Acetaldehyde	174^a	66	190^a	114	13^b	23	24^b	29
Methyl acetate	4^{ab}	3	1^a	2	7^b	6	2^a	3
Ethyl acetate	71^a	25	22^b	17	52^c	22	42^c	12
Methanol	71^a	9	79^a	13	58^b	17	78^a	13
2-Butanol	2^a	4	$0{,}4^a$	1	0^a	0	0^a	0
1-Propanol	55^a	26	47^a	10	34^b	8	30^b	11
Isobutanol	53^a	11	60^a	5	29^b	11	33^b	10
1-Butanol	8^a	2	7^a	1	0^b	0	0^b	0
2-Methyl-1-butanol	47^a	6	45^a	5	26^a	5	30^a	10
3-Methyl-1-butanol	228^a	35	187^b	23	130^c	13	159^d	38
Acetoin	31^a	21	32^a	19	15^b	11	9^b	17

Superscripts indicate the results of pairwise comparison using Tukey's *B* test for a confidence interval of 99%. Identical superscripts non-significant differences.

selected (92.1% of the variance). The projected data on a figure PC1 versus PC2 shows, as in the IMS case, a separation in terms of alcohol content. In order to improve separability, LDA has been applied on the PCA model, and a kNN classifier with hold out validation has been used to quantify separability. However the classification rate on the validation set only reaches 75% with confidence interval [55.1%, 89.3%] at confidence level 95%.

Similar pattern recognition development has been applied to a new data set consisting of the areas of the 36 peaks from the chromatogram. After removing outliers, the final data set consists of a matrix with 101 samples and 36 variables. The estimation data set encloses 72 samples (71.3% of the total of samples) and validation set 29 samples (28.7% of the total of samples). PCA has been used on autoscaled estimation set. 17 PCs which explain 91.2% of the variance were selected. The four groups show good separability as it can be seen in Fig. 7 where PC1 versus PC2 are represented. The score plot resulted to the use of the combined PCA–LDA strategy with PCA dimension 17 reduced is shown in Fig. 8. Using 3NN with hold out validation, a percentage of 96.5% of good classification is obtained, with a confidence interval [88.2%, 99.9%] at $P = 0.05$ confidence level. Just a single sample, corresponding to *Valdepeñas* wine was bad labeled.

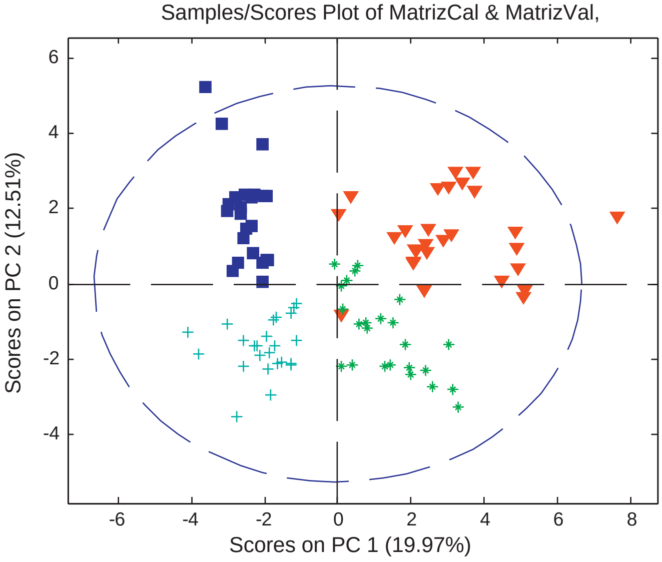

Fig. 7. PCA results for the exploratory analysis from the four classes of wine samples analyzed by GC using the areas of 36 peaks from chromatogram. Samples: *Montilla-Moriles* wines: ▲; *Jerez* wines: *; *Condado de Huelva* wines: ■; *Valdepeñas* wines: +.

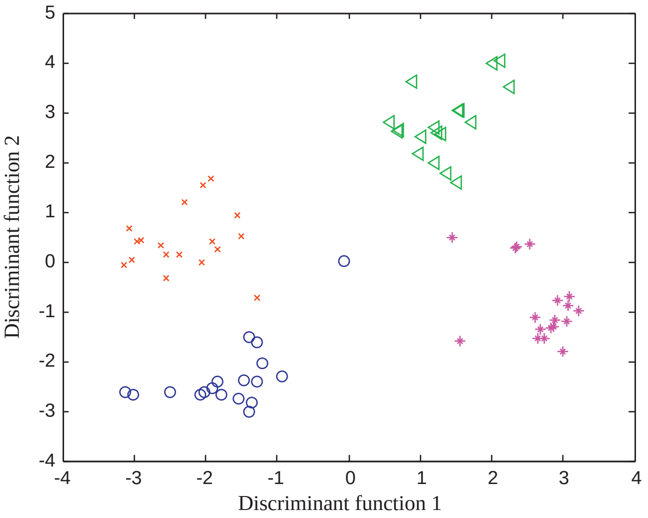

Fig. 8. Score plot for the LDA obtained using 17 PCs of wine samples analyzed by GC using the areas of 36 peaks from chromatogram. Samples: *Montilla-Moriles* wines:$^\circ$; *Jerez* wines: ×; *Valdepeñas* wines: *; *Condado de Huelva* wines: Δ.

3.4. *Comparison of the main features of the CFS–GPS–UV–IMS and GC–FID methods*

Table 4 summarizes the main characteristics of the methods developed in this work. Although a good classification of the white wine samples has been achieved with both methods, with the CFS–GPS–UV–IMS method none sample pretreatment is required while in the chromatographic method a prior dilution and addition of the internal standard to the wine sample are necessary before analysis. Others advantages of the IMS technique are the short analysis times and its lower cost of the CFS–GPS–UV–IMS compared with a GC system.

Table 4
Comparison of the most salient features of the methods developed for the classification of white wine samples.

Variable	CFS–CPS–UV–IMS	GC–FID
Sample volume	10 mL	0.5 μL
Analysis time	2 min	20 min
Sample preparation	None	Dilution and addition internal standard
Time generating of headspace	15 min	–
Costs	Low	High
Classification	92%	96.5%

Table 5
Compounds analyzed by GC–FID$^{(1)}$ and CFS–CPS–UV–IMS$^{(2)}$.

Compounds	Ionization potential (eV)	Boiling temperature (°C)	Vapor pressure (mmHg to 20 °C)	GC$^{(1)}$	IMS$^{(2)}$
Acetaldehyde	10.23	20.08	756.8	√	√
Methyl acetate	10.25	57	165	√	√
Ethyl acetate	10.01	77	73	√	√
Methanol	10.85	64.7	97.7	√	x
2-Butanol	9.88	99	12.5	√	x
1-Propanol	10.22	97.1	14.9	√	x
Isobutanol	10.02	108	8	√	x
1-Butanol	9.99	118	5	√	x
2-Methyl-1-butanol	>10.6	128	3	√	x
3-Methyl-1-butanol	>10.6	132	2	√	x
Acetoin	>10.6	148	2.7 (25 °C)	√	x

√: It can be identified by GC or IMS. X: It cannot be identified by GC or IMS.

The same set of samples were previous analyzed using capillary electrophoresis (CE) and the results obtained were already published [37]. With this separation technique the better classification was achieved making a chemometric treatment of the CE data according to their alcohol content. The results obtained with IMS and GC present the advantage over CE results that the obtained classification were independent from the alcohol content present in the wine samples analyzed.

3.5. Possible identification of the profile of the wine samples analyzed by IMS

The wine present a huge number of volatile compounds, such as, alcohols, esters, aldehydes or ketones [38]. Although the objective of the work was not to carry out the identification of the volatiles compounds presents in the wine samples, the profile of the spectra from wine samples obtained by CFS–GPS–UV–IMS was studied to identify some analytes. To achieve this goal, all the compounds identified by GC–FID method were studied to check if these analytes could also be determined by UV–IMS. In Table 5, the boiling point, ionization potential and vapor pressure to ambient temperature of these compounds are summarized. Only the compounds with an ionization potential lower that 10.6 eV can be ionized by a UV lamp. Therefore, all the compounds with an ionization potential lower that 10.6 eV could be determined using the proposed IMS method but only three compounds (acetaldehyde, methyl acetate and ethyl acetate) shown signal. Although the other compounds have an ionization potential lower that 10.6 eV their boiling points are above 80 °C. Moreover these compounds have a very low vapor pressure at room temperature for all these reasons, these compounds were not identified in the ion mobility spectrum. In Fig. 9, the ion mobility spectrum of a wine and the ion mobility spectrum of the same wine spiked with a mixture of acetaldehyde, methyl acetate and ethyl acetate and with an alcoholic degree similar to the present in a wine are shown. IMS has the disadvantage that it is difficult to identify peaks in their spectra due to ion–molecule reactions and its low resolving power, but by comparing the ion mobility spectrum of a wine with one spiked wine sample shows an increase can be seen in the bands from spectrum between 15 and 27 ms. Therefore, in this preliminary study we could confirm that acetaldehyde, ethyl acetate and methyl acetate contribute to obtaining the fingerprint of a wine sample analyzed using the proposed method.

For this reason a correlation between the IMS and GC data from wine samples analyzed were developed using partial least-squares (PLS) regression with full cross validation (random subsets) and mean center prior of the data. Coefficient of determination in cross validation (R^2) and the standard error in cross validation (RMSECV) were 0.86 and 58.70 mg L^{-1} respectively. This study could contribute to confirm that three volatile chemical compounds present in wine (acetaldehyde, methyl acetate and ethyl acetate) could be measured using IMS technique. Further development with large data sets will be required to test the predictive ability of the IMS calibration models developed.

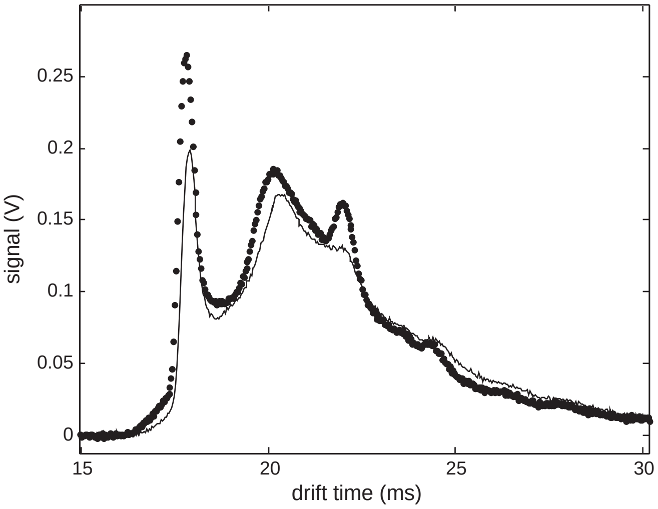

Fig. 9. Ion mobility spectra of a wine (solid line) and of a wine sample with mix of three compounds – acetaldehyde, methyl acetate, ethyl acetate (dotted line).

4. Conclusion

In this work, a vanguard analytical system (CFS–GPS–UV–IMS) has been proposed for extraction in-line of volatile compounds present in liquid samples. This method has been applied to the analysis of white wine samples from different origins and different alcohol content. In this way, characteristics profiles from each group of wine samples have been obtained. Later a detailed chemometric treatment was carried out to classify the different wine samples. A good classification was obtained by firstly reducing the data dimensionality by PCA followed by LDA and finally using a kNN classifier.

On the other hand, these same wine samples have been analyzed using a chromatographic method using GC–FID. Later a chemometric treatment has been carried out too. Using the data of the areas from all the peaks (36 peaks) from the chromatogram and applying a PCA, all the wine samples have been classified correctly. Moreover LDA and later a kNN classifier have been used to get a good classification too.

Therefore, in this work a vanguard method has been compared (CFS–GPS–UV–IMS) with a rearguard method (GC–FID) obtain satisfactory results with both. Although GC is a more reliable, precise, sensitive and selective technique, its cost is higher compared with UV–IMS equipment, moreover the speed of response obtained using IMS. We recommend the proposed method as a quick, efficient, reliable, cheap and for all of characteristics, CBOs could use IMS as a technique to check the authenticity of their wines in order to prevent frauds.

Acknowledgments

The authors are grateful to the wine makers and CBOs that supplied some of the studied wine samples. They also want to thank Dr. B. Palenzuela for the chromatographic analysis from the wine samples and Dr. S. Garrigues for the construction of the gas-phase separator.

References

[1] M. Valcarcel, S. Cardenas, Anal. Chem. 25 (2005) 67–74.
[2] M.N. Arvanitoyannis, E.P. Katsota, E.H. Psarra, S. Soufleros, Kallithraka, Food Trends Sci. Technol. 10 (1999) 321–336.
[3] J. Lozano, T. Arroyo, J.P. Santos, J.M. Cabellos, M.C. Hornillo, Sensors Actuators B 133 (2008) 180–186.
[4] B. Medina, Food Authentication, Chapman & Hall, London, 1996.
[5] J. Lozano, J.P. Santos, M.C. Hornillo, Talanta 67 (2005) 610–616.
[6] D. Kruzlicova, J. Mocak, B. Balla, J. Petka, M. Farkova, J. Havel, Food Chem. 112 (2009) 1046–1052.
[7] V. Ferreira, P. Fernandez, J. Cacho, Food Sci. Technol. 29 (1996) 251–259.
[8] A. Cuadros-Inostroza, P. Giavalisco, J. Hummel, A. Eckardt, L. Willmitzer, H. Peña-Cortés, Anal. Chem. 82 (2010) 3573–3580.
[9] S.M. Rocha, F. Rodrigues, P. Coutinho, I. Delgadillo, M.A. Coimbra, Anal. Chim. Acta 563 (2004) 257–262.
[10] R. Perestrelo, A. Fernandez, F.F. Albuquerque, J.C. Marques, J.S. Camara, Anal. Chim. Acta 563 (2006) 154–164.
[11] R. Lopez, M. Aznar, J. Cacho, V. Ferreira, J. Chromatogr. A 966 (2002) 167–177.
[12] R. Lopez, N. Ortin, J.P. Perez-Trujillo, J. Cacho, V. Ferreira, J. Agric. Food Chem. 51 (2003) 3419–3425.
[13] M. Gil, J.M. Cabellos, T. Arroyo, M. Prodanov, Anal. Chim. Acta 563 (2006) 145–153.
[14] A. Garcia, M. Aleixandre, J. Gutierrez, M.C. Horrillo, Sensors Actuators B 132 (2006) 911–916.
[15] M. Aleixandre, J. Lozano, J. Gutierrez, I. Sayago, M.J. Fernandez, M.C. Horrillo, Sensors Actuators B 131 (2008) 71–76.
[16] A.Z. Berna, S. Trowell, D. Clifford, W. Cynkar, D. Cozzolino, Anal. Chim. Acta 648 (2009) 146–152.
[17] W. Cynkar, R. Dambergs, P. Smith, D. Cozzolino, Anal. Chim. Acta 660 (2010) 231–277.
[18] M. Peris, L. Esduder-Gilabert, Anal. Chim. Acta 638 (2009) 1–15.
[19] J.A. Ragazzo-Sanchez, P. Chalier, D. Chevalier-Lucia, M. Calderon-Santoyo, Sensors Actuators B 140 (2009) 29–34.
[20] M. Penza, G. Cassano, Food Chem. 86 (2004) 283–296.
[21] L. Arce, M. Menendez, R. Garrido-Delgado, M. Valcarcel, Trends Anal. Chem. 20 (2008) 139–150.
[22] M. Menendez, R. Garrido-Delgado, L. Arce, M. Valcarcel, J. Chromatogr. A 1215 (2008) 8–14.
[23] R. Alonso, V. Rodriguez-Estevez, A. Dominguez-Vidal, M.J. Ayora-Cañada, L. Arce, M. Valcarcel, Talanta 76 (2008) 591–596.
[24] S. Armenta, N.M. Coelho, R. Roda, S. Garrigues, M. de la Guardia, Anal. Chim. Acta 580 (2006) 216–222.
[25] M. McCooye, L. Ding, G.J. Gardner, C.A. Fraser, J. Lam, R.E. Sturgeon, Z. Mester, Anal. Chem. 75 (2003) 2538–2542.
[26] M. McCooye, Z. Mester, Rapid Commun. Mass Spectrom. 20 (2006) 1801–1808.
[27] M. McCooye, B. Kolakowski, J. Boison, Z. Mester, Anal. Chim. Acta 627 (2008) 112–116.
[28] B. Kolakowski, D.A. Agostino, C. Chernier, Z. Mester, Anal. Chem. 79 (2007) 8257–8265.
[29] B. Ells, D.A. Barnett, R.W. Purves, R. Guevremont, J. Environ. Monit. 2 (2000) 393–397.
[30] G.A. Eiceman, L. Garcia-Gonzalez, Y. Wang, B. Pittman, Talanta 39 (1992) 459–467.
[31] W. Cynkar, R. Dambergs, P. Smith, D. Cozzolino, Anal. Chim. Acta 660 (2010) 227–231.
[32] R. Gutierrez-Osuna, IEEE Sensors J. 2 (2002) 189–201.
[33] A. Webb, Stat. Pattern Recogn. (2002).
[34] B. Efron, R. Tibshirani, Stat. Sci. 1 (1986) 54–77.
[35] PLS Toolbox 5.2.2 Eigenvector Research Inc., 2010.
[36] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size consideration in pattern recognition practice in Handbook of Statistics, Amsterdam, The Netherlands, 1982.
[37] R. Garrido-Delgado, S. Lopez-Vidal, L. Arce, M. Valcarcel, J. Sep. Sci. 32 (2009) 3809–3816.
[38] J.M. Jurado, O. Ballesteros, A. Alcázar, F. Pablos, M.J. Martín, J.L. Vilchez, A. Navalón, Anal. Bioananl. Chem. 390 (2008) 961–970.