DEVELOPMENT AND VALIDATION OF MULTIVARIATE QUALITATIVE METHODS IN THE FOOD FIELD

Maria Isabel López Vilardell

Dipòsit Legal: T 893-2015

Development and Validation of Multivariate Qualitative Methods in the Food Field

Ph.D. Thesis presented by

Maria Isabel López Vilardell

Supervisors:

Dr. Itziar Ruisánchez Capelástegui

Prof. María Pilar Callao Lasmarías

Department of Analyticial and Organic Chemistry

UNIVERSITAT ROVIRA I VIRGILI

Tarragona, 2015

UNIVERSITAT ROVIRA I VIRGILI

DEPARTAMENT DE QUIMICA ANALITICA I QUIMICA ORGANICA

C/ Marcel.lí Domingo s/n
Campus Sescelades
43007 Tarragona
Tel. 34 977 55 97 69
Fax 34 977 55 84 46
e-mail: secqaqo@urv.net

WE STATE that the present study, entitled "**Development and validation of multivariate qualitative methods in the food field**", presented by *Maria Isabel López Vilardell* for the award of the degree of Doctor, has been carried out under our supervision at the Department of Analytical Chemistry and Organic Chemistry of this university, and that it fulfils all the requirements to be eligible for the European Doctorate Award.

Tarragona, 27th March 2015

Doctoral Thesis Supervisors:

Dr. Itziar Ruisánchez Capelástegui

Prof. María Pilar Callao Lasmarías

To the great sense of humour of my family,
to the endless patient of my partner and
to my apparent lack of common sense.

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT AND VALIDATION OF MULTIVARIATE QUALITATIVE METHODS IN THE FOOD FIELD
Maria Isabel López Vilardell
Dipòsit Legal: T 893-2015

Acknowledgments.

This thesis wouldn't have been possible without the support from lots of people.

Firstly, I would like to thank my thesis supervisors; Dr. Itziar Ruisánchez and Prof. M. Pilar Callao. I really appreciate the opportunity you gave me when you offered me a place in your group, the time you have given to me, the patience you have had, and all you have taught me.

During the thesis, I was offered the opportunity to stay at the University of Genoa under the supervision of Dr. Paolo Oliveri. Despite the rain, it was a pleasure to spend those months working in Genoa with all of you.

I do not want to forget the members of the Chemometrics, Qualimetrics and Nanosensors group or all my PhD colleges; those who are gone now and those who are still there. I warmly thank you guys!

I wish to thank the people of Girona for the support they gave me. My life would not have been the same without you 'wevins'! We have always had fun! Also the people I have found here in Tarragona deserve a big thank you.

And last but not least, I want to dedicate this thesis to my family and my partner.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. 1. Scope

Qualitative methods have constituted an effective and potential tool over the years in routine laboratories, mainly for its screening potential. They have been applied to different fields such as environmental, clinical or food analysis to deal with quality and safety. These methods are used to address problems that require a binary response (yes/no). In this way, the number of quantitative characterizations, which are usually more expensive, is reduced and decision-making is made quicker by the use of simple and low-cost analytical instrumentation **[1]**.

It is quite usual to use 'screening method' as a synonym for 'qualitative method' despite this may not always be the case. Screening methods generally implies low-time of analysis, permitting a high throughput of samples at low cost, making them suitable for routine analysis. Moreover, they have to demonstrate low false compliance rate (generally < 5%) **[2]**.

Qualitative analysis can be achieved with a specific measurement (univariate) or with multiple non-specific signals used as a fingerprint (multivariate), which, after chemometric treatment, produces the binary response. Spectroscopic data are one of the most common data sources used in multivariate methods. Infrared (IR) and Near-Infrared (NIR) spectroscopies are the most widely used, although Ultraviolet (UV) and Fluorescence spectroscopies are also common.

Other instrumental techniques less common in the analytical field need to be explored since these new spectroscopic data can have a potential impact on multivariate methods. Examples of such techniques are Surface-enhanced Raman Spectroscopy (SERS), the use of which has increased in the field of electronics, and Nuclear Magnetic Resonance (NMR), which is widely used in the omic and clinical fields. Their

analytical applications have been increasing and some studies have focused on both qualitative and quantitative analysis.

A key step in the development of multivariate qualitative methods is the selection of a suitable chemometric treatment. Classification techniques have been widely used to obtain mathematical rules aimed at characterising samples with respect to categorical properties. The techniques can be used, for example, to distinguish between groups of samples according to their designation of origin. The many classification methods that exist can be divided into two main blocks depending on their modelling or discriminant power, as illustrated in Figure 1-1 **[3]**.

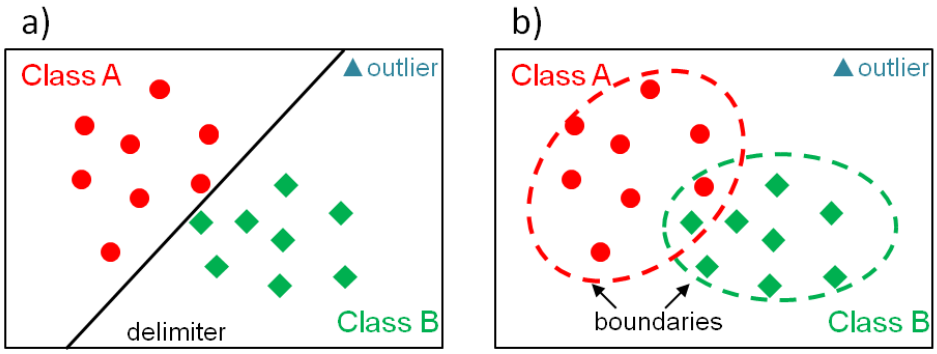

Figure 1-1. Scheme of two class classification using a discriminant (a) and modelling (b) approach.

Discriminant methods divide space into separate regions, each of which corresponds to one class section. At least two classes need to be defined. A wide range of discriminant techniques exist in the chemometric literature, including Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLS-DA).

Discriminant analysis gives excellent results regarding many analytical problems. However, a weakness of this approach is that it is difficult to

deal with outliers or ambiguous samples because all samples must belong to one of the predefined class. Moreover, delimiters have to be reformed if new classes are introduced. Furthermore, some problems, such as those related to food fraud, cannot be tackled from a discriminant perspective since this would require the collection of a representative set of all possible products that can be used to make frauds. As this is rarely feasible in practice, the set is under-representative, which inevitably leads to biased decision rules.

Class-modelling methods, on the other hand, build independent models for each class from the samples belonging to the class to be modelled. This allows a new sample to be assigned to one, more than one, or none of the predefined classes. The main drawback with this approach lies in the complicated interpretation of the results when many classes are predefined. The two most common techniques in this category are Soft Independent Modelling of Class Analogy (SIMCA) and Unequal Dispersed Classes (UNEQ).

Both types of classification techniques are mainly used to address multiclass problems (targeted approaches) in which at least two classes are defined. As stated earlier, when the samples from one of the classes are extremely difficult or impossible to collect, the modelling of just one class (the untargeted approach) is a suitable strategy. Class-modelling methods are the only ones that can tackle these kinds of problems by modelling just one class. However, the application of this strategy is less extended.

An important aspect of the implementation of qualitative methods is the development and application of validation protocols. Validation, a critical step for ensuring reliability in the results, has to be carried out in any analytical application.

Validation protocols are relatively available for univariate qualitative analysis but their multivariate analogues are not so common in routine analysis. One reason for this is that multivariate validation protocols are not so well established in the legislative framework defined in Commission Decision CD/657/EC 2002 **[2]**. Over the last few years, however, several efforts have been made to develop and implement such protocols since they can be used complementary to univariate qualitative analysis.

Multivariate qualitative methods can be applied to numerous disciplines, including medicine, environmental, and food fields. This thesis focuses on food fraud since this is a problem that gravely concerns both the general population and the authorities.

The food industry is a complex business that supplies the world's food demand. Dealing with large quantities of food can endanger food quality and lead to the use of non-regulated food additives, which represents a safety concern **[4,5]**. Such practices, which are largely motivated by financial profits, care not only end consumers but also producers and distributors.

Of the multiple kinds of food fraud, adulteration is gaining interest as an emerging risk, given the increasingly complex and global nature of food supply chains. Food adulteration and contamination involves intentionally adding a non-declared substance to a food in order to improve some of its properties (taste, appearance, length of conservation, etc.) and/or change its composition in order to increase the volume of production. Detecting food adulteration is important for economic reasons but it is especially important when the non-declared substance involves a health risk.

Another key issue in food fraud is food authentication. This type of fraud involves non-compliance with established legislative standards, the implementation of unacceptable processing practices (e.g. freezing) and/or the mislabelling of geographical, botanical or species origin. Due to the increasing complexity and length of the food chain and the increased public sensitivity regarding the origin of food, traceability has become a cornerstone of European Union policy. Similarly, fraudulent practices have forced government institutions to establish regulatory measurements and develop tools to ensure that foods are both of a high quality and safe to be eaten when they reach the consumer.

1. 2. Objectives

The general purpose of this doctoral thesis is to further develop multivariate qualitative methods and its validation. This includes studying and applying several spectroscopic techniques that are generally used with low sample pre-treatment and novel classification approaches in the food field. The experimental developments that shape this overall objective lead to three separate sub-objectives:

1. To evaluate Raman signal in the Surface-enhanced Raman spectroscopy (SERS) modality for its use in multivariate approaches.

Unlike other comparable spectroscopic techniques such as infrared (IR) or near-infrared (NIR) spectroscopy, Raman signal is, despite its possibilities, not so much used in multivariate analysis today. Our research group has already performed some initial work. This thesis extends this work, firstly by characterising the SERS support and secondly by evaluating SERS as a potential data source for its use in multivariate qualitative analysis.

2. To develop multivariate qualitative methods based on untargeted modelling.

Most studies based on multivariate qualitative analysis have focused on the multiclass approach (targeted) rather than the one-class approach (untargeted) when dealing with analytical problems. This doctoral thesis exploits the significant possibilities of the one-class approach to address food fraud problems. In order to accomplish it, well-established class-modelling techniques such as SIMCA and UNEQ have been used in addition to a newly developed one (partial least square density modelling, PLS-DM).

3. To establish validation protocols for multivariate qualitative analysis in accordance with the indications of the European directives and to calculate the performance parameters associated with them.

The implementation of multivariate qualitative methods is constantly growing. However, there is no established worldwide criterion for their validation. This thesis applies generic univariate concepts to the multivariate environment and presents a way to establish performance parameters with quantitative connotation such as the *unreliability region* and limits related to concentration.

1. 3. Structure of the thesis

This thesis is divided into five chapters: Chapter 1 the introduction, Chapter 2 to 4 the core of it and Chapter 5 the conclusions. The scientific contributions are presented in the three main chapters (Chapters 2 to 4), all three with an introduction and results section:

- ✓ Introduction: It includes the ***state-of-the-art*** subsection, in which a recent bibliographic research has been carried out. Moreover, those theoretical fundamentals that the authors consider not enough explained or established in the papers but relevant for a proper understanding of the work are presented in the ***background*** subsection.

- ✓ Results: It is presented in published papers format. These papers have been edited to provide the thesis a uniform format.

Given the thesis' structure, including a theoretical chapter can be seen as redundant. Thus, this chapter has been omitted to make the reading more bearable.

Chapter 1: Introduction. This chapter introduces the general scope of the thesis, defines its objectives and details its structure.

Chapter 2: Analytical study of SERS spectroscopy. This chapter is related to the first objective and assesses the use of Surface-enhanced Raman spectroscopy (SERS) as data source for multivariate analysis. The *state-of-the-art* includes a recent overview of the application fields of SERS and its use in multivariate analysis, in addition to some review of the studied analyte: Sudan dye. The *background* subsection gives theoretical fundamentals about Raman spectroscopy and Sudan dyes structure. The results are presented by the published paper; *M.I. López, I. Ruisánchez, M.P. Callao, Spectrochim. Acta. A., 111 (2013) 237–241*.

Chapter 3: Class-modelling approach for adulteration & authentication. This chapter is related to the second objective and describes some food applications of multivariate qualitative analysis. It is mainly focused on untargeted modelling. The *state-of-the-art* reviews the use of discriminate versus class-modelling approach to finally focus on

the one-class modelling approach. The *background* subsection includes a comparative study of the class-modelling techniques used in the chapter. The results are presented by the two published papers; *M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, Food Chemistry, 147 (2014) 177-181* and *P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, S. Lanteri, Anal. Chim. Acta, 851 (2014) 30-36*.

Chapter 4: Validation protocols for multivariate qualitative analysis. This chapter is related to the third objective and proposes a validation methodology for multivariate qualitative applications. The *state-of-the-art* presents the actual validation scenario for analytical methodologies, including qualitative and quantitative analysis as well as univariate and multivariate approaches. The *background* subsection is presented as a tutorial paper, which includes an overview of method validation applied to qualitative analysis, including both univariate and multivariate analysis; *M.I. López, M.P. Callao, I. Ruisánchez, Anal. Chim. Acta, Submitted*. The results are presented by the published paper; *M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, Anal. Chim. Acta, 824 (2014) 28–33*.

Chapter 5: General conclusions. This chapter contains the general conclusions of the thesis.

Appendix. This part of the thesis contains: (A) summary of main abbreviations used in this thesis, (B) list of papers presented by the author in this thesis, (C) contributions made to international meetings attended during this period and (D) research stay and training courses attended during this period.

1. 4. References

[1] R. Muñoz-Olivas, *TrAC Trends Anal. Chem.* **2004**, *23*, 203.

[2] C. R. (EC) No. 2002/657/EC, *Off. J. Eur. Communities* **2002**, 8.

[3] R. G. Brereton, *Chemometrics for Pattern Recognition*, **2009**.

[4] D. Cozzolino, E. Corbella, H. E. Smyth, *Appl. Spectrosc. Rev.* **2011**, *46*, 523.

[5] G. Vinci, R. Preti, A. Tieri, S. Vieri, *J. Sci. Food Agric.* **2013**, *93*, 439.

CHAPTER 2

ANALYTICAL STUDY OF SERS SPECTROSCOPY

2. 1. Introduction

2.1.1. *State-of-the-art*

Raman Spectroscopy

Since the *Surface Enhanced Raman Spectroscopy (SERS)* was discovery several decades ago it has experienced substantial changes, ranging from theoretical and technical developments to practical analytical applications **[1]**. Figure 2-1 shows the percentage of recent publications (since 2003) that have used SERS, grouped in different application fields. 'Not specified' category corresponds to those results which were found in the chemistry field but was neither included in univariate not multivariate after adding the corresponding specific keywords for each category.

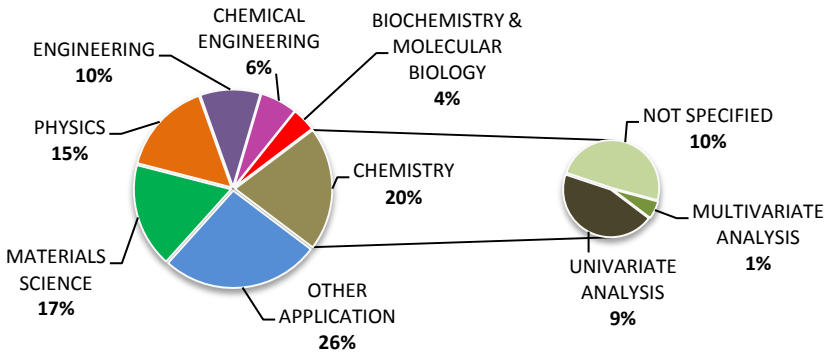

Figure 2-1. Percentage of scientific publications (between 2003 and mid-2014) obtained by searching the scientific database *ISI Web of Knowledge (WoK)* and using 'SERS' or 'Surface Enhanced Raman Spectroscopy' as keywords.

This pie chart indicates a modest percentage of published work conducted in chemistry and biochemistry fields but only some of them

address analytical purposes. Thus, Raman spectroscopy can be still considered an emerging technique in the analytical field. In comparison with univariate analysis, multivariate analysis constitutes just a small proportion of those analytical applications and Table 2-1 lists some recent works, including examples of quantitative and qualitative analysis.

Table 2-1. Examples of applications of SERS in multivariate analysis using different chemometric techniques.

Study	Chemometric technique	Analyte	Sample	Reference
Quantification	Multivariate curve resolution (MCR)	Antibiotics	Urine	**[2]**
Quantification	Partial least-squares regression (PLS), artificial neural networks (ANNs), and support vector regression (SVR)	Sudan	Food	**[3]**
Classification	Soft independent modelling of class analogy (SIMCA)	Viruses	Food/ Water	**[4]**
Optimization	Experimental design (factorial design)	-	SERS roughness	**[5]**
Unsupervised trace detection	Principal component analysis (PCA)	Chemicals	Explosives	**[6]**
		Sudan	Food	**[7]**

Sudan dyes

The analytical problem to which SERS is applied follows a previous study conducted by our research group that was based on the detection of Sudan dyes, a potential adulterant commonly used in spices **[7]**.

Since the reporting of the first adulteration case in 2003, interest has increased in the development and application of methods for detecting and analysing Sudan dyes and their derivates in foodstuffs as well as in the study of their effect on human health. In 2005 the European Food Safety Authority (EFSA) reviewed the toxicology of numerous dyes found illegally in foods in the European Union (EU). Particularly with regard to Sudan I, there is strong evidence both of genotoxicity and carcinogenicity. Because of structural similarities between Sudan I and the other Sudan dyes, the latter ones are presumed to have the same deleterious effects **[8,9]**. The EU therefore issued Decision 2005/402/EC, which requires chilli, curcuma and palm oil products imported into Europe to be tested for Sudan dyes **[10]**.

The pie chart in Figure 2-2 shows the percentage of scientific publications (since 2003), classified in different fields, using “Sudan dye” as keyword.

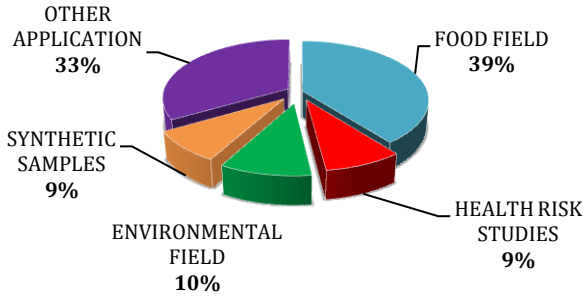

Figure 2-2. Percentage of scientific publications (between 2003 and mid-2014) obtained by searching the scientific database *ISI Web of Knowledge (WoK)*.

Several methods have been proposed for determining the presence of Sudan dyes and quantifying them. Many of these methods are based on chromatographic techniques with different detection systems and pre-

extraction methods **[11–14]**. Advances have also been made in the direct analysis of Sudan dyes based on surface-sensitive techniques such as SERS, Electrochemistry, or Quenched Fluorescence methods, where the analytical response is sensitive to the surface or support used **[15–17]**.

2.1.2. *Background*

Raman Spectroscopy

Raman spectroscopy is based on the inelastic scattering, or Raman scattering, of monochromatic light, usually from a laser source. Vibrational, rotational, and other low-frequency changes caused by laser radiation therefore provide a fingerprint of the sample under study. This technique is non-destructive, relatively simple to perform and requires no special sample preparation. However, as its signal is normally too weak, later research has been conducted to improve it.

Several Raman modalities are available. These include *Confocal Raman Microscopy*, which improves space resolution, and *Resonance Raman (RR) spectroscopy*, which has been extensively used in the field of bioinorganic chemistry to elucidate the coordination environment of metals. This chapter focuses on *Surface Enhanced Raman Spectroscopy (SERS)*, which deals with molecules that are adsorbed onto metal surfaces or colloids that, under certain conditions, greatly enhance the Raman signal, thus increasing sensitivity.

The modality used in this thesis is a combination of Confocal Raman Microscopy with the signal enhancement of the SERS substrate, shown in Figure 2-3.

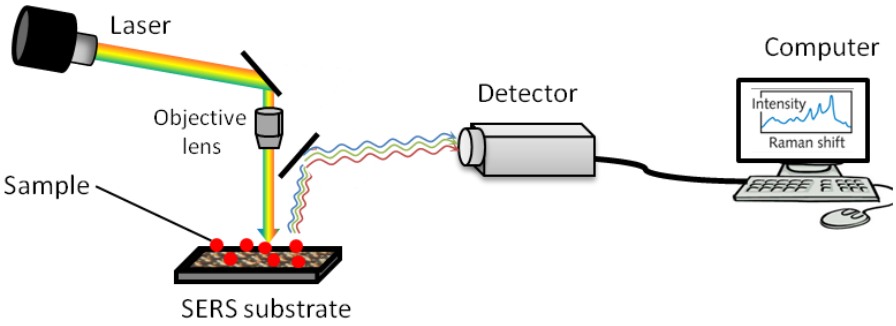

Figure 2-3. Schematic drawing of a Confocal Raman Microscopy combined with SERS system.

Sudan dyes

Sudan dyes belong to a family of synthetic azo orange-red dyes that are used to give colour to industrial materials such as plastics, oils, and waxes. The low price of Sudan dyes leads to them being illegally used to enhance and maintain the colour of certain foods, especially spices. Figure 2-4 shows the structure of four commonly used Sudan dyes.

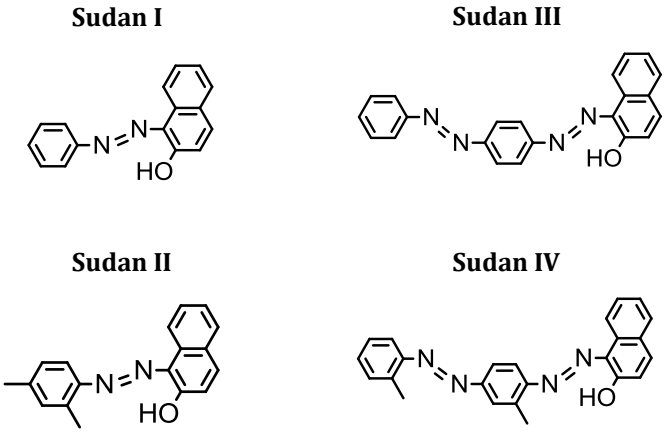

Figure 2-4. Chemical structure of Sudan I, II, III, IV dyes.

The common structure motif in the family of Sudan dyes is represented by Sudan I which consists in a beta-napthol ring linked by an azo group in the alfa position to a phenyl ring. Methylation in one *ortho*-position and the *para*- position of the phenyl ring produces Sudan II.

On the other hand, Sudan III is directly related to Sudan I, the phenyl group bears and additional azo functionality in the *para*-position which connects it to another phenyl group. Finally, Sudan IV preserves the backbone from Sudan III but methylation occurs in one *ortho*- position of each phenyl ring.

2. 2. Results

FIGURES OF MERIT OF A SERS METHOD FOR SUDAN I DETERMINATION AT TRACES LEVELS

M.I. López, I. Ruisánchez, M.P. Callao

Spectrochim. Acta. A., 111 (**2013**) 237–241.

Spectrochim. Acta. A., 111 (2013) 237–241

Figures of merit of a SERS method for Sudan I determination at traces levels

M. Isabel López, Itziar Ruisánchez, M. Pilar Callao

Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovia i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

ABSTRACT

A substrate for Surface-Enhanced Raman Scattering spectroscopy (SERS), electropolished Al, is proposed as a tool for a rapid and low cost determination of Sudan I. This dye has been used as an additive in some foodstuffs but it is now banned because of the health risk associated with its carcinogenic and mutagenic properties. Despite the presence of fluorescence, Raman spectra of Sudan I can be obtained using excitation lasers at 633 and 785 nm. To get rid of the spectral noise and fluorescence background, Savitzky-Golay smoothing and polynomial corrections were applied, respectively. The Raman signal was proved to be enhanced. A linear dependence was found between the logarithmic intensity at 1598 cm^{-1} peak versus the logarithmic concentration. The figures of merit were studied obtaining high sensitivity and low detection limits (10^{-7} M). A multivariate exploratory analysis (PCA) was used to study the ability of SERS to distinguish Sudan I from other similar compounds. Therefore, results show that SERS is a potential tool to determine Sudan I quickly and effectively.

***Keywords**: SERS, Sudan I determination, Figures of merit, Raman spectroscopy*

1. Introduction

Since Surface-Enhanced Raman Spectroscopy (SERS) was discovered by Fleischmann et al. in 1974 [1], it has been increasingly used as a powerful method for the fast detection and direct analysis of chemicals [2-5]. SERS spectroscopy detects lower analytical concentrations than conventional Raman because it significantly amplifies Raman effects by

several orders of magnitude when analyte molecules are adsorbed onto nanoparticles [6] or metal surfaces with nanoscale roughness [7]. The enhancement effect is assumed by two generally accepted mechanisms: *chemical enhancement*, which is the transference of charge between the adsorbed molecules and the metal substrate; and *electromagnetic enhancement*, considered as the main effect, which is associated with the large local field enhancement close to metallic surfaces when localized surface plasmon resonances are excited by the laser [8-10].

Various methods have been used to obtain a SERS substrate that can provide a huge enhancement. The preparation of metal colloid solutions, predominantly made of Ag and Au, are well established [11-13]. Also, rough metal surfaces prepared by chemical etching or electrochemical treatment, are used as SERS active substrates. The later, give irregular structures but they are easy to produce with relatively good Raman signal [8, 9].

The SERS applications cover current areas in science, such as physical chemistry, analytical chemistry, biomedicine and materials. In fact, most current investigation focuses on theoretical and technological development [8, 14, 15] and application in bioscience [5, 16], but few studies have been made in analytical chemistry. Some of them use qualitative approaches [3, 17-21] whereas others are based on a multivariate quantitative analysis [22, 23].

Our group has recently proposed SERS as screening tool for detecting Sudan I (1-phenylazo-2-naphthol) in culinary spices [3]. Sudan I is a synthetic azo orange-red dye, used as a colouring agent in such commercial applications. Unfortunately, in some countries, it has been used for commercial benefits as an additive to improve the colour of some foodstuffs. It has been considered as genotoxic [24] and classified as class three carcinogen by the International Agency for Research on Cancer (IARC) [25].

Spectrochim. Acta. A., 111 (2013) 237–241

The previous work was focused in the qualitative detection of the Sudan I and the goal of the present work is to develop the quantitative methodology. Two excitations lasers were proved and for both the same protocol have been established. Firstly, a pre-treatment of the spectra was done to get rid of the fluorescence and spectral noise. Secondly, the most notable analytical figures of merit such as reproducibility, sensitivity, limit of detection, and selectivity in front of other Sudan together with the enhancement factor (EF), was studied in this work for the quick determination of Sudan I at low concentration levels. All parameters have been established with two different excitation lasers. We consider that the establishment of figures of merit is required for SERS method to become a current method for trace analysis.

2. Experimental

2.1 Al substrate

Aluminium (Al) foils from Goodfellow (99.999% Al) were electropolished under 20 V in a 1:4 (v:v) mixture of perchloric acid ($HClO_4$) 60 wt% and ethanol (EtOH) for 4 min at 5ºC.

2.2 Reagents and samples

Sudan I standard was purchased from ACROS (Geel, Belgium) and the other Sudan dyes were purchased from SIGMA (St. Louis, MO, USA). HPLC grade chloroform was provided by SDS (Carlo Erba Reagents SDS S.A., Spain).

A set of stock solution was prepared within the concentration range of $4.0x10^{-4}$ M to $2.01x10^{-6}$ M. To obtain the SERS spectra, 5µL of each solution was dropped onto the aluminium SERS substrate and dried at 30ºC for 15 minutes to get rid of the solvent completely. For the selectivity study, a solution of other Sudan dyes was prepared in the same way as for Sudan I.

The Al support can be reused numerous times. The analyte was removed by washing with $CHCl_3$ drops followed by a chloroform bath for half an hour.

2.3 SERS spectra collection

SERS spectra were collected using a Renishaw in Via Reflex Raman confocal microscope (Gloucestershire, UK), equipped with an Ar-ion laser at 514 nm, an He-Ne laser at 633 nm, a diode laser emitting at 785 nm and a Peltier-cooled CCD detector (-70ºC) coupled to a Leica DM-2500 microscope. Calibration was carried out daily by recording the Raman spectrum of an internal Si standard. Spectra were recorded with the accumulation of three scans of ten seconds each one and using a 50x low working distance to focus the laser light on the samples. The Raman data were the averaged value of five different points on the surface.

2.4 Data treatment

As a spectral pre-processing, vertical offset correction and Savitzky-Golay smoothing **[26]** was applied to all data sets to suppress the instrumental noise. Then, the presence of fluorescence was removed by polynomial baseline correction. For selectivity, principal component analysis (PCA) **[27]** was used as an exploratory analysis tool. All data treatment was carried out using MATLAB 6.5 (The MathWorks, MA, USA).

3. Methodology

3.1 Preliminary SERS studies

For *excitation laser selection*, some working lasers were checked to obtain a well defined Raman spectrum. Then, *reproducibility* was studied

by measuring a sample of Sudan I at several random places on the SERS substrate. Moreover, substrate-to-substrate reproducibility was studied by comparing SERS spectra obtained from two supports.

3.2 SERS performance

The *enhancement factor (EF)* is the parameter used to assess the SERS performance. It shows the capacity of a SERS substrate to amplify the Raman signals with respect to a non-SERS substrate. This magnitude (EF) was estimated according to the following equation **[10]**:

$$EF = \frac{I_{SERS} \cdot N_{Ref}}{I_{Ref} \cdot N_{SERS}} \qquad Eq.\ 1$$

where I_{SERS} and I_{Ref} are the Raman intensities of the same band under SERS and reference Raman conditions, respectively. N_{SERS} and N_{Ref} are the number of Sudan I molecules illuminated by the laser focus spot under SERS and reference Raman conditions, respectively.

3.3 Intensity and concentration relationship

The univariate linear dependence of SERS intensity over Sudan I concentration was studied at a robust peak with high intensity. Since a good relationship between both parameters was not possible, a data transformation to a logarithmic relationship was carried out as it is described in the following equation **[17]**:

$$\log I = b_o + b_1 \cdot \log C \qquad Eq.2$$

where I is the intensity, C is the concentration and b_0 and b_1 are the regression coefficients.

3.4 Figures of merit

Sensitivity can be defined as the variation of intensity caused by the variation in the concentration. Thus, sensitivity (ΔI/Δc) was assessed considering the increase in intensity obtained at two Sudan I concentration levels relating to this increment of concentration.

Limit of Detection (LOD) is the lowest analyte concentration likely to be reliably distinguished. Traditionally, the LOD concentration has been calculated measuring replicates of a blank sample from which a limit signal is obtained by Eq. 3 that is finally converted in concentration (LOD) by means of linear regression.

$$I_{lim} = \bar{I}_{blank} + 3 \cdot SD_{blank} \qquad Eq.\ 3$$

where I_{lim} is the limit intensity, above which is considered there is analyte, $\bar{I}_{blank}$ is the mean value of blank replicates at a specific intensity and SD_{blank} is the standard deviation.

In this study, a sample which contains low concentration of Sudan I giving not recognized Raman spectrum was used instead of blank. Then, the LOD was calculated from the log-log regression line established between intensity and concentration described in section 3.3 under Methodology (Eq. 2).

Selectivity can be defined as the ability of SERS to distinguish Sudan I from other similar compounds. It was carried out by a multivariate exploratory analysis (PCA) with the spectra of the Sudan I and the ones of potential interferers, all measured in the same spectroscopic conditions.

4. Results and Discussion

4.1. Preliminary SERS studies

First, a test was carried out to evaluate if a spectra could be obtained with available lasers (514 nm, 633 nm and 785 nm). The results are shown in Fig. 1. The spectra recoded at 633 nm and at 785 nm showed good signals although fluorescence was present. Since a lack of spectrum was obtained using the excitation laser at 514 nm, this laser was subsequently kept out of the following experimentation.

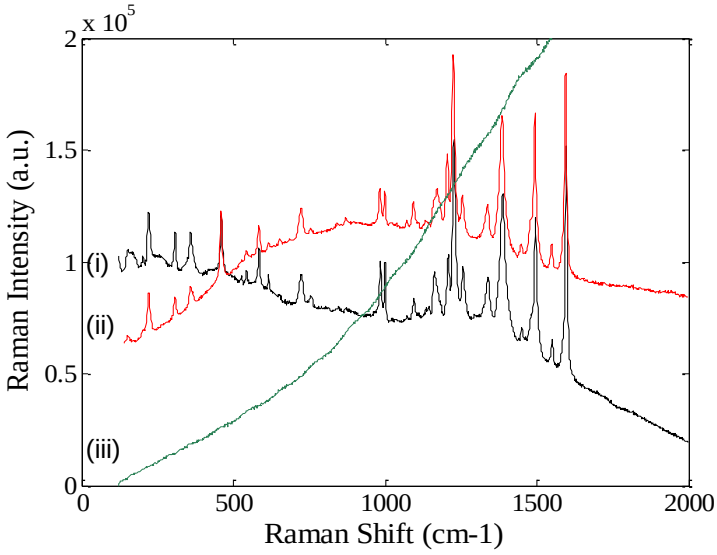

Fig. 1. *Row SERS spectra of Sudan I obtained at $4x10^{-4}$ M using 785 nm (i), 633 nm (ii) and 514 nm (iii) as the excitation laser.*

As a second step, the reproducibility was studied by measuring the spectra of Sudan I at five different random points of the substrate. As there was some vertical displacement, the raw spectra were corrected by forcing all the spectra to end at the same point. This was done by subtracting the minimum intensity at 2000 cm^{-1}. Fig. 2(i) shows the

baseline-corrected spectra obtained working at 633 nm and it can be seen that the main variations are due to the effect of fluorescence.

To get rid of the florescence, some corrections have to be made to the spectra. The raw spectra shown in Fig. 1 (recorded with both lasers) are both affected by fluorescence, as expected. The proposed correction was based on fitting a polynomial function through specified points (marked with arrows in Fig. 2), selected because no signal should be present. Once the polynomial had been adjusted, it was subtracted from the spectra. In our particular case, a third-order polynomial was the optimal choice. The corrected spectra are shown in Fig. 2(ii) where it is possible to see that the florescence background was removed and that the Raman peaks were not significantly altered.

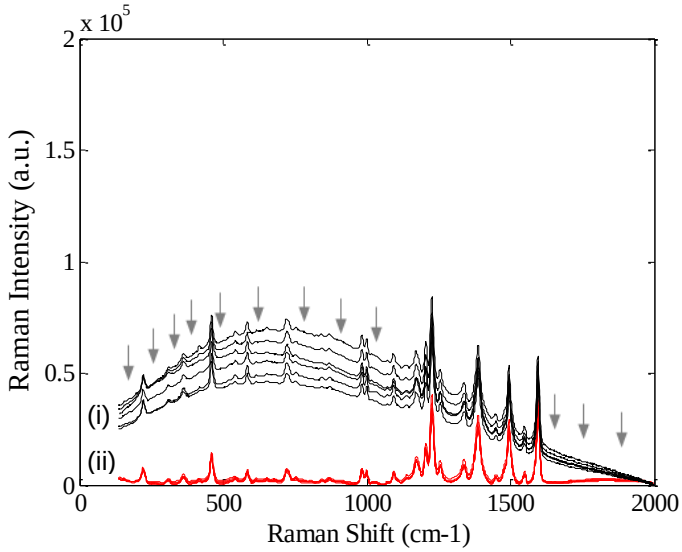

Fig. 2. *SERS data (i) and baseline corrected data (ii) of Sudan I at $4x10^{-4}$ M obtained at five randomly selected places with 633 nm excitation laser.*

Similar results (not shown) were obtained with the 785 nm excitation laser, although the resulting polynomial is different due to the different fluorescence behaviour.

To estimate the reproducibility, standard deviations (SDs) for raw and corrected spectra were computed. Fig. 3 shows the SD values along the Raman shift on the same scale as the raw spectrum for better understanding of the real variation. Moreover, for the sake of clarity, the figure also includes an amplification showing that SD followed the same shape of spectrum and that major variability was found between 1200 and 1700 cm^{-1} were the major peaks appeared while it is really low between 200 and 1000 cm^{-1}. In particular, the magnitude of SD was greater for uncorrected spectrum since it was affected by fluorescence.

The relative SD of the major spectral peaks were lower than 5%. Therefore, the reproducibility of the Al substrate was acceptable. The results for the substrate-to-substrate reproducibility were similar.

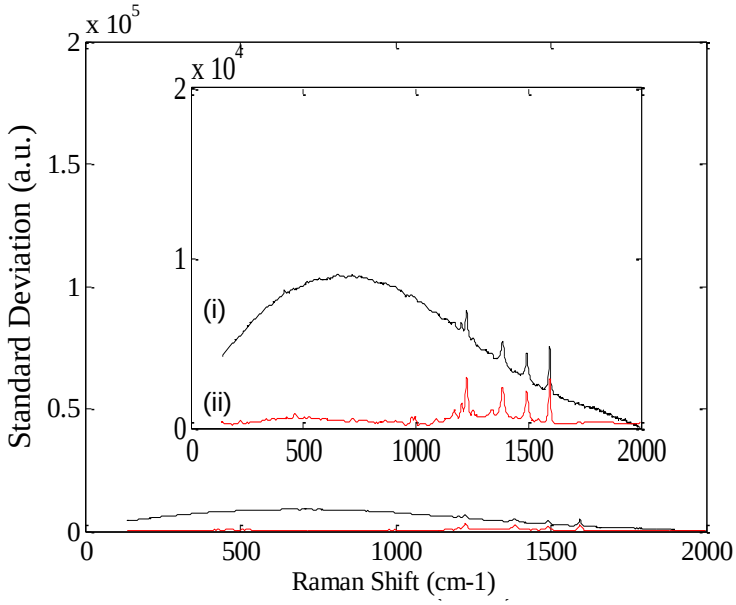

Fig. 3. *Standard deviation (SD) and SERS data (i) and baseline corrected data (ii) of Sudan I at 4x10^{-4} M obtained with 633 nm excitation laser.*

4.2 SERS performance

The intensity used to estimate the enhancement factor (section 3.2 under Methodology, Eq. 1) was the peak intensity measured at 1598 cm^{-1} since it is a robust peak with high intensity. The reference condition was the corresponding Raman spectra of Sudan I on a glass substrate. Assuming that molecules are uniformly dispersed, the *N* values can be calculated by the following equation:

$$N = \frac{V \cdot C \cdot 6.02 \cdot 10^{23}}{A} \cdot Spot \qquad Eq.4$$

where *V* is the volume deposited on the SERS or reference support which was 5 µL. *C* is the concentration, $4x10^{-3}$ M was used for reference whereas the SERS concentration was $4x10^{-5}$ M. *A* is the areas of dispersion, which were 5 mm^2 and 7 mm^2 for reference and SERS, respectively. Finally, *Spot* is the area lit up by the laser, which is related to the wavelength of the laser and the numerical aperture of the microscope in the following way; 1.22·λ/NA. The objective used in this study was 0.75 NA./50x.

Table 1 summarizes the EF together with its respective errors obtained for the Al support at 633 nm and 785 nm laser excitation. In general terms, SERS was enhanced by a power of three and four order of magnitude at 633 nm and 785 nm, respectively. These EF values could be considered as normal for this kind of supports **[9]**.

4.3 Intensity and concentration relationship

The dependence of SERS spectral intensities over concentration was studied at low concentration, taking five repetitions of each standard. The logarithmic relationship (Eq. 2) was established between both parameters considering the intensity values at 1598 cm^{-1} peak because of a good univariant linear regression was not possible.

Spectrochim. Acta. A., 111 (2013) 237–241

Table 1. *Analytical characteristics of SERS support with 633 nm and 785 nm laser excitation.*

	633 nm	**785 nm**
EF	$6x10^3 \pm 2x10^3$	$1x10^4 \pm 9x10^3$
Working range (M)	$2.0x10^{-6}$ - $4.0x10^{-5}$	$4.0x10^{-6}$ - $8.0x10^{-5}$
Slope	1.09	0.95
Intercept	9.29	8.43
Correlation, r	0.8648	0.9405
LOD (M)	$2.6x10^{-7}$	$9.4x10^{-8}$
Sensitivity(ΔI/Δc)	$5.7x10^8$	$4.0x10^8$

The results for both lasers are shown in Table 1. The second row shows the specific concentration range used in this study, the one at 633 nm being slightly lower than the one at 785 nm. As far as the regression parameters are concerned, there are no significant differences between the data provided by the two lasers. Both regression models have acceptable correlations, as is expected for Raman measurement, but the results obtained at 785 nm are slightly better.

4.4 Figures of merit

The sensitivity (ΔI/Δc) was assessed considering the decrease in intensity obtained when the concentration was decreased from $4.0x10^{-5}$ M to $4.0x10^{-6}$ M. As Table 1 shows, sensitivities were comparable for both lasers.

The LOD was calculated from the established regression line explained in section 3.4 under Methodology. The limit intensity signal to be converted into a concentration is the intensity at 1598 cm^{-1} peak plus three times the standard deviation corresponding to the Sudan I sample at $1x10^{-7}$ M which presents not distinguishable spectrum (Fig. 4). Table 1 shows the LOD results for both studied lasers, sensitivities were high and detection limits were low being the LOD at 785 nm slightly lower.

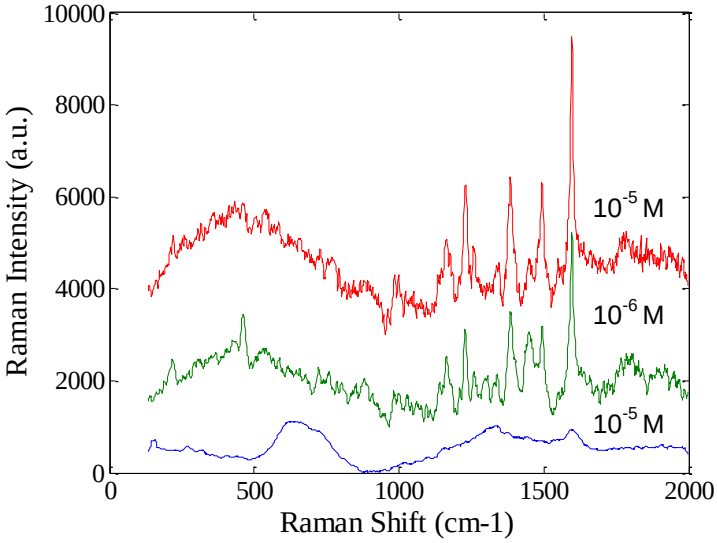

Fig. 4. *Spectra of Sudan I at different concentration levels. Laser wavelength 633 nm.*

Selectivity was assessed by collecting five spectra of Sudan II, III and IV dyes in the same spectroscopic conditions as Sudan I at 785 nm. Fig. 5 shows the spectra for all four Sudan dyes along with their respective molecular structure. In their structure, Sudan III and IV have additional azo and benzene groups. On the other hand, Sudan II and IV have two methyl groups more than Sudan I and III. It can be seen that the spectra of the four Sudan dyes have some similarities because part of the structure is the same. However, no specific peak can be assigned for each of the four Sudan dyes analysed.

The amount of 20 samples have been analysed, 5 for each Sudan dye. An exploratory analysis based on PCA was made to visualize the distribution of the samples after the spectra were mean centred. Fig. 6 shows the score 3D plot of the third principal components (PC) which accounted for over 84% of the total variance of the data. It can be seen

Spectrochim. Acta. A., 111 (2013) 237–241

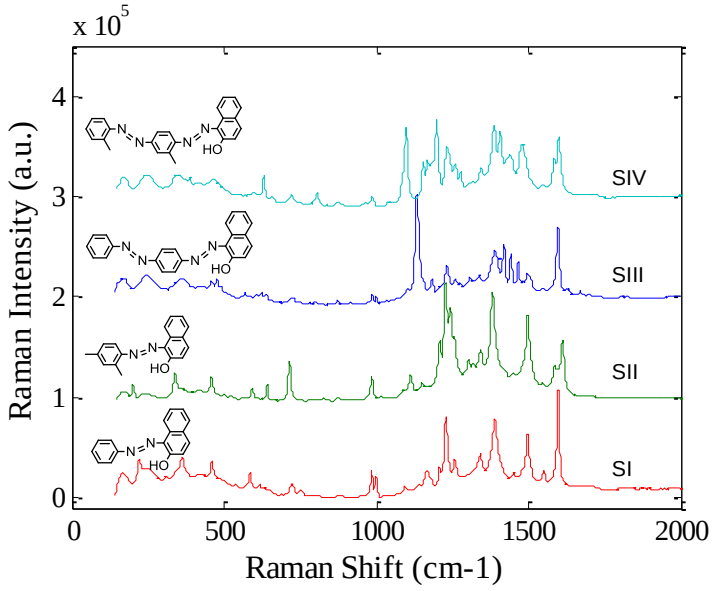

Fig. 5. *SERS spectra of Sudan I, II, III and IV together with the corresponding molecular structures. Laser wavelength 785 nm.*

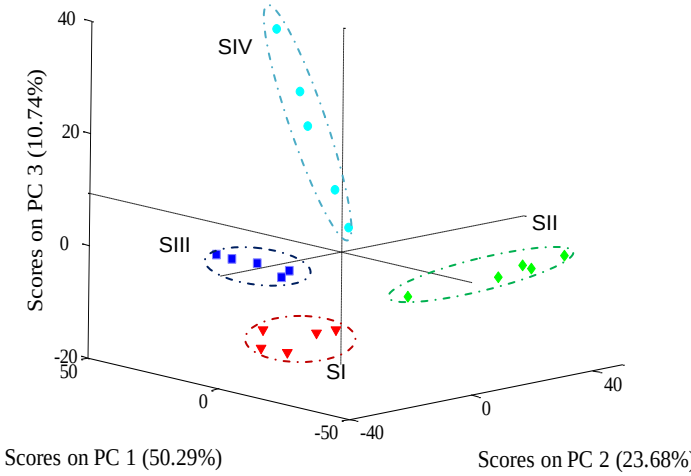

Fig. 6. *PCA score plot from SERS spectra acquired from four Sudan. The ellipses are drawn as a guide and have no statistical meaning.*

from this plot that the first PC (50.29% of total variance) was not enough to distinguish between dyes. However, when the second PC was included (23.68% of total variance), Sudan II could be distinguished from the other dyes. Moreover, the other dyes tended to group together. Finally, when the third PC was included (10% of total variance) Sudan IV could be differentiated from the other dyes. To differentiate between clusters of Sudan I and Sudan III, the three PCs had to be considered.

On the basis of these results, it seems that SERS is selective enough to detect Sudan I in matrixes that contain the studied interferences.

5. Conclusions

Analytical methods with electropolished aluminium as a SERS substrate and excitation lasers of 633 and 785 nm had been established for Sudan I determination.

The Raman signals were significantly enhanced (around an enhancement factor of 10^4) at both 633 and 785 nm lasers. The enhancement was slightly higher with 785 nm laser excitation. Fluorescence was removed by polynomial baseline correction. Repeated measures on the same substrate and different substrates showed that reproducibility was good.

Our work has also demonstrated that a linear relation can be established between the logarithmic intensity versus the logarithmic concentration for the vibrational band at 1598 cm^{-1}. Low detection limits around $3x10^{-7}$ M can be obtained.

The study of selectivity against potential interferences (substances of similar molecular structure) by multivariate analysis (PCA) showed that the technique can be used to detect Sudan I in the presence of interferences.

These results together with the recent implementation of portable Raman systems open up the way to quick and low-cost detection of Sudan I using a simple SERS substrate.

6. Acknowledgments

This work was supported by Generalitat de Catalunya under project 2009 SGR 549. M.I López thanks the Rovira i Virgili University for providing the doctoral fellowship (2011BRDI-06-09). We would like to thank the Nanoelectronic and Photonic Systems Group from the Rovira i Virgili University for providing the SERS support.

7. References

[1] M. Fleischmann, P.J. Hendra, A.J. McQuillan, *Chemical Physics Letters*. 26 (1974) 163-166.

[2] L. Guerrini, S. Sanchez-Cortes, V.L. Cruz, S. Martinez, S. Ristori, A. Feis, *J. Raman Spectrosc*. 42 (2011) 980-985.

[3] C.V. Di Anibal, L.F. Marsal, M.P. Callao, I. Ruisánchez, *Spectrochim. Acta Part A*. 87 (2012) 135-141.

[4] H. Tang, D. Fang, Q. Li, P. Cao, J. Geng, T. Sui, X. Wang, J. Iqbal, Y. Du, *J. Food Sci*. 77 (2012) T105-T109.

[5] V.P. Drachev, M.D. Thoreson, V. Nashine, E.N. Khaliullin, D. Ben-Amotz, V.J. Davisson, V.M. Shalaev, *J. Raman Spectrosc.* 36 (2005) 648-656.

[6] S. Miljanić, A. Dijanošić, Z. Meić, *Spectrochim. Acta Part A*. 75 (2010) 1008-1012.

[7] K.G.M. Laurier, M. Poets, F. Vermoortele, G.D. Cremer, J.A. Martens, H. Uji-I, D.E. De Vos, J. Hofkens, M.B.J. Roeffaers, *Chem. Commun.* 48 (2012) 1559-1561.

[8] E. Le Ru, P. Etchegoin, *Principles of Surface-Enhanced Raman Spectroscopy: and related plasmonic effects*, Elsevier Science, 2008.

[9] R.J.C. Brown, M.J.T. Milton, *J. Raman Spectrosc.* 39 (2008) 1313-1326.

[10] P.L. Stiles, J.A. Dieringer, N.C. Shah, R.P. Van Duyne, *Annu. Rev. Anal. Chem.* 1 (2008) 601-626.

[11] S. Mohapatra, S. Siddhanta, D.R. Kumar, C. Narayana, T.K. Maji, *Eur. J. Inorg. Chem*. 2010 (2010) 4969-4974.

[12] T. Huang, F. Meng, L. Qi, *J. Phys. Chem. C.* 113 (2009) 13636-13642.

[13] P.N. Sisco, C.J. Murphy, *J Phys Chem A.* 113 (2009) 3973-3978.

[14] Y. Fang, H. Yang, P. Jiang, D.D. Dlott, *J. Raman Spectrosc.* 43 (2012) 389-395.

[15] N. Guillot, M.L. de la Chapelle, *J. Quant. Spectrosc. Radiat. Transf.* 113 (2012) 2321-2333.

[16] X.X. Han, Y. Ozaki, B. Zhao, *TrAC, Trends Analyt. Chem.* 38 (2012) 67-78.

[17] C. Zhu, G. Meng, Q. Huang, Z. Huang, *J. Hazard. Mater.* 211–212 (2012) 389-395.

[18] X. Zhou, Y. Fang, P. Zhang, *Spectrochim. Acta Part A.* 67 (2007) 122-124.

[19] M.A. De Jesús, K.S. Giesfeldt, M.J. Sepaniak, *J. Raman Spectrosc.* 35 (2004) 895-904.

[20] J.D. Driskell, S. Shanmukh, Y.-. Liu, S. Hennigan, L. Jones, Y.-. Zhao, R.A. Dluhy, D.C. Krause, R.A. Tripp, *IEEE Sensors J.* 8 (2008) 863-870.

[21] S.J. Clarke, R.E. Littleford, W.E. Smith, R. Goodacre, *Analyst* 130 (2005) 1019-1026.

[22] A. Lorén, C. Eliasson, M. Josefson, K.V.G.K. Murty, M. Käll, J. Abrahamsson, K. Abrahamsson, *J. Raman Spectrosc.* 32 (2001) 971-974.

[23] W. Cheung, I.T. Shadi, Y. Xu, R. Goodacre, *J. Phys. Chem. C.* 114 (2010) 7285-7290.

[24] M. Stiborová, V. Martínek, H. Rýdlová, P. Hodek, E. Frei, *Cancer Res.* 62 (2002) 5678-5684.

[25] International Agency for Research on Cancer (Ed.), *Monographs in the evaluation of carcinogenic risk of chemicals to man.* Vol 8. Some aromatic azo compounds, Lyon, France, 1975.

[26] A. Savitzky, M.J.E. Golay, *Anal. Chem.* 36 (1964) 1627-1639.

[27] R.G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, Wiley, 2003.

2. 3. References

[1] C. L. Haynes, A. D. McFarland, R. P. Van Duyne, *Anal. Chem.* **2005**, *77*.

[2] M. B. Mamián-López, R. J. Poppi, *Anal. Bioanal. Chem.* **2013**, *405*, 7671.

[3] W. Cheung, I. T. Shadi, Y. Xu, R. Goodacre, *J. Phys. Chem. C* **2010**, *114*, 7285.

[4] C. Fan, Z. Hu, L. K. Riley, G. A. Purdy, A. Mustapha, M. Lin, *J. Food Sci.* **2010**, *75*, M302.

[5] R. M. Jarvis, W. Rowe, N. R. Yaffe, R. O'Connor, J. D. Knowles, E. W. Blanch, R. Goodacre, *Anal. Bioanal. Chem.* **2010**, *397*, 1893.

[6] S. Botti, S. Almaviva, L. Cantarini, A. Palucci, A. Puiu, A. Rufoloni, *J. Raman Spectrosc.* **2013**, *44*, 463.

[7] C. V Di Anibal, L. F. Marsal, M. P. Callao, I. Ruisánchez, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2012**, *87*, 135.

[8] T. E. Journal, *Journal, Efsa* **2005**, 1.

[9] T. M. Fonovich, *Drug Chem. Toxicol.* **2013**, *36*, 343.

[10] T. H. E. Commission, O. F. The, E. Communities, **2005**, *2003*, 2004.

[11] W. Siangproh, K. Sonamit, S. Chaiyo, O. Chailapakul, *Anal. Lett.* **2013**, *46*, 1705.

[12] J. Zhang, J. Shao, P. Guo, Y. Huang, *Anal. Methods* **2013**, *5*, 2503.

[13] H. Yan, J. Qiao, Y. Pei, T. Long, W. Ding, K. Xie, *Food Chem.* **2012**, *132*, 649.

[14] R. Rebane, I. Leito, S. Yurchenko, K. Herodes, *J. Chromatogr. A* **2010**, *1217*, 2747.

[15] S. Liu, R. Que, X.-L. Wu, *New J. Chem.* **2013**, *37*, 3131.

[16] M. Zhou, X. Chen, Y. Xu, J. Qu, L. Jiao, H. Zhang, H. Chen, X. Chen, *Dye. Pigment.* **2013**, *99*, 120.

[17] J. Zhang, M. Wang, C. Shentu, W. Wang, Y. He, Z. Chen, *J. Electroanal. Chem.* **2012**, *685*, 47.

CHAPTER 3

CLASS-MODELLING APPROACH FOR ADULTERATION & AUTHENTICATION

3. 1. Introduction

3.1.1. *State-of-the-art*

As has been stated previously in Chapter 1, talking about multivariate qualitative methods generally means using classification techniques, which can be divided into two main blocks. One block focuses on class-modelling analysis (also known as 'soft modelling'), which models each class independently. The other main block refers to discriminant analysis (also referred to as 'hard modelling') that aims to divide data space up into separate regions, each of which corresponds to one class. In this last case, at least two classes must be defined.

A bibliographic search over the last ten years shows that whereas only a small amount of research has looked at the class-modelling approach, a substantial body of literature can be found regarding the discriminant approach. Figure 3-1 is a bar chart showing the percentage of scientific contributions corresponding to each approach.

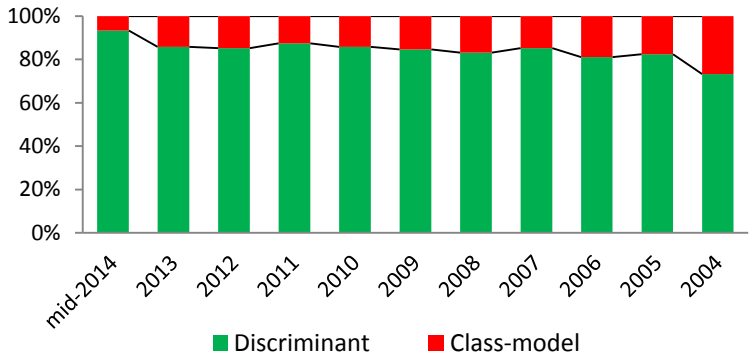

Figure 3-1. Percentage of scientific publications (from 2004 to mid-2014) obtained through a search on the scientific database *ISI Web of Knowledge (WoK)*, using 'discriminant' or 'class-modelling' related keywords in the analytical field.

According to the published research, among the reasons why the discriminant approach is preferred is the fact that they were developed first **[1]** and they are intuitive in the sense that a sample is always assigned to a one of the predefined classes. This last characteristic can be both an advantage and a drawback. Let us consider an example, the authentication of edible olive oil, for which two categories are defined: '*hojiblanca olive oil*' and '*arbequina olive oil'*. As mentioned, new samples will be classified to the closest class, i.e. assigned to the class whose characteristics are the most similar. Because they are always classified, the detection of outlier sample (e.g. a blend of both olive oils) is almost difficult. Alternatively, the analyst could consider the class-modelling approach, which allows samples to be classified into any defined category. However, this outline could be also classified in both categories, giving an ambiguous or inconclusive result. In fact, there is no established rule about whether the discriminant approach or the class-modelling approach is more effective because both approaches have their advantages and disadvantages.

Focusing on class-modelling techniques, they offer the possibility of modelling only one class (untargeted modelling) and are thus more useful when samples of only one class are available because it is impossible to cover all other areas. For example, when authenticating Priorat wine, the 'Priorat wine' class can be defined from that samples, but how can the other category ('not Priorat wine') be defined? Using only Spanish or Mediterranean wines for the 'not Priorat wine' class is inappropriate because this would exclude wines from all around the world from the 'not Priorat wine' category. This is a case where it is impossible to cover all the possibilities.

The untargeted approach basically builds a class from the training set samples and detects which new samples resemble this training set. Although this is not a new approach, its application has recently been increased. Figure 3-2 shows the rise over the last ten years in studies that have adopted the one-class modelling approach in chemical applications.

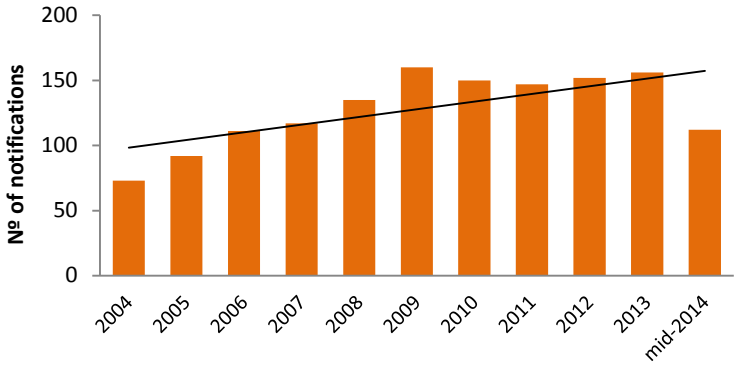

Figure 3-2. Number of papers (from 2004 to mid-2014) obtained through a search on the scientific database *ISI Web of Knowledge (WoK)*, using 'one-class classification' related keywords.

The one-class approach has been applied in many different ambits **[2]**. Of these, it is particularly suited to the detection of food fraud because this is a situation where it is impossible to take into account all existing types of fraud.

3.1.2. *Background*

There are several class-modelling techniques described in the literature. The two most widely used are Soft Independent Modelling of Class Analogies (SIMCA) **[3]** developed by Wold in 1970s, and Unequal Dispersed Classes (UNEQ) **[4]** introduced by Massart in 1986.

Nonetheless, other modelling techniques have also been developed. Recent examples from the last decade include Fuzzy Grid Encoded Independent Modelling for Class Analogies (FIMCA) **[5]**, Support Vector Machines (SVM) **[6,7]** and One-Class Partial Least Squares (OCPLS) **[8,9]**. All of these have been successfully applied to food fraud problems.

Classification techniques used in this thesis are all class-modelling techniques. They differ in the way the data are processed, in the way boundaries are set and/or in the steps used to obtain the output from the classification. Figure 3-3 gives a schematic idea of the types of models obtained for each technique and Table 3-1 summarizes the main differences between them.

UNEQ and SIMCA are PCA-based methods, although UNEQ can also be applied to original variables when the ratio object/variable is large enough. Thus, the distances to the centroid of the class are measured in the principal components (PCs) space, which are new orthogonal uncorrelated linear combinations of the original variables that describe the maximum variance. The differences between them are outlined in the model definitions:

- UNEQ defines the class model as a centroid vector, which is the mean vector of the training set samples. The boundaries define the class space and are determined by the critical T^2 Hotelling, thus giving an ellipse (two variables or PCs, see Figure 3-3) or a hyper-ellipsoid (more than two variables or PCs). define

- SIMCA defines the class model as the resulting combination of vectors obtained from the projections of the training set samples to each PC. In this way, the model shapes a vector (one PC), a plane (two PCs, see Figure 3-3) or a hyper-plane (three or more

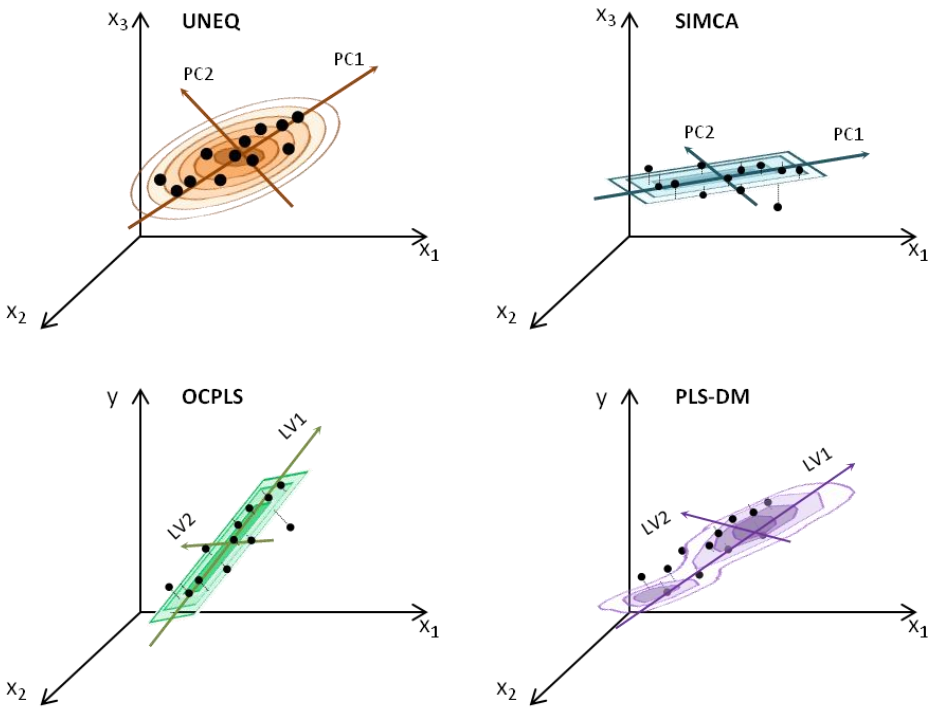

Figure 3-3. Examples of two dimensional models obtained with the different class-modelling techniques used in this thesis. Inner ellipses and boxes correspond to the class space at different confidence levels.

Table 3-1. Comparison of class-modelling techniques.

	UNEQ	**SIMCA**	**OCPLS**	**PLS-DM**
Data distribution	Normal-like	Normal-like	Normal-like	Complex-like
Compression data	None/PCA	PCA	PLS	PLS
Residuals consideration	No	Yes	Yes	Yes
Data centering	Yes	Yes	No	Yes

PCs). Finally, the class space is defined by a pre-set distance to the model and by the distance of the full multidimensional residuals.

In contrast, OCPLS and PLS-DM perform the variable reduction by using the PLS method which maximises covariance between the multivariate data matrix, for example spectral data, (**X**-bloc) and the response vector that indicates to which category the samples belong (**y**-bloc). As shown in Figure 3-3, the new variables are called latent variables (LV). These two methods differ in the **y**-bloc input:

- OCPLS considers **y**=1 (constant). According to reference **[8]**, **X**-bloc should not be column-centred otherwise all the features would be orthogonal to **y**. It should be noted that the class model and class space definition are similar in certain ways to those of SIMCA, except for the PCs-space, which in this method is LVs-space (see Figure 3-3).

- PLS-DM computes **y** as an estimation of sample density on the basis of inter-sample distances in the multivariate space. For this method, the class model is defined by a probability density distribution in which the residuals are used to define the class space.

PLS-DM is a density-based method that can be used for complex sample distributions whereas the other techniques have are all probabilistic and distance-based methods. Moreover, they require a multivariate normal sample distribution. This difference is shown in Figure 3-3 where PLS-DM is depicted with irregular geometrical shapes but the others have different but always regular geometric figures.

3. 2. Results Part A

MULTIVARIATE SCREENING IN FOOD ADULTERATION: UNTARGETED VERSUS TARGETED MODELLING

M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez

Food Chemistry, 147 (**2014**) 177-181.

Food Chemistry, 147 (2014) 177-181

Multivariate screening in food adulteration: untargeted versus targeted modelling

M. Isabel López[a], Esther Trullols[b], M. Pilar Callao[a], Itziar Ruisánchez[a]

[a]*Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovia i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain*

[b]*La Morella Nuts, S. A. U. Camí Ample s/n 43392 Castellvell del Camp., Tarragona, Spain*

ABSTRACT

Two multivariate screening strategies (untargeted and targeted modelling) have been developed to compare their ability to detect food fraud. As a case study, possible adulteration of hazelnut paste is considered. Two different adulterants were studied, almond paste and chickpea flour. The models were developed from near-infrared (NIR) data coupled with soft independent modelling of class analogy (SIMCA) as a classification technique. Regarding the untargeted strategy, only unadulterated samples were modelled, obtaining 96.3% of correct classification. The prediction of adulterated samples gave errors between 5.5% and 2%. Regarding targeted modelling, two classes were modelled: Class 1 (unadulterated samples) and Class 2 (almond adulterated samples). Samples adulterated with chickpea were predicted to prove its ability to deal with non-modelled adulterants. The results show that samples adulterated with almond were mainly classified in their own class (90.9%) and samples with chickpea were classified in Class 2 (67.3%) or not in any class (30.9%), but no one only as unadulterated.

***Keywords**: Multivariate screening, Untargeted modelling, SIMCA classification, Food fraud, Hazelnut, Adulteration*

1. Introduction

Fraud in food sector has been common since ancient times. Food adulteration tends to be economically motivated and is achieved through

the addition, substitution or removal of food ingredients [1]. It is an issue that concerns not only consumers, but producers and distributors as well.

In the recent past, food fraud has become more sophisticated due to the use of unconventional or synthetic adulterants, which has resulted in growing concern about associated health risks [2]. To guarantee food safety and quality, most analytical strategies are based on the knowledge of the contaminants [3,4]. The ever-increasing range of analytes that can be used in food fraud together with the impossibility of covering them all, make these strategies not always suitable for food adulteration problems. Otherwise, covering the widest possible range of analytes usually requires sophisticated analytical equipment such as chromatographic or mass spectrometry devices [5–7].

There is increasing demand for the development of fast, easy-to-use and low-cost analytical methods to test for adulteration. Methods based on spectroscopic techniques offer these advantages and their combination with multivariate chemometric techniques turns into a powerful tool for adulteration testing [7,8]. Some examples of ultraviolet (UV) [9], infrared (IR) [10], Raman [11,12], fluorescence [13] and nuclear magnetic resonance (NMR) [14,15] techniques can be found in the literature. Within the field of spectroscopy, one of the most widely used techniques in the food industry is near-infrared (NIR) spectroscopy, which has shown successful results in testing food quality as well as food adulteration [16–20].

Several multivariate techniques are based on their discriminating power, such as the *K*-nearest neighbour algorithm [9], linear discriminant analysis [4] and partial least squares discriminant analysis [14,20]. All of these methods require at least two classes to be defined, which implies *a priori* knowledge of the possible adulterant. Other multivariate techniques are based on their modelling ability. Soft independent modelling of class analogy (SIMCA) and unequal dispersed classes (UNEQ) are the most

well known of these [21,22]. Unlike the discriminant techniques, these approaches can be used in problems that have only one class of interest. Despite the potential of these class-modelling techniques, they are not commonly used in one-class modelling approaches [21–23].

In this paper, two methodologies for multivariate screening are proposed considering their different purpose. When the goal is to test whether a sample is adulterated or not, regardless of which adulterant might be present, a one-class modelling strategy is proposed. Therefore, a SIMCA model is established considering only the unadulterated class (untargeted strategy). Nevertheless, in some cases of food adulteration, one might have information about the possible adulterant, most of the cases based on experience or literature information. Therefore targeted modelling is proposed in which, in addition to the unadulterated class, a class for the known possible adulterant is modelled.

In this work, the feasibility of NIR spectroscopy coupled with SIMCA was assessed to test for adulteration, using a hazelnut paste problem as a case study. Hazelnuts and their derivatives (oils and pastes) are widely used as ingredients in many desserts, ice creams and chocolates or can be eaten alone as a snack. The hazelnuts price depends on the market and several ingredients can be added to reduce it being one of the often used paste or flour almond, since it is usually much cheaper. However, adulteration studies for this kind of nut are not extended in the literature [24]. There is, therefore, a need to develop analytical procedures to verify the quality of hazelnut paste and food safety, identifying any type of fraud motivated primarily by economic gain. For this work, two different adulterants were considered: (1) a similar product (almond paste) and (2) an unexpected product (chickpea flour), both present in a low percentage in the hazelnut paste.

2. Materials and methods

2.1 Samples

A total of 28 row hazelnuts (*Corylus avellana*) from different geographical origins were provided by *La Morella Nuts S.A.U*. The set contains sixteen samples from Spain, seven from Italy, three from Georgia and two from Azerbaijan. The nuts were first processed in a microwave and then ground until the oil exuded in a food processor, resulting in a coarse paste.

The adulterants used in this work were ecological chickpea flour from a commercial supplier and almonds from *La Morella Nuts S.A.U*. Adulterated samples were obtained from non-contaminated samples by adding almond paste or chickpea flour at 7%. In addition, a representative sub-set of 14 samples of each adulterant was selected using Kennard-Stone algorithm **[25]** to obtain new samples with other adulteration percentages (3% and 5%).

2.2 Instrumentation and software

NIR measurements were performed using a Bruker VECTOR 22/N in diffuse reflectance working condition. Spectra were recorded in the 3650-12000 cm^{-1} range at 8 cm^{-1} resolution. The spectral profile of the sample was acquired as the mean of 32 scans recorded during the rotation of the cylinder.

Data measured was processed with Matlab 6.5 software (Version 6.5, The Math Works, Inc., Natick, USA) and PLS Toolbox 3.5 (Eigenvector Research Incorporate).

3. Data analysis

3.1 Pre-processing

An offset correction was applied to the original data to eliminate any vertical shift in the spectra. It involves subtracting the absorbance value at 10538 cm^{-1} where no peak is observed from the values of absorbance at each wave number. This treatment was applied at each individual spectrum.

3.2 Exploratory analysis

Principal component analysis (PCA) was used as an unsupervised exploratory analysis tool to visualise the sample distribution in the multivariate space. Data was autoscaled before the chemometric treatment.

PCA decomposes the multivariate response arranged in an **X** matrix into a product of two new matrices as indicated in the Eq. 1 [26]:

$$\boldsymbol{X} = \boldsymbol{T_k}\boldsymbol{P_k^T} + \boldsymbol{E} \qquad Eq.1$$

$\boldsymbol{T_k}$ being the matrix of scores, $\boldsymbol{P_k}$ the matrix of loadings, $\boldsymbol{k}$ the number of factors included in the model and $\boldsymbol{E}$ the matrix of residuals, which contains the information not retained by the model.

3.3 Modelling analysis

SIMCA, introduced by Wold in 1976 [27], is based on PCA, as each class is modelled independently from any other. In order to assess how a given sample fits in each model, two scalar statistics are evaluated, *Q* and the *Hotelling* T^2.

Q_i-statistic is defined as the sum of squares of the residuals and can be calculated according to Eq. 2 **[26]**. Therefore, it is related to the amount of original information not included in the model.

$$Q_i = e_i e_i^T \qquad Eq.2$$

where e_i is the residual of sample *i* after applying the SIMCA model. The limit of the *Q*, Q_{lim}, can be calculated for the model under construction at a specific significance level (α), often set to 0.05.

The *Hotelling* T_i^2 measures the information of each sample within the SIMCA model and is calculated by means of Eq. 3 **[26]**. It therefore provides a measurement of how well each sample fits the model.

$$T_i^2 = I(x_i - \bar{x})S^{-1}(x_i - \bar{x})' \qquad Eq.3$$

where *I* is the number of samples in the training set, x_i the multivariate measurement of a sample *i*, $\bar{x}$ the column mean value of the training set and *S* the corresponding standard deviation. The limit of T^2 can be calculated at a specific α, often set to 0.05 (T^2_{lim}).

The reduced *Hotelling* T^2 (T_r^2) and the reduced *Q-statistic* (Q_r) values can be calculated from the ratio between the corresponding statistic of the sample *i* and the corresponding limit at α=0.05.

According to the “basic SIMCA”, boundaries are defined considering the reduced statistics. A sample must have values under 1 for both statistics to be considered “within the model”.

A newer version called “restrictive-SIMCA” takes the distance defined by Eq. 4 **[28]**. In this case, the boundary considered as “within the model” forms a semi-circle with a radius of 1.

Food Chemistry, 147 (2014) 177-181

$$d_{ij} = \sqrt{(Q_r)^2 + (T_r^2)^2} \qquad Eq.4$$

where d_{ij} is de distance of a sample *i* to a model *j*.

4. Results and Discussion

Fig. 1a shows the corrected NIR data of the unadulterated hazelnut samples. A slightly vertical variability between samples can be seen where one of them has a spectrum more differentiated from the others. Fig. 1b displays the unadulterated spectrum of a randomly chosen sample together with the corresponding spectra of the spiked samples with almond paste and chickpea flour at 7%. No visual difference between unadulterated and adulterated samples can be observed. Therefore, a multivariate treatment of spectral data by chemometrics is needed to distinguish between them.

4.1. Untargeted modelling

Firstly, an exploratory analysis of the unadulterated samples was made using PCA. Fig. 2 shows the PCA scores, being an 84.3% of the total variance explained by the first two principal components. Looking at the first PC, the figure shows that one of the samples is clearly far away of main block. So, this sample could be considered suspicious due to its different behaviour within the main PC. Regarding the second PC, three samples have slightly higher values than the main block, which can be explained by their specific geographical origin (Georgia).

All unadulterated samples were used to build the SIMCA model at 95% of confidence level, which was validated by leave-one-out cross validation. The number of PCs required to build the SIMCA model was determined plotting the percentage of correct classification versus the number of PC's which usually increases until no significance change is

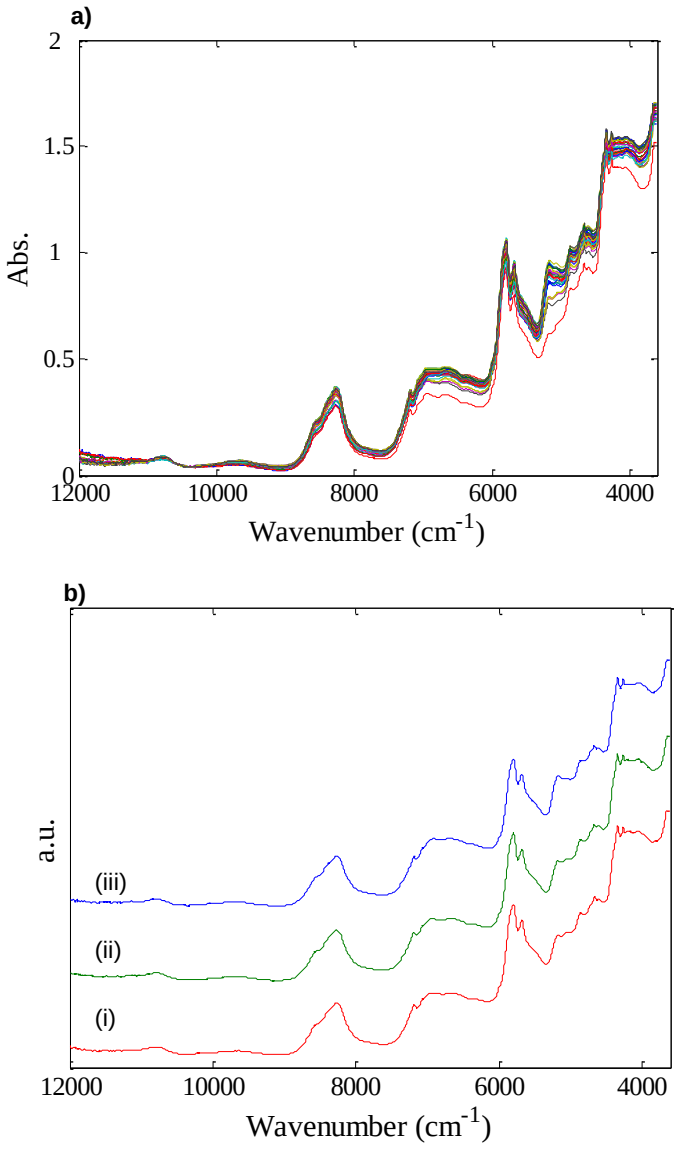

Fig. 1. *(a) Vertical off-set corrected NIR data of unadulterated hazelnut pastes. (b) NIR spectra of a randomly chosen unadulterated sample (i), adulterated at 7% with almond paste (ii) and adulterated at 7% with chickpea flour (iii). For the sake of clarity, a vertical shift was applied to the adulterated samples.*

observed when another PC is added. The usual criterion is to choose the minimum number of PC which gives similar correct classification value.

Accordingly, in our particular case the model was developed considering the first three PCs which presents 96.4% of correct classification.

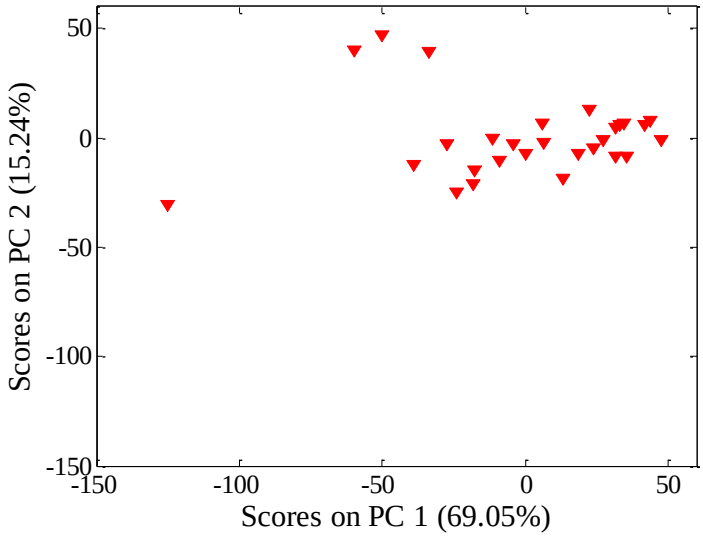

Fig. 2. *PCA score plot from corrected NIR data of hazelnut samples.*

Fig. 3 shows the results of T_r^2 versus Q_r values for the unadulterated samples. The acceptance of a sample as "within the model" is marked by the square area, whereas samples plotted at any point outside of this area are rejected. If more restrictive criterion is imposed, then the area is reduced to the marked semicircle. As shown, one sample has a high T_r^2 value and is out of the bounds defined by both approaches. As the suspicious sample was the same sample in both the exploratory analysis (the visual inspection of spectra and PCA) and the SIMCA model, it was got rid of the data set because keeping the sample could negatively affect the modelling process. On the other hand, another sample is

placed at the limits of the restrictive criterion, but very close to the boundary. In this case, this sample was maintained in the development of the model

So, a new 3PCs-based SIMCA model without the outlier was built at 95% of confidence level. This model was validated using two strategies: leave-one-out cross validation and training (20 samples) and test set (7 samples) randomly selected. The prediction results showed not significant differences and, considering the reduce number of available samples, we think that the leave-one-out cross validation is more reliable.

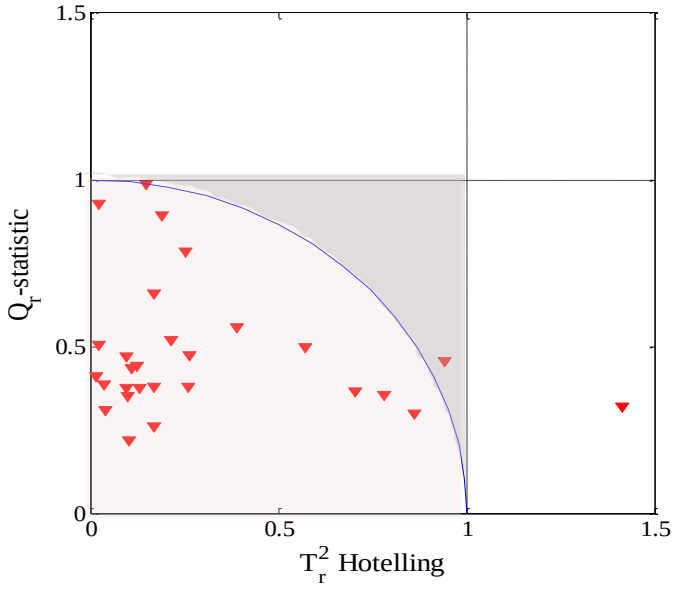

Fig. 3. *Plot of the resulting T_r^2 Hotelling and Q_r-statistic parameters from untargeted model, based on 3-PC space.*

Table 1 shows the classification results obtained with the untargeted model in the predictions of the 27 unadulterated samples and 54 samples of each adulterant (27, 14 and 13 samples with 7%, 5% and 3% of adulterant, respectively). The results show a correct classification of

about 95% is achieved whereas the success in the prediction ability is about 74%. Regarding the predictions of adulterated samples, a general success of more than 94% is achieved. No relevant changes were observed when basic and restrictive SIMCA criteria were considered.

Table 1. *Classification ability obtained using the untargeted model. Percentage of success in the application of untargeted model.*

	Basic-SIMCA (%)	**Restrictive-SIMCA (%)**
Hazelnut samples	74.1 (96.3)[a]	69.7 (92.6)[a]
Almond adulterated samples	94.5	94.5
Chickpea adulterated samples	98.2	100

[a] Percentage of correct classification in brackets

These percentages of success have different meanings and consequences based on whether the samples were adulterated or not. Around 25% of the unadulterated samples were not considered as such. This kind of error has economic consequences, as these samples would be unnecessarily removed from the market and a confirmatory or further analysis would probably be required. On the other hand, 5% of the samples adulterated with almond paste and about 1% of the samples adulterated with chickpea were considered unadulterated. This error is more important than the previous one since it implies fraud not only for consumers, but also for producers and distributors. Since almonds are more closely related to hazelnuts than chickpeas are, they are more likely to be mistaken for hazelnuts, which is why the error rate is slightly lower for chickpeas. No trends related to the percentage of adulteration present in the sample were found in neither of adulterants. Nevertheless, the results of the analysis satisfactorily distinguished between adulterated and unadulterated samples.

Finally, a deeper look at the T_r^2 and Q_r statistics shows that the adulterated samples are considered "not in the model" mainly due to their Q_r value, which means that the adulterated samples have information not included in the model: the contribution of the adulterant in the spectra (results not shown).

4.2 Targeted modelling

In this part of the study two classes were modelled: Class 1, corresponding to the unadulterated hazelnut samples, and Class 2, corresponding to the samples adulterated with almond paste. This adulterant was selected for modelling because its use has been reported in the literature. Based in our experience is one of the most common adulterants used to decrease the final hazelnut price, being one of the main reasons its similarity to hazelnut. But it is not the only possible hazelnut adulterant reported so, to check the model prediction behaviour in front of an unknown or not modelled adulterant, an unexpected adulterant has been chosen as is the case of using chickpea, which of course also allows decreasing the final hazelnut price.

The SIMCA classification models were built in accordance with the methodology described in section 4.1 (exploratory analysis and selection of number of PCs), being each class developed based on the first three PCs. Because in the previous study both SIMCA strategies (basic and restrictive) gave similar results, the use of the restrictive approach was not justified. Therefore, the following results are given based on the "basic SIMCA".

The classification results are presented as a contingency table in Table 2. It can be seen that unadulterated hazelnut samples were mainly classified in their own class (Class 1), but among these, there were a large number that so are in Class 2 (double classification). The samples adulterated with almond were mainly classified in their own class, and a

few were double classified or considered as “not in any model”. Samples adulterated with chickpea were mainly assigned to Class 2 (adulterated samples) or “not in any class” (samples different from the modelled ones). No trends related to level of adulteration were found in neither of adulterants.

Table 2. *Contingency table obtained from whole samples.*

	Hazelnut samples	Almond adulterated samples	Chickpea adulterated samples
In Class 1	10/27	0/0	0/0
In Class 2	1/27	49/54	36/54
Double classification (In Class 1 and Class 2)	16/27	3 /54	1/54
Not in any class	0/0	2/54	17/54

Double classified samples should undergo a confirmatory analysis. Prior to this, the in-depth study of the distance between the sample and the models could perhaps help to define the assignation. Fig. 4 shows the distances of the doubly classified samples to each class (according to Eq. 4, section 3.3). Most of the unadulterated samples (triangle) are closer to their own class (Class 1), with values between 0.3 and 0.5, than to Class 2, with values over 0.6. They therefore most likely belong to Class 1. However, four unadulterated samples are plotted equidistantly between the two classes and close to the line. So, in this case the study of the distance could not help to the final assignation of them. This additional information could significantly reduce the number of samples that need to undergo a confirmatory analysis.

Among the samples adulterated with almond paste (circles) only one sample is plotted closer to its own class (0.5 to Class 2 compared to 0.9 to Class 1), so it can be considered as properly assigned to Class 2.

Meanwhile the distance of the other two to both models is similar. Finally, the adulterated chickpea sample is nearly equidistant to both models, but with quite high distance values (between 0.9 - 1), so it is near the boundaries of both models. It has to be remarked that the four adulterated samples with double classification are the same four samples wrongly assigned by the untargeted SIMCA model. With the targeted model they were identified as ambiguous due to the double classification, whereas this was not possible with the untargeted model because once a sample is assigned as not adulterated (fits the model), no further studies are conducted.

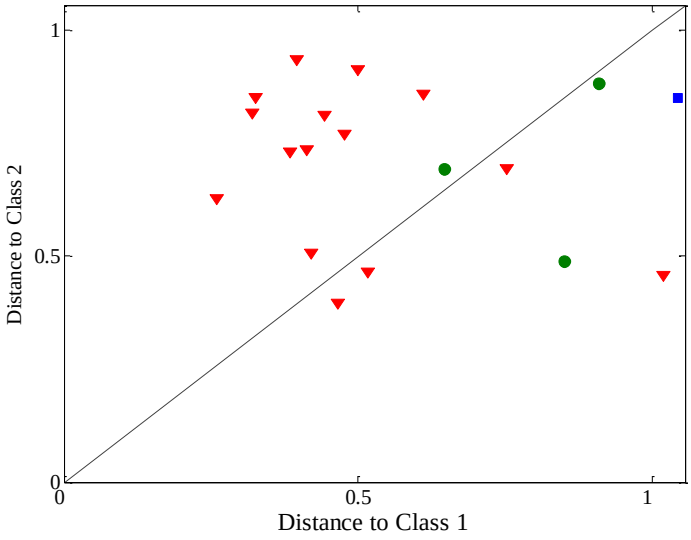

Fig. 4. *SIMCA distance to each model in targeted modelling. The upper triangle corresponds to less distance to the unadulterated class whereas the bottom triangle represents less distance to the adulterated model. The dashed line indicates the same distance to both models. Unadulterated samples (▼), adulterated with almond paste (●) and adulterated with chickpea flour (■).*

5. Conclusions

A multivariate screening method based on NIR spectroscopy coupled with the SIMCA technique was developed to test for adulteration in hazelnut pastes by means of two approaches: untargeted and targeted modelling. Both approaches gave satisfactory results.

If the main interest is to find out whether a sample is adulterated or not, untargeted modelling is a good approach. Some of its advantages are that the classification model can be established by only analysing the unadulterated samples. The model is easy to implement and, once it has been established, it can be used to predict any sample, even when there is a lack of knowledge about the potential adulterants.

Targeted modelling is a suitable approach in situations where testing is required for a common or known adulterant. One of the main advantages of this approach is that it can provide detailed information about the problem. Our results show suitable assignations even though the two defined classes are very similar, as hazelnut was the primary basis (more than 93%) in all of them. When a sample containing a non-modelled adulterant was predicted, it was not assigned to the unadulterated class, which is the most important. In our particular case, some of those samples were not assigned to any class, although most of the samples of this type were wrongly assigned to the adulterated class.

Comparing both approaches, untargeted modelling is more restrictive, as it offers a single response: a sample either fits or does not fit the model. On the other hand, with targeted modelling a sample can have several responses or assignations: fits the model of just one class, fits the model of more than one class (double classification) or does not fit any model. Therefore, the targeted model provides more information about the tested samples than the untargeted model.

6. Acknowledgments

This work was supported by Generalitat de Catalunya under project 2009 SGR 549. M.I. López thanks Rovira i Virgili University for providing her doctoral fellowship (2011BRDI-06-09).

7. References

[1] J.C. Moore, J. Spink, M. Lipp, *J. Food Sci.* 77 (2012) R118–R126.

[2] R. Krska, A. Becalski, E. Braekevelt, T. Koerner, X.-L. Cao, R. Dabeka, et al., *Anal. Bioanal. Chem.* 402 (2012) 139–162.

[3] S.A. Haughey, S.F. Graham, E. Cancouët, C.T. Elliott, *Food Chem.* 136 (2013) 1557–1561.

[4] L. Mannina, M. D'Imperio, D. Capitani, S. Rezzi, C. Guillou, T. Mavromoustakos, et al., *J. Agric. Food Chem.* 57 (2009) 11550–11556.

[5] C.D. Calvano, C. De Ceglie, A. Monopoli, C.G. Zambonin, *J. Mass Spectrom.* 47 (2012) 1141–1149.

[6] V. Di Stefano, G. Avellone, D. Bongiorno, V. Cunsolo, V. Muccilli, S. Sforza, et al., *J. Chromatogr. A.* 1259 (2012) 74–85.

[7] R. Karoui, J. De Baerdemaeker, *Food Chem.* 102 (2007) 621–640.

[8] P.K. Ghosh, D.S. Jayas, *Sens. Instrum. Food Qual. Saf.* 3 (2009) 3–11.

[9] C. V Di Anibal, M. Odena, I. Ruisánchez, M.P. Callao, *Talanta.* 79 (2009) 887–892.

[10] Q. Zhang, C. Liu, Z. Sun, X. Hu, Q. Shen, J. Wu, *Food Chem.* 132 (2012) 1607–1613.

[11] C. V Di Anibal, L.F. Marsal, M.P. Callao, I. Ruisánchez, *Spectrochim. Acta Part A* 87 (2012) 135–141.

[12] E.C. López-Díez, G. Bianchi, R. Goodacre, *J. Agric. Food Chem.* 51 (2003) 6145–6150.

[13] M.R. Kunz, J. Ottaway, J.H. Kalivas, C.A. Georgiou, G.A. Mousdis, *J. Agric. Food Chem.* 59 (2011) 1051–1057.

[14] R. Consonni, L.R. Cagliani, C. Cogliati, *Talanta.* 88 (2012) 420–426.

[15] C. V Di Anibal, M.P. Callao, I. Ruisánchez, *Talanta.* 86 (2011) 316–323.

[16] D. Behmer, *Food Sci. Technol.* 24 (2010) 47–48.

[17] F. Guy, S. Prache, A. Thomas, D. Bauchart, D. Andueza, *Food Chem.* 127 (2011) 1280–1286.

[18] P. Oliveri, G. Downey, *TrAC, Trends Anal. Chem.* 35 (2012) 74–86.

[19] T. Szigedi, J. Lénárt, M. Dernovics, S. Turza, M. Fodor, *Int. J. Food Sci. Technol.* 47 (2012) 436–440.

[20] R. Vitale, M. Bevilacqua, R. Bucci, A.D. Magrì, A.L. Magrì, F. Marini, *Chemom. Intell. Lab. Syst.* 121 (2013) 90–99.

[21] P. Oliveri, M.C. Casolino, M. Casale, L. Medini, F. Mare, S. Lanteri, *Anal. Chim. Acta.* 761 (2013) 46–52.

[22] L. Xu, Z.-H. Ye, S.-M. Yan, P.-T. Shi, H.-F. Cui, X.-S. Fu, et al., *Anal. Chim. Acta.* 754 (2012) 31–38.

[23] L. Xu, C.-B. Cai, D.-H. Deng, *J. Chemom.* 25 (2011) 568–574.

[24] J.S. Bonvehi, F.V. Coll, *Int. J. Food Sci. Technol.* 44 (2009) 456–466.

[25] M. Daszykowski, B. Walczak, D.L. Massart, *Anal. Chim. Acta.* 468 (2002) 91–103.

[26] A. Rius, M.P. Callao, F.X. Rius, *Analyst.* 122 (1997) 737–741.

[27] S. Wold, *Pattern Recognit.* 8 (1976) 127–139.

[28] C. Durante, R. Bro, M. Cocchi, *Chemom. Intell. Lab. Syst.* 106 (2011) 73–85.

3. 3. Results Part B

PARTIAL LEAST SQUARES DENSITY MODELING (PLS-DM) – A NEW CLASS-MODELING STRATEGY APPLIED FOR THE AUTHENTICATION OF OLIVES IN BRINE BY NEAR-INFRARED SPECTROSCOPY

P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini,

S. Lanteri

Analytica Chimica Acta, 851 (**2014**) 30-36

Analytica Chimica Acta, 851 (2014) 30-36

Partial least squares density modeling (PLS-DM) – A new class-modeling strategy applied to the authentication of olives in brine by near-infrared spectroscopy

Paolo Oliveri[a,Ŧ], M. Isabel López[b,Ŧ], Maria Chiara Casolino[a], Itziar Ruisánchez[b], M. Pilar Callao[b], Luca Medini[c], Silvia Lanteri[a]

[a]*Department of Pharmacy, University of Genoa, Via Brigata Salerno, 13, I-16147 Genoa, Italy*

[b]*Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovira i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain*

[c]*Special Company for Professional Training and Technological and Commercial Promotion of the Chamber of Commerce of Savona, Regione Rollo, 98, I-17031 Albenga (SV), Italy*

ABSTRACT

A new class-modeling method, referred to as partial least squares density modeling (PLS-DM), is presented. The method is based on partial least squares (PLS), using a distance-based sample density measurement as the response variable. Potential function probability density is subsequently calculated on PLS scores and used, jointly with residual Q statistics, to develop efficient class models. The influence of adjustable model parameters on the resulting performances has been critically studied by means of cross-validation and application of the Pareto optimality criterion. The method has been applied to verify the authenticity of olives in brine from cultivar Taggiasca, based on near-infrared (NIR) spectra recorded on homogenized solid samples. Two independent test sets were used for model validation. The final optimal model was characterized by high efficiency and equilibrate balance between sensitivity and specificity values, if compared with those obtained by application of well-established class-modeling methods, such as soft independent modeling of class analogy (SIMCA) and unequal dispersed classes (UNEQ).

***Keywords**: Class-modeling; One-class classifier; Density estimation; Partial least squares (PLS); Potential functions.*

Ŧ These two authors contributed equally.

1. Introduction

Class-modeling and discriminant classification methods are widely employed to build mathematical models aimed at characterizing samples with respect to qualitative properties. Discriminant classification techniques are used to determine to which class, among a number of pre-defined classes, a sample most probably belongs, by setting a delimiter between the pairs of classes. Each new sample is then always assigned to one of the categories, even in the case of samples that do not belong to any class studied. Also, class-modeling techniques can be used for multiclass classification but, in this case, each new sample can be either assigned to one, more than one or none of the predefined classes. The suitable selection of the classification strategy depends on the problem to be solved and it may represent an important issue to be considered when working with these techniques [1]. As a matter of fact, the modeling of a single class of interest, to verify whether a sample is compatible or not with the characteristics of that class [2,3], is only allowed by the class-modeling techniques, which – for this reason – are also referred to as one-class classifiers [4] or untargeted modeling methods [5].

In multi-class classification, the discriminant approach is followed more frequently than the class-modeling one. A discriminant classification method which has gained increasing attention in the last years is based on partial least squares (PLS) regression, and it is usually referred to as discriminant PLS (D-PLS) or PLS discriminant analysis (PLS-DA) [6–8]. In the recent years, a number of attempts have been addressed to develop class-modeling techniques exploiting the advantages offered by the PLS method [9–12].

In particular, a method called one-class PLS (OC-PLS) has been recently presented, in which a PLS model is built using a constant response (y=1), i.e., identical values for all of the training samples belonging to the

class of interest [13]. Hotelling's T^2 and Q statistics are used to verify compliance of test samples with the class model. Such a method proved to be efficient on different data sets [14–16]. Nonetheless, some issues have to be considered. First of all, PLS regression on column-centered *x*-block data leads to degenerate solutions when the response variable is constant [13]. Secondly, this strategy gives equal importance to all samples in the class model definition, without taking into account class heterogeneity. Finally, the use of T^2 statistics on the PLS scores implies the underlying hypothesis of a normal distribution.

In order to manage data with non-normal and non-uniform distributions, some class-modeling methods have been proposed. Among the most efficient, potential function methods (PFM) [17], a family of probabilistic non-parametric techniques, define the class model by empirically estimating a probability density distribution for a class of interest [18,19]. An important limitation of such techniques is related to the impossibility of direct application to data sets with high variable dimensionality. In fact, the higher the number of variables, the lower the reliability in the estimation of the probability density, as well as the establishment of the critical value for the decision rule. In order to overcome this hurdle, PFM are commonly applied after unsupervised variable reduction by means of principal component analysis (PCA) [19].

In the present study, a new PLS-based class-modeling strategy is presented, called partial least squares density modeling (PLS-DM), which combines the features of PLS and PFM, together with Q statistics, to obtain highly efficient class models. The method was applied on a set of near-infrared (NIR) spectra recorded on samples of olives in brine, with the purpose of verifying the authenticity of olives from cultivar Taggiasca. In this application – like in most of the cases involving verification of food authenticity claims – the focus was on a single class (cultivar *Taggiasca*). In such a case, the discriminant approach would require the collection of

two sets of training samples: one representative of the *Taggiasca* olives and a second representative of the entire production of all of the olives potentially usable to make frauds. Such a condition is rarely realizable in practice, and collected sets of non-compliant samples would be under-representative of the whole non-compliance possibilities. This inadequacy would inevitably lead to biased decision rules, the outcomes of which being heavily dependent on those samples included in the non-compliant set. For this reason, in such a case, decisions regarding sample conformity based on class-modeling strategies are more robust and suitable than those based on discriminant approaches [20].

Olives from cultivars *Leccino* and *Coquillo*, being morphologically very similar to *Taggiasca*, are suspected to be used in fraudulent manufactures and, therefore, they were considered in this study as potential adulterants.

Model performances are evaluated in terms of sensitivity and specificity and by application of the Pareto optimality criterion. Results are compared with those achieved by unequal dispersed classes (UNEQ) [21], soft independent modeling of class analogy (SIMCA) [22] and OC-PLS [13]. UNEQ and SIMCA are the class-modeling techniques most commonly applied in chemometrics, while OC-PLS represents the most recent modeling method based on PLS.

2. Theory

2.1 Partial least squares

Partial least squares (PLS) is a multivariate regression technique which computes directions in the space of the predictors (**X**) characterized by the maximum covariance with the response variable (**y**). Such directions, called latent variables (LVs), are employed to define the regression model:

$$\mathbf{X}_{I,V} = \mathbf{T}_{I,L} \cdot [\mathbf{P}_{L,V}]^T + \mathbf{E}_{I,V} \qquad Eq.\ 1$$

$$\mathbf{y}_{I,1} = \mathbf{U}_{I,L} \cdot [\mathbf{q}_{L,1}]^T + \mathbf{f}_{I,1} \qquad Eq.\ 2$$

where *I* is the number of samples, *V* is the number of original predictor variables, **T** and **U** are the matrices of scores (i.e., projections) of **X** and **y**, respectively, **P** and **q** contain the loading terms, **E** and **f** contain the error terms. The most appropriate number of LVs (*L*) is usually determined in a cross-validation (CV) cycle, by studying the evolution of the quality parameter – such as the prediction error – as a function of the increasing number of LVs [23].

2.2 Residuals and confidence limit

Q statistics is related to the residuals, i.e. the fraction of the information about samples not explained by the *L* latent variables retained in the final PLS model:

$$Q_i = e_i e_i^T \qquad Eq.\ 3$$

where e_i is the vector of residuals of sample *i* after applying the PLS model. Its confidence limit, Q_α, is computed according to Jackson [24]:

$$Q_\alpha = \theta_1 \left[\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \qquad Eq.\ 4$$

where z_α is the value of the standard normal deviate corresponding to the upper (1-α) percentile, and θ_j terms and h_0 are defined as:

$$\theta_j = \sum_{l=L+1}^{\min(I,V)} \lambda_l^j \quad \text{for} \quad \mathrm{j} = 1,2,3 \qquad Eq.\ 5$$

and

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \qquad Eq.\ 6$$

The θ_j terms are the sums of the eigenvalues (λ) raised to the j^{th} power for the LVs that have not been retained in the regression model. It has to be remarked that Eq. (4) assumes the residuals to be normally distributed, a condition that is approximately verified, as it can be demonstrated **[24]**.

2.3 Potential function methods

Potential function methods (PFM) for class-modeling are based on the estimation of the probability density distribution for a class *c* composed by I_c objects. Each object *i* belonging to the class is assumed to contribute to the global class probability as electrical charges contribute to a potential field. So, the global probability function is determined as the sum of such individual contributions in the multivariate space normalized by I_c **[25,26]**. Compliance of a new object with class *c* is determined by comparing its probability density value with the critical value of the probability density distribution at a selected confidence level **[27]**.

The global probability function, $f(x)$, is commonly obtained by summing the individual contributions $f_i(x)$ defined as:

$$f(x) = \sum_{i=1}^{I_C} f_i(x) = \sum_{i=1}^{I_C} \frac{1}{I_c} \cdot \prod_{v=1}^{V} \frac{1}{as_v\sqrt{2\pi}} e^{-\frac{1}{2}\cdot\frac{(x_v - x_{iv})^2}{(as_v)^2}} \qquad Eq.\ 7$$

where *a* is the smoothing coefficient that applies to all of the variables within class *c*, and s_v is the standard deviation of variable *v* within the training set of class *c*. The smoothing coefficient cooperates in determining the shape of the distribution, being higher the smoothness when increasing *a* (which usually ranges between 0 and 1.5).

In order to define the class boundary, the critical value (f_α) of the probability density distribution $f(x)$, at a selected confidence level (1-α), is obtained from the critical value of the chi-squared distribution (χ^2_α) by the so-called equivalent determinant method [17], according to:

$$f_\alpha = \frac{1}{(2\pi)^{\frac{V}{2}} \left|\hat{\mathbf{C}}\right|^{\frac{1}{2}}} \cdot e^{-\frac{\chi^2_\alpha}{2}} \qquad Eq.\ 8$$

where V is the number of variables and $\left|\hat{\mathbf{C}}\right|$ is the estimation of the determinant corresponding to the variance-covariance matrix of the multivariate normal distribution equivalent to the probability distribution estimated by PFM, computed as:

$$\left|\hat{\mathbf{C}}\right|^{\frac{1}{2}} = \frac{1}{2^V \pi^{\frac{V}{2}} \frac{\sum f(x)}{I}} \qquad Eq.\ 9$$

According to Forina et al. [17], the term “equivalent” is used in this context to indicate distributions with the same mean value.

3. Data and methodology

3.1 Olive samples and data sets

Representative samples of olives in brine from the cultivars under study were collected by the Special Company for Professional Training and Technological and Commercial Promotion of the Chamber of Commerce of Savona (Albenga, Italy). Samples from 2010–2011 and from 2011–2012 harvests were randomly split into a training set formed by 187 samples (83 *Taggiasca* and 104 *Leccino* and *Coquillo*) and a test set containing the remaining 21 samples (9 *Taggiasca*, 12 *Leccino* and *Coquillo*). Samples from 2012-2013 harvest were used as an external

test formed by a total amount of 25 samples (19 *Taggiasca* and 6 from different cultivars).

Olives were washed with water, dried with tissue paper and pitted. Then, the pulp was ground and 30 g were placed in a Petri dish and analyzed.

3.2 Data acquisition and data treatment

NIR measurements on olive samples were performed in the chemical laboratory of Albenga, by a FT-NIR Thermo Scientific spectrometer (Thermo Scientific, AntarisII™ FT-NIR Analyzer). Spectra were recorded in the 4,000–10,000 cm^{-1} range, at 4 cm^{-1} resolution. Samples were analyzed in the reflection mode using standard glass Petri dishes with 9 cm diameter. The spectral profile of the sample was acquired as the mean of 64 scans recorded during rotation of the glass dish. Systematic differences among Petri dishes – mainly due to small variations in glass thickness – were corrected by dividing point by point the reflectance spectrum of each sample by the spectrum of a certified reference material (Spectralon®) with 99% reflectance in the entire NIR region, recorded on the same dish. The whole analytical procedure was repeated on three different aliquots of each sample and the resulting average spectrum was submitted to data analysis.

All the chemometric data processing was performed by means of in-house Matlab routines (The MathWorks, Inc.).

3.3 Partial least squares density modeling (PLS-DM)

The strategy deployed for developing PLS-DM models is schematized in Fig. 1.

Initially, a PLS model is developed using analytical data as **X** predictor matrix and a density vector as the **y** response vector. The response

value (y_i) – for each sample *i* of the training set of the class to be modeled – is computed as an estimation of sample density (d_i), based on inter-sample distances in the multivariate space. In more detail, all the Euclidean distances from the sample *i* to each of the other training samples are computed. Such distances are, therefore, ordered, and the density value (d_i), obtained as the sum of the *k* smallest (i.e., lowest-order) distances, was studied varying *k*. Moreover, parameter *k* influences the smoothness of density function, which evolves from a sharper to a smoother shape while increasing *k*.

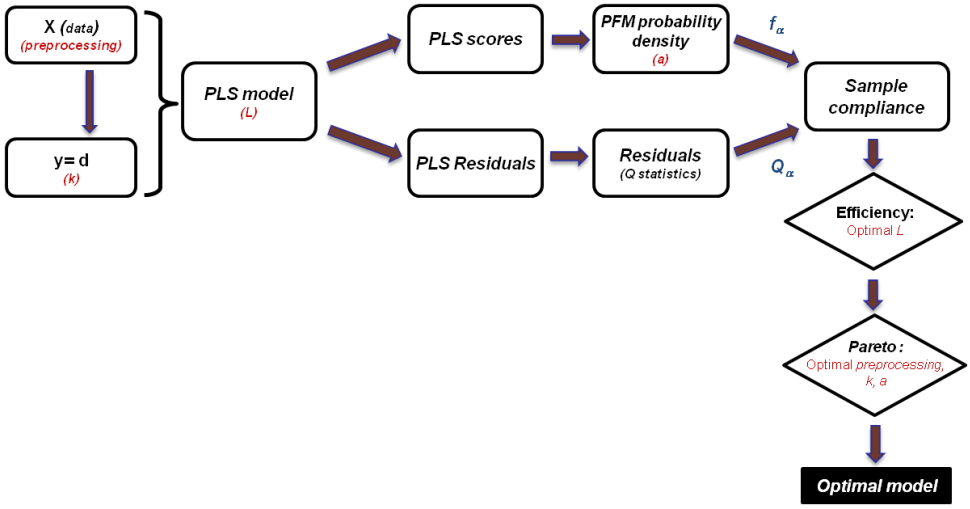

Fig. 1. *Scheme of the PLS-DM strategy. Red parameters between brackets show the features to be optimized.*

After PLS modeling, the PLS scores of the training set on the first *L* latent variables selected are used as input to estimate the PFM probability density of the class, with different smoothing coefficients (*a*). Then, the critical value, f_α, of the probability density distribution is computed, at a preselected confidence level (1-α). In addition, the PLS residuals are used to compute the critical value of *Q* statistics, Q_α, at the same

confidence level. In this way, compliance of each object with the class model is granted when it satisfies both f_α and Q_α criteria.

The algorithm calculates models with all of the different parameter combinations – i.e., the *k* distance, the *a* smoothing coefficient and the *L* latent variables as well as the suitable X-block *pre-processing*. Then, the procedure selects the optimal parameter combination by: (1) fixing the number of *L* through the efficiency criterion and (2) evaluating the rest of parameters applying the multicriteria Pareto's decision method.

Efficiency gives a global evaluation of the modeling performances, computed in this study as the geometric mean of sensitivity and specificity. Sensitivity is defined as the percentage of samples belonging to the modeled class which is correctly accepted by the class model whereas specificity is the percentage of samples not belonging to the modeled class which is correctly rejected by the model **[28]**.

According to Pareto-optimality method, each class model can be represented in a bidimensional scatter plot reporting sensitivity and specificity on the two axes, respectively. A theoretical example of Pareto diagram is shown in Fig. 2. The triangle represents the ideal – usually utopic – solution. A point on this graphic is defined as Pareto efficient when no other points show better results in one criterion without showing worse results in the other one. Point A is dominated (Pareto inefficient) by point B as it has higher values in both objectives (i.e. sensitivity and specificity). Moreover, point A is dominated by point C because both of the points have the same specificity, but point C has a higher sensitivity value. The Pareto front is the dashed line that connects all the Pareto efficient results **[29]**. In the end, the operator chooses the final optimal solution among the Pareto efficient solutions, looking for the best compromise between sensitivity and specificity, and depending on what is preferable for a particular problem under study.

Analytica Chimica Acta, 851 (2014) 30-36

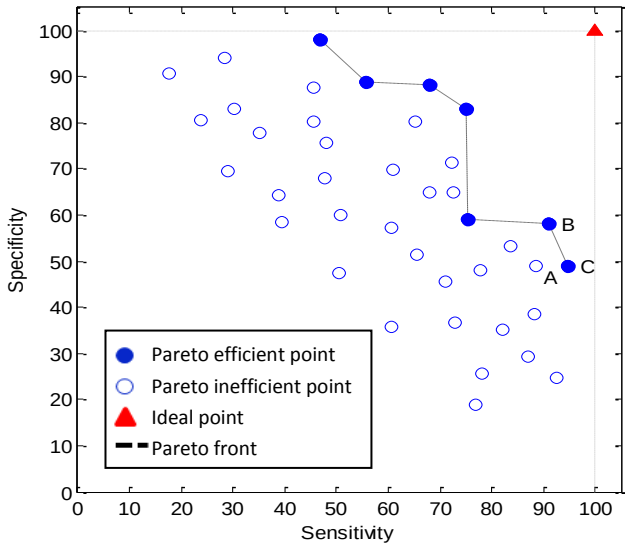

Fig. 2. *Example of a Pareto diagram.*

All the optimization steps in model building were performed by means of a cross-validation (CV) procedure with five deletion groups (Venetian-blind scheme), using only the training set samples. The optimal values for the different parameters were selected by examining all the different outcomes in terms of model efficiency and optimal solutions according to Pareto methodology. The optimal class model was validated by means of the test sets described above, to obtain final sensitivity and specificity estimations.

4. Results and Discussion

4.1. Spectra

The raw NIR spectra of olive samples from different cultivars and different harvests are shown in Fig. 3a. The principal differences observable among spectra consist in baseline shifts and effects ascribable to light scattering phenomena. To correct for such undesired

effects, data were pre-processed by the standard normal variate (SNV) transform [30]. eliminate regions that previous studies demonstrated to be non-informative for the characterization of olives in brine [31]. Corrected data are depicted in Fig. 3b.

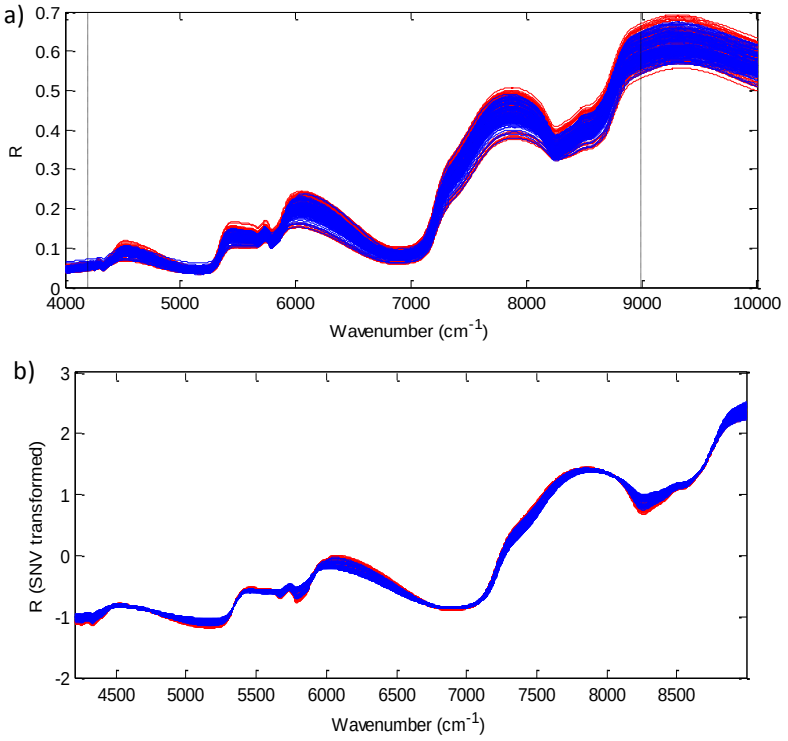

Fig. 3. *(a) Original NIR spectra acquired for all the Taggiasca (red), Leccino and Coquillo (blue) olive samples. Vertical lines mark the boundaries of the reduced spectral range. (b) SNV corrected spectra.*

4.2. PLS-DM

According to the methodology described (Fig. 1), several parameters were settled in order to define the optimal model – namely, the pre-processing, *k*, *L*, and *a*. In more detail, four possibilities were considered for variable pre-processing: no pre-processing, mean centering, scaling and autoscaling. As for the *k* parameter, integer values from 1 to 7 were

considered. The number *L* of latent variables was varied from 1 to 15. Finally, the smoothing coefficient *a* was varied from 0.3 to 0.8, with 0.1 increments. These parameters were varied within the specified ranges and the results of all of the combinations were tested by a cross-validation scheme to select the optimal model.

As a first step, the optimal number of LVs was selected, for each condition, by considering the maximum efficiency of the resulting class model, evaluated by cross-validation. By way of example, one of the evaluated models was randomly selected to illustrate how sensitivity, specificity and efficiency evolve at increasing the LV number (Fig. 4). In this particular case, 4 LVs were selected, according to the maximum efficiency criterion. Examination of such profiles may allow excluding solutions potentially associated to data overfitting.

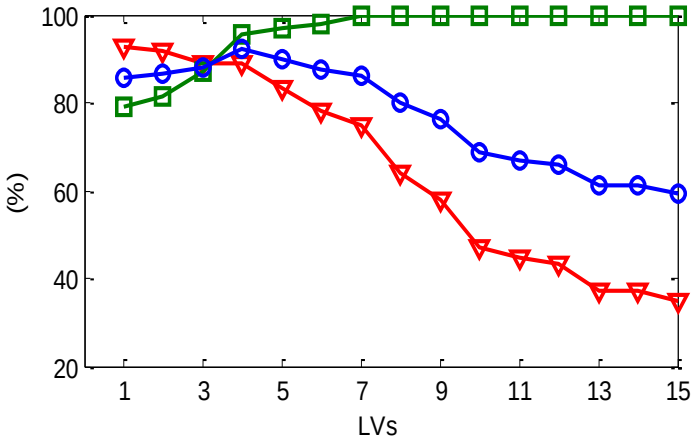

Fig. 4. *Example of evolution of sensitivity (red triangles), specificity (green squares) and efficiency (blue circles) at increasing the LV number.*

In a second step, models at fixed LVs were evaluated by the Pareto diagram. The effect of each parameter on the Pareto results was analyzed in detail. Fig. 5a depicts the effect of varying *k* from 1 to 7. In this case, no clear tendencies were noticeable; however, optimal Pareto

solutions (Table 1) are associated with lower *k* values (ranging from 1 to 3). Regarding the effect of the pre-processing, more marked tendencies can be observed (see Fig. 5b). In particular, it can be noticed that results corresponding to scaled (S) or autoscaled data (A) are characterized by higher sensitivity values than original or mean-centered data (N and M, respectively). Fig. 5c shows the effect of the smoothing coefficient *a* employed in the potential function probability estimation, increasing from 0.3 to 0.8, with 0.1 increments, codified by capital letters from A to F. A definite positive correlation between *a* values and sensitivities of resulting class models can be observed. Such outcomes could be expected since the smoothing coefficient determines the shape of probability density-based class spaces and, specifically, the class boundaries become larger at increasing *a* values. For the same reason, a concomitant reduction of specificity values should be expected. Nevertheless, the contribution of *Q* statistics softens this effect, maintaining elevate specificity values.

Pareto solutions, summarized in Table 1, correspond to the Pareto front (points connected by the black line in Fig. 5). The final model was selected among the solutions laying on the front, looking for a balance between sensitivity and specificity, whose certified producers have to be recognized as compliant by the models. Accordingly, three coincident solutions marked in bold are the most suitable, characterized by 96.4% sensitivity and 93.1% specificity in cross-validation. It has to be remarked that such three equivalent optimal solutions correspond to slight variations on the smoothing parameter (a = 0.6, 0.7 and 0.8, respectively). The following optimal conditions were chosen: k = 1, preprocessing = scaling; L = 2; a = 0.7.

Finally, it can be noted that slight variations of model parameters lead to not significantly different results, and such a feature can be considered as a significant proof of robustness of PLS-DM models towards soft perturbations of the adjustable parameters.

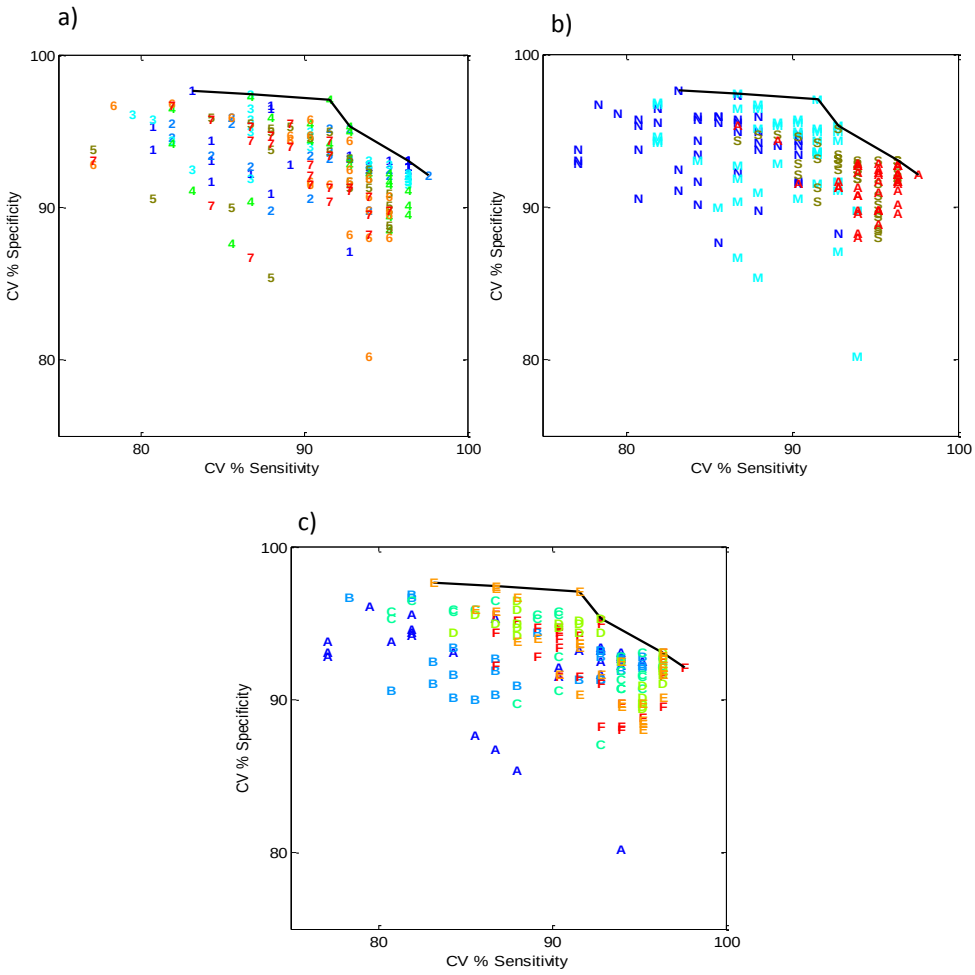

Fig. 5. *Sensitivity vs. specificity Pareto diagram. The effect of the parameters on the results is evidenced by coding: (a) k: integers from 1 to 7; (b) pre-processing: N = no- preprocessing, M = mean centering, S = scaling, and A = autoscaling; (c) a: A = 0.3, B = 0.4, C = 0.5, D = 0.6, E = 0.7, F = 0.8. Pareto front is marked with the black solid line.*

4.3. Validation

The final model was applied on the test sets in order to evaluate its performances in the prediction of new samples, as well as the stability along time. Fig. 6 illustrates the results of the optimal PLS-DM class

model, showing compliance of samples according to both the PFM probability density and *Q* statistics.

Table 1. *Conditions corresponding to the points on the Pareto front.*

Pre-processing	*k*	*a*	LV	% Sensitivity	% Specificity
Autoscaling	2	0.8	1	97.6	92.1
Scaling	**1**	**0.6**	**2**	**96.4**	**93.1**
Scaling	**1**	**0.7**	**2**	**96.4**	**93.1**
Scaling	**1**	**0.8**	**2**	**96.4**	**93.1**
Mean-centering	4	0.6	4	92.8	95.4
Mean-centering	4	0.7	6	91.6	97.1
Mean-centering	3	0.7	7	86.7	97.5
No pre-processing	1	0.7	8	83.1	97.7

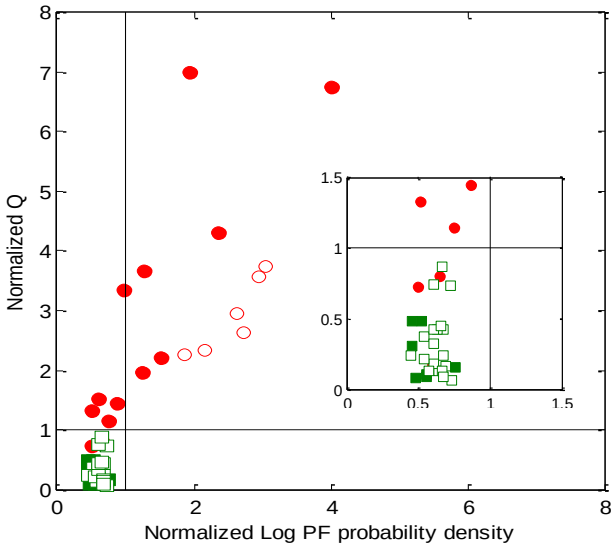

Fig. 6. *Sample compliance with the model according to the probability density and the residual Q statistics criteria. Taggiasca samples are represented by solid (test set) and empty (external test set) green squares, while samples from other cultivars are represented by solid (test set) and empty (external test set) red circles.*

On the other hand, results achieved by the PLS-DM method were compared with those obtained by application of well-established class-modeling methods (i.e., SIMCA and UNEQ) and of OC-PLS on the same data sets. Models obtained with all of these methods were optimized by the same cross-validation scheme used for PLS-DM (data not shown) and the final models were validated on the same test sets. The outcomes of all the optimal models are summarized in Table 2.

As it can be noticed looking at the first two columns of Table 2, PLS-DM was able to provide more balanced results, in terms of cross-validation sensitivity and specificity, compared with those obtained by the other modeling techniques studied. Also considering validation on the test sets, overall higher efficiency value was obtained by PLS-DM. In particular, results achieved on the external test set reflect the stability of the model developed overtime, showing 100% of both sensitivity and specificity for PLS-DM.

Table 2. *Results obtained in cross-validation and on the two test sets with different class-modeling methods. (SENS, sensitivity; SPEC, specificity).*

	PLS-DM[a]		**SIMCA[b]**		**UNEQ[c]**		**OC-PLS[d]**	
	% SENS	% SPEC	% SENS	% SPEC	% SENS	% SPEC	% SENS	% SPEC
Training (CV)	96.4	93.1	65.1	97.7	96.4	84.6	95.2	91.3
Test set	100	91.7	77.8	100	100	83.3	100	83.3
External test set	100	100	63.2	100	100	100	73.7	100

[a] Data were scaled and the model was built with $k=1$, $L = 2$ and $a=0.7$.
[b] Data were autoscaled and the model was built with 2 PCs.
[c] Data were autoscaled and the model was built with 3 PCs.
[d] Data were not pre-processed and the model was built with $L = 10$.

5. Conclusions

The new class-modeling method described in the present study combines the supervised features of PLS regression, the efficient modeling power of potential function techniques and the analysis of residuals, typical of the SIMCA method. Application on a complex data set of NIR spectra of olives in brine showed that the new strategy described is appropriate for authentication of cultivar *Taggiasca*, being able to provide highly efficient models. Elevate balance between sensitivity and specificity values was achieved, as well as high stability overtime if compared with classical class-modeling methods.

The effect of parameters that can be adjusted in the model was thoroughly studied, evidencing the key role of data preprocessing (basic step in any chemometric treatment) and of the proper setting of the smoothing coefficient *a*. At the same time, models developed by the PLS-DM method showed to be robust and stable towards soft perturbations of such parameters. Therefore, PLS-DM can be considered as a foremost untargeted method to be efficiently used for modeling complex analytical data.

6. Acknowledgments

Financial support by the Italian Ministry of Education, Universities and Research (MIUR) is acknowledged – National Project PRIN 2010AXENJ8 (CUP: D31J12001090001). M.I López thanks the Rovira i Virgili University for providing her doctoral fellowship (2011BRDI-06-09). Dr. Xu Lu (China Jiliang University) is acknowledged for kindly providing OC-PLS Matlab routines.

7. References

[1] P. Oliveri, G. Downey, *TrAC - Trends Anal. Chem.* 35 (2012) 74–86.

[2] P.B. Harrington, *Anal. Chem.* 86 (2014) 4883–4892.

[3] M. Forina, P. Oliveri, S. Lanteri, M. Casale, *Chemom. Intell. Lab. Syst.* 93 (2008) 132–148.

[4] R.G. Brereton, *J. Chemom.* 25 (2011) 225–246.

[5] M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, *Food Chem.* 147 (2014) 177–81.

[6] M. Barker, W. Rayens, *J. Chemom.* 17 (2003) 166–173.

[7] Z. Wang, P. Chen, L. Yu, P. de B. Harrington, *Anal. Chem.* 85 (2013) 2945–53.

[8] C. V. Di Anibal, M. Odena, I. Ruisánchez, M.P. Callao, *Talanta.* 79 (2009) 887–92.

[9] M.C. Ortiz, L. Sarabia, R. García-Rey, M.D.L. de Castro, *Anal. Chim. Acta.* 558 (2006) 125–131.

[10] R. Díez, L. Sarabia, M.C. Ortiz, *Anal. Chim. Acta.* 585 (2007) 350–60.

[11] B. Álvarez-Sánchez, F. Priego-Capote, J. García-Olmo, M.C. Ortiz-Fernández, L.A. Sarabia-Peinador, M.D. Luque de Castro, *J. Chemom.* 27 (2013) 221–232.

[12] N. Rodríguez, M.C. Ortiz, L.A. Sarabia, A. Herrero, *Anal. Chim. Acta.* 657 (2010) 136–46.

[13] L. Xu, S.-M. Yan, C.-B. Cai, X.-P. Yu, *Chemom. Intell. Lab. Syst.* 126 (2013) 1–5.

[14] L. Xu, S.-M. Yan, C.-B. Cai, X.-P. Yu, *Food Anal. Methods.* 6 (2013) 1568–1575.

[15] L. Xu, D.-H. Deng, C.-B. Cai, *J. Agric. Food Chem.* 59 (2011) 10461–9.

[16] D.-H. Deng, L. Xu, Z.-H. Ye, H.-F. Cui, C.-B. Cai, X.-P. Yu, *J. Am. Oil Chem. Soc.* 89 (2012) 1003–1009.

[17] M. Forina, C. Armanino, R. Leardi, G. Drava, *J. Chemom.* 5 (1991) 435–453.

[18] C. Armanino, M.C. Casolino, M. Casale, M. Forina, *Anal. Chim. Acta.* 614 (2008) 134–42.

[19] P. Oliveri, V. Di Egidio, T. Woodcock, G. Downey, *Food Chem.* 125 (2011) 1450–1456.

[20] P. Oliveri, G. Downey, *Compr. Anal. Chem.* 60 (2013) 317-338.

[21] M.P. Derde, D.L. Massart, *Anal. Chim. Acta.* 184 (1986) 33–51.

[22] Q. Zang, D.A. Keire, R.D. Wood, L.F. Buhse, C.M. V Moore, M. Nasr, et al., *Anal. Chem.* 83 (2011) 1030–9.

[23] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.

[24] J.E. Jackson, G.S. Mudholkar, *Technometrics.* 21 (1979) 341–349.

[25] D. Coomans, D.L. Massart, I. Broeckaert, A. Tassin, *Anal. Chim. Acta.* 133 (1981) 215–224.

[26] B.W. Silverman, Density Estimation for Statistics and Data Analysis, first ed., 1986, Chapman and Hall, New York.

[27] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, 1998.

[28] M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, *Anal. Chim. Acta.* 827 (2014) 28–33.

[29] S. O'Hagan, W.B. Dunn, M. Brown, J.D. Knowles, D.B. Kell, *Anal. Chem.* 77 (2005) 290–303.

[30] R.J. Barnes, M.S. Dhanoa, S.J. Lister, *Appl. Spectrosc.* 43 (1989) 772–777.

[31] P. Oliveri, M.C. Casolino, M. Casale, L. Medini, F. Mare, S. Lanteri, *Anal. Chim. Acta.* 761 (2013) 46–52.

3. 4. References

[1] R. A. Fisher, *Ann. Eugen.* **1936**, *7*, 179.

[2] R. G. Brereton, *J. Chemom.* **2011**, *25*, 225.

[3] Q. Zang, D. A. Keire, R. D. Wood, L. F. Buhse, C. M. V Moore, M. Nasr, A. Al-Hakim, M. L. Trehy, W. J. Welsh, *Anal. Chem.* **2011**, *83*, 1030.

[4] M. P. Derde, D. L. Massart, *Anal. Chim. Acta* **1986**, *184*, 33.

[5] P. de B. Harrington, *Anal. Chem.* **2014**, *86*, 4883.

[6] M. Chica, P. Campoy, *J. Food Eng.* **2012**, *112*, 50.

[7] R. G. Brereton, G. R. Lloyd, *Analyst* **2010**, *135*, 230.

[8] L. Xu, S.-M. Yan, C.-B. Cai, X.-P. Yu, *Chemom. Intell. Lab. Syst.* **2013**, *126*, 1.

[9] L. Xu, C.-B. Cai, D.-H. Deng, *J. Chemom.* **2011**, *25*, 568.

CHAPTER 4

VALIDATION PROTOCOLS FOR MULTIVARIATE QUALITATIVE ANALYSIS

4. 1. Introduction

4.1.1. *State-of-the-art*

Validation is a basic requirement to guarantee the quality, and trustworthiness of the results obtained by any analytical method. Although it is assumed as a key point for any analytical method, there is little literature on qualitative methods and it is mainly focused on univariate approach.

Since most of the research focuses on developing new 'in-house' analytical methods, the state-of-the-art will also focus on the validation proposals, guidelines and/or protocols of those methods.

A bibliographic search of the last ten years shows how keywords as 'validation protocol', 'validation method' or 'validation guideline' are increasingly found within the scientific contribution's titles. Figure 4-1 shows the rise of publication in the chemical field by the last ten years. This search has focused on validation, understood as quality assurance.

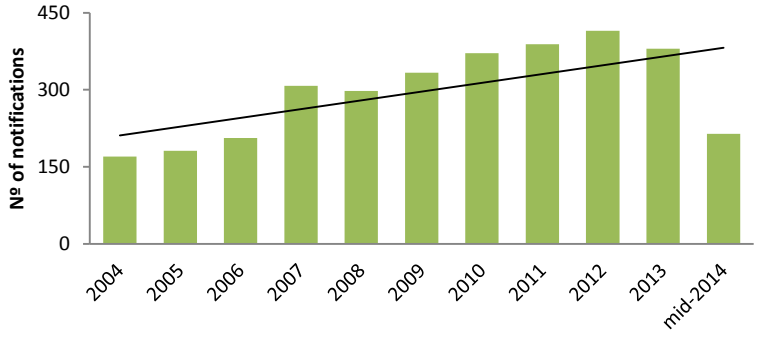

Figure 4-1. Number of scientific communication obtained through a search on the scientific database *ISI Web of Knowledge (WoK)* using 'validation' related words in the title of the scientific communications.

Based on above results, a deep search has been carried out adding specific keywords to determine the percentage referred to quantitative and qualitative analysis, specifying in each case the univariate and multivariate ratio. Figure 4-2 depicts the results in a pie chart format, showing that method validation studies, guidelines and procedures have focused mainly on quantitative methods of analysis although several efforts have also been made to develop validation protocols for qualitative ones. Nonetheless, there are a big percentage of results that after adding specific keywords, they are not included to any of four specified categories. They have been enclosed in the category called 'other'.

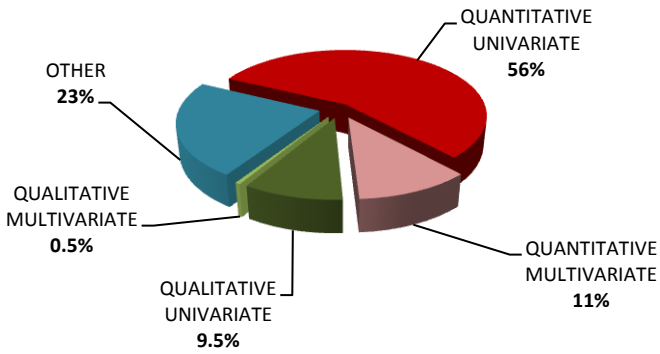

Figure 4-2. Percentage of scientific publications (between 2004 and mid-2014) obtained by searching the scientific database *ISI Web of Knowledge* (WoK).

Since the objective of this chapter focuses on the multivariate qualitative validation, a deeper check of the references found in this category was carried out. The results showed that papers use the term 'validation' to refer the use of training/test sets or cross-validation strategies. Some recent examples can be found in the literature **[1,2]**.

4.1.2. *Background*

A TUTORIAL ON THE VALIDATION OF QUALITATIVE METHODS: FROM THE UNIVARIATE TO THE MULTIVARIATE APPROACH

M.I. López, M.P. Callao, I. Ruisánchez

Analytica Chimica Acta, **Submited**

Analytica Chimica Acta, Submitted

A tutorial on the validation of qualitative methods: from the univariate to the multivariate approach

M. Isabel López, M. Pilar Callao, Itziar Ruisánchez

Chemometrics, Qualimetric and Nanosensors Group, Department of Analytical and Organic Chemistry, Rovia i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

ABSTRACT

This tutorial provides an overview of the validation of qualitative analytical methods, with particular focus on their main performance parameters, for both univariate and multivariate methods. We discuss specific parameters (sensitivity, specificity, false positive and false negative rates), global parameters (efficiency, Youden's index and likelihood ratio) and those parameters that have a quantitative connotation since they are usually associated to concentration values (decision limit, detection capability and unreliability region).

Some methodologies that can be used to estimate these parameters are also described: the use of contingency tables for the specific and global parameters and the performance characteristic curve (PCC) for the ones with quantitative connotation. To date, PCC has been less commonly used in multivariate methods.

To illustrate the proposals summarized in this tutorial, two cases study are discussed at the end, one for a univariate qualitative analysis and the other for multivariate one.

***Keywords**: Method validation, Qualitative analysis, Screening, Multivariate classification techniques, Performance parameters*

1. Introduction

Laboratories have to guarantee the quality and trustworthiness of the results of any analytical method. Then, its validation is fundamental to ensure the reliability, traceability or comparability of results.

Most analytical problems require the amount of one or more substances present in a sample to be determined (quantitative analysis). Other analytical problems require semi-quantifiable or non-quantifiable information: i.e. to authenticate a substance/product or verify if a substance is present above or below a pre-established concentration level (cut-off value). In these cases, using qualitative methods that provide a binary response (positive/negative) might be suitable. They have commonly been used in systems that require immediate decisions to be taken since they are an appealing alternative to quantitative analysis, which generally gives more but often unnecessary sample information and requires a greater investment of money and/or time. For some time now, qualitative methods have been increasingly developed and applied in such fields as clinical medicine, biology and chemistry [1–4].

The performance of quantitative methods has been the subject of numerous studies, which have resulted in the production of international guidelines. By contrast, there is still no consensus about the validation protocol and the terminology used for qualitative methods. Several authors have tried to make proposals or guidelines about various aspects of the validation of qualitative methods using the information available in the literature [5–7]. In 2005, the International Union of Pure and Applied Chemistry (IUPAC) promoted a project that aimed to draft an internationally harmonized protocol (guidelines) for the organisation and interpretation of collaborative trials for the validation of qualitative methods [8]. All the effort that has been made (and is still being made) focuses mainly on univariate analytical methods whereas the multivariate ones are hardly developed.

This tutorial presents an overview about the validation of qualitative methods, both univariate and multivariate, focusing on the performance parameters and the strategies used to establish them.

2. Fundamentals: general terminology

2.1. Method Validation

The development and validation of a method are closely related since performance parameters are often evaluated as part of method development. When approaching an analytical problem, analysts have to consider several issues, which are schematised in Fig. 1 **[9,10]**. The problem's solution must be regarded as a cyclic and iterative process of checking and evaluating the method, which does not stop until the method is deemed capable of meeting the requirements. The process starts with the study of the analytical problem at hand, what is known about it and what the analytical requirements are. The analytical method that best responds to these requirements must be chosen. When no existing analytical method responds to the requirements, then an existing methodology has to be redesigned or a new one developed. Before a method is validated, it must first be assessed whether it satisfies the requirements (fits the purpose) or not. If it does, the method has to be validated. The method is considered to be fully validated when all the requirements are satisfied and the whole process has been documented.

According to the Handbook for the Quality Assurance of Metrological Measurements, *"method validation consists of documenting the quality of an analytical procedure, by establishing adequate requirements for the performance criteria, such as accuracy, precision, detection limit, etc. and by measuring the values of these criteria"* **[11]**. Thus, the validity of a method must be proven in its documentation, which must describe how the method is performed, which parameters have been investigated during the validation process, what the results of the validation study are.

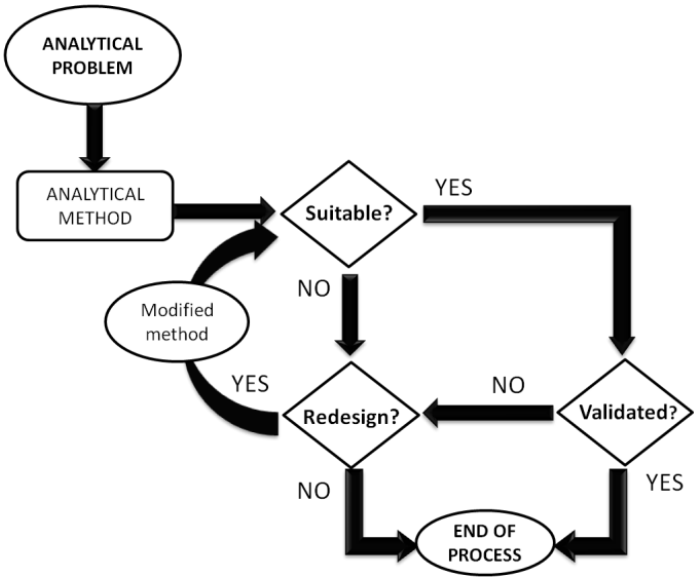

Fig.1. *Diagram of method development and validation (adapted from EUROCHEM The Fitness for Purpose of Analytical Methods [9]).*

In ISO/IEC 2005, method validation became the *"confirmation by examination and provision of objective evidence that the particular requirements for a specified intended use are fulfilled"* **[12]**. In general terms, it establishes the concept of 'fitness-for-purpose' since it evaluates the fitness of the analytical method for its purpose. As a result, the performance characteristics to be established depend on the requirements of the analytical problem.

In this regard, validation is considered as the process of ensuring that an analytical procedure is reliable and can fulfil the expectations of a particular application. In short, it means that it can be used with confidence.

2.1.1. Validation level

Depending on the needs of the laboratory, several levels of validation can be considered.

Single-laboratory method validation: Laboratories have to take internal measures to ensure the quality of the data they provide. There are several circumstances in which internal validation is carried out: to ensure the viability of methods developed in-house, to assess a method developed in other laboratories and to estimate the quality of long-term results. Internal quality control is also considered as internal validation **[13]**.

Most of the research work carried out has been on single-laboratory validation since laboratories are continuously modifying and improving methodologies to achieve, for example, lower detection limits, to consider novel interferences or to reduce time and costs. This tutorial will also focus on single-laboratory validation.

Interlaboratory method validation: Different laboratories agree to carry out the same analytical trial under the supervision of a coordinator, who sets out the goals, the conditions and, obviously, the parameters to be studied. The aims of an interlaboratory study can be: (1) to assess the performance of an analytical method and (2) to compare laboratories. The process followed is schematized in Fig. 2.

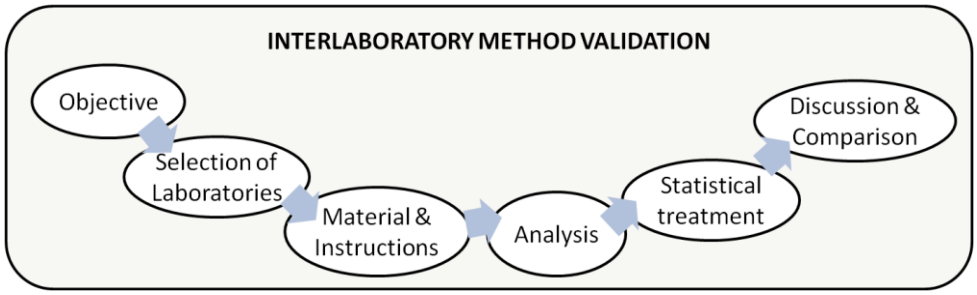

Fig.2. *Diagram of the steps followed for an interlaboratory trial.*

Collaborative studies are the best way to assess and verify the quality of the work done by a laboratory on the validation of a method. They can be used only after the method has already been fully validated in single-

laboratory trials [10,14]. These studies enable all of the participants to determine parameters such as bias, precision or robustness and compare them with statistically assessed results.

Generally, higher levels of validation require a greater investment of time and money, and the final results are of greater quality in terms of trustworthiness, reliability and consistency. So, it is fundamental to decide what level of validation is the most suitable.

2.1.2. Performance parameters

Performance parameters are a set of measurable attributes that define the quality of an analytical method. Thus, methods must be validated by establishing their performance parameters, which depend on the type of analytical method.

Quantitative performance parameters are established on the basis of statistical fundamentals. Since qualitative methods are based on binary response (positive/negative), their performance parameters cannot be established using the same fundamentals. Instead, they have to be established on the probabilities that arise from four possible binary response scenarios [5]:

- True positive (TP) result, when the qualitative method gives a positive output for a sample that is positive.

- False positive (FP) result, when the qualitative method gives a positive output for a sample that is negative.

- True negative (TN) result, when the qualitative method gives a negative output for sample that is negative.

- False negative (FN) result, when the qualitative method gives a negative output for a sample that is positive.

Table 1 shows the performance parameters considered in both qualitative and quantitative analyses. Depending on (1) the nature of the analytical problem, (2) the analytical method purpose or (3) the level of validation; the required performance parameters to be estimated could be several or all of the parameters presented in Table 1. It should be pointed out that quantitative performance parameters will not be discussed in this tutorial.

Table 1. *Quality performance parameters.*

Quantitative	Qualitative
✓ Accuracy: **trueness**, precision	✓ **Trueness**
✓ Uncertainty	✓ False positive (FP) and false negative (FN) rates
✓ Sensitivity and specificity	✓ Sensitivity and specificity
✓ Range and linearity	✓ Efficiency, Youden's Index and Likelihood ratio
✓ Limits: limits of detection/quantification	✓ Limits: Decision limit/Detection capability and other related terminology
✓ **Selectivity/interferences**	✓ Unreliability region
✓ **Ruggedness or robustness**	✓ **Selectivity/interferences**
✓ **Stability**	✓ **Ruggedness or robustness**
	✓ **Stability**

The parameters in bold are evaluated in the same way in both types of analysis. They have been extensively defined in EC/657/2002 **[15]**. For instance, trueness is achieved by using certified reference material or, when this is not possible, another reliable reference.

Qualitative parameters that are not in bold are derived directly or indirectly from binary response and, therefore, from the four scenarios

defined above. Worthy of special mention are the underlined parameters which have the same name in both types of analytical method although the concepts they represent and their evaluations are slightly different.

Sensitivity, in quantitative methods, indicates how the response changes when the analyte concentration varies whereas in qualitative methods it refers to the ability of the method to recognise truly positive samples (and so is directly related to the TP response).

The same occurs with the term specificity. In quantitative methods, it refers to the ability of a method to distinguish between the analyte being measured and other substances whereas in qualitative methods it is the ability of a method to detect truly negative samples (and so is directly related to the TN response).

False positive and false negative rates assess the probability of error, which is directly related to the FP and FN responses, respectively. They complement sensitivity and specificity (i.e., *sensitivity = 1- false negative rate*). On the other hand, efficiency, Youden's Index and the likelihood ratio assess the overall suitability of the method since they are a combination of true responses (directly related to the parameters of sensitivity and specificity). Further information will be given in section 3 and Table 5.

According to 2002/657/EC, detection capability (CCβ) is *"the smallest amount of a substance that can be reliably detected, identified and/or quantified in a sample with a statistical certainty of 1 – β"* **[15]**. β error is "*the probability that the tested sample is truly non-compliant, even though a compliant measurement has been obtained (false compliant decision)"* **[15]**. This probability of error is usually set at 5% (significance level 0.05). To clarify concepts, let us consider the case of a contaminant regulated by legislation. For instance, a sample is non-compliant when the contaminant is present and compliant when it is not. Therefore, CCβ is

the concentration limit at which the qualitative method detects the contaminant (it is present) with a 5% of error of stating that the contaminant is not present when in fact it is (false compliant decision or false negative result). Belter et al. **[16]** has recently published a review of the wide range of terms used for this parameter, the most widespread of which is "limit of detection".

According to 2002/657/EC, decision limit (CCα) means "*the limit at and above which it can be concluded with an error probability of α that a sample is non-compliant"* **[15]**. *α* error is also usually set at 5% (significance level 0.05) and it is defined as *"the probability that the tested sample is compliant, even though a non-compliant measurement has been obtained (false non-compliant decision)"* **[15]**. Following the same example discussed above, CCα is the concentration limit at which the qualitative method detects the contaminant (it is present) with a 5% of error of stating that the contaminant is present when in fact it is not (false non-compliant decision or false positive result). This parameter is also referred to as threshold, cut-off, critical value, limit of detection **[16,17]**. Note that this tutorial shall refer to decision limit with the abbreviation *DL*.

It should be pointed out that in the literature the term limit of detection is used to refer to both CCβ and DL. This may be due to the fact that in quantitative analysis it is defined as "*the smallest amount or concentration of analyte in the test sample that can be reliably distinguished from zero"* **[13]**. It is usually estimated by simultaneously considering both probabilities of committing error, *α* and β. In qualitative analysis, due to the particularity of binary response, the probabilities of false positive (error *α*) and false negative (error β) are independent. As a consequence, two concentration limits are defined, one for each kind of error. The one referred to as limit of detection will depend on what is the main interest: to restrain the errors of saying that a sample contains a substance when in fact it does not (false compliant) or to restrain the

errors of saying that a sample does not contain a substance when in fact it does (false non-compliant).

The unreliability region is defined by the two limits: DL and CCβ. In between those two limits, the probability of making a wrong decision is higher than a fixed percentage, usually 5% (false positive and false negative rates). In this regard, unreliability could be related to uncertainty in quantitative analysis. But, unreliability cannot be considered as dispersion around a value as the response in qualitative analysis is not quantifiable **[18]**.

The unreliability region and limit parameters have a quantitative connotation since they are associated to the amount of substance. For this reason, they cannot be established for qualitative analysis based on categorical propriety such as a food authentication problem. Further information will follow in section 3.

2.2. Qualitative analysis

Qualitative analysis has been defined by several recognised international organisations. The International Union of Pure and Applied Chemistry (IUPAC) stated that it is *"the analysis in which substances are identified or classified on the basis of their chemical or physical properties"* **[19]**.

Other organisms such as the U.S. Food and Drug Administration (FDA) and the Association of Official Analytical Chemists (AOAC International) have reformulated the definition. Thus, qualitative analysis is a "*method in which substances are identified or classified on the basis of their chemical, biological or physical properties. Its response is either the presence or absence of the analyte(s) in question, detected either directly or indirectly in a specified test"* **[20,21]**.

As can be easily inferred from the definitions, qualitative analysis is characterised by its binary response (positive/negative outputs). Although

it is not specifically mentioned in the official definitions, qualitative analysis might also be related to a categorical propriety of the samples instead to the presence or absence of an analyte(s).

Nowadays, it is quite usual to use 'screening method' as a synonym for 'qualitative method' even though this may not always be the case. Screening methods *"are used to detect the presence of a substance or class of substances at the level of interest. These methods have the capability for a high sample throughput and are used to sift large numbers of samples for potential non-compliant results. They are specifically designed to avoid false compliant results".* Legislation recommends a false compliance rate lower than 5% (1% for banned substances and 5% for substances with a maximum permitted level) **[15]**. Thus, they generally involve short analysis times, which lead to a high throughput of samples at low cost and mean that they are suitable for routine analysis. In this scenario, non-compliant samples are usually submitted to confirmatory analysis if the specific amount of substance present in the sample needs to be known. This requires an in-depth study of the sample, which is time and cost consuming.

Depending on the nature of the data used, we will refer to univariate or multivariate methods:

Univariate qualitative methods provide a binary response from only one analyzed variable. This variable should provide enough information to solve the analytical problem at hand (Fig. 3a) **[22]**. The binary response can be obtained from a specific instrumental signal (i.e., an absorbance value at specific wavelength). It can also be obtained through visual observation of colour change or development (i.e., test kits which are prepared to detect/identify substances at a specific threshold concentration).

In the same way, multivariate qualitative methods provide a binary response from two or more analyzed variables (Fig. 3b) [22]. These variables might come from an instrumental signal (for example, an absorbance spectrum, hence, a vector of absorbance values recorded in a defined wavelength range) or from a non-instrumental signal (for example, sensory panels). Since the analysed variables are non-specific, a data treatment step is always required to obtain the binary response. Data are treated using chemometric tools, mainly by the application of classification techniques.

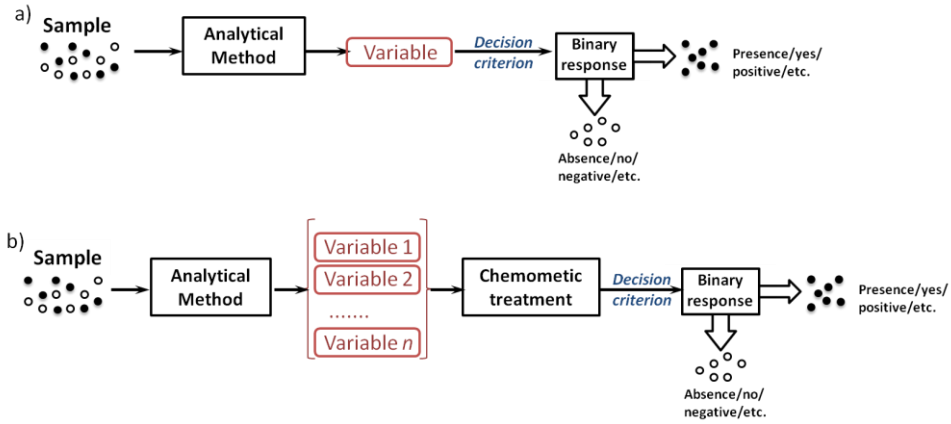

Fig. 3. *Scheme of a qualitative analysis using (a) univariate and (b) multivariate approaches.*

Regardless of the number of variables analyzed, binary responses — positive/negative outputs — are obtained using a decision criterion, which can be related to either a quantifiable value or to a categorical property (see Table 2). In the first case, there is a threshold, generally imposed by regulation or clients. In the absence of any other verifiable information, it is set by the analysts on the basis of their knowledge of the analytical problem.

A common qualitative analysis (univariate or multivariate) is the detection of an analyte above/below a threshold concentration. One particular case

is the detection and identification of compounds, which imply the presence/absence of the analyte. In this case, the threshold is set at concentration value equal zero.

Another goal of the qualitative analysis is related to a sample compliment/not compliment regarding any intrinsic property (categorical property), i.e. authentication problem: sample assignation to a protected designation of origin (PDO). Since the decision criterion is not related to a quantifiable value, it has no associated threshold value. This kind of analysis has to be carried out using the multivariate approach.

Table 2. *Types of qualitative analysis.*

Focus of the study	Decision Criterion	Threshold	Data	Examples
Analyte/Index	Related to quantifiable value	Above/below a certain value ≠ 0 Presence/absence (value = 0)	Univariate/ Multivariate	Maximum permissible amount (e.g. Content of biodiesel in diesel) Banned substances (e.g. doping in sport)
Sample	Related to categorical property	-	Multivariate	Authentication (e.g. Protected Designation of Origin, PDO)

3. Univariate qualitative analysis

Univariate qualitative analyses are used to detect a substance or group of substances. Thus, the decision criterion — how positive and negative responses are defined — is directly related to a threshold concentration. The analysis of samples that have the substance at the threshold level will provide a specific signal (threshold signal): (i) if the signal is instrumental, the output is generally positive when the *sample signal* ≥

threshold signal (i.e., contaminated sample) whereas the output is negative when the *sample signal* < *threshold signal* (i.e., non-contaminated sample); ii) if the signal is visual, the output is positive when colour is developed while the output is negative if there is no colour, or the other way around.

A full validation process consists of estimating both mandatory and any other performance parameters by analysing truly positive and negative samples (Table 3). The parameters related to concentration limits are established after mandatory performance parameters achieve satisfying values.

Table 3. *Summary of the main methodologies that can be used to evaluate a qualitative method.*

Validation step	Performance Parameters	Methodology
Mandatory	FP and PN rates, sensibility and specificity	Bayes' Theorem
		Statistical Hypothesis Test
	Efficiency, Youden's Index and Likelihood ratio	Contingency Table
(if applicable)	Unreliability region, Limits ($x_{cc\beta}$, x_{DL})	Performance characteristic curves
		Statistical Hypothesis Test

Several methodologies can be used to estimate the mandatory performance parameters: Bayes' theorem, statistical hypothesis tests and contingency tables. They provide an overall characterization of the qualitative method at a specific–concentration level (static situation), and they all estimate the same performance parameters. Nonetheless, the terminology in each case is slightly different.

This tutorial will explain how contingency tables can be used to estimate mandatory performance parameters. Detailed information on Bayes' theorem **[23]** and statistical hypothesis tests **[24,25]** can be found elsewhere.

Contingency tables are based on Bayes' theorem and are widely used since they are easy to work with and can be applied to solve any qualitative analytical problems. They consist of a 2x2 table which is obtained by analysing actual samples (positive and negative) which are then compared with the outcomes of the qualitative analysis (Table 4). The expressions used to calculate the main performance parameters are shown in Table 5. Sensitivity, specificity, false positive and false negative rates are obtained from the frequencies of each respective response divided by the total number of samples. Global indexes (efficiency, Youden's index and the likelihood ratio) are obtained by a combination of the previous parameters.

If the mandatory performance parameters satisfy the requirements stipulated by the analyst, additional performance parameters can be estimated when the threshold value is related to a quantifiable property (see Tables 2 and 3). This tutorial will explain how *performance characteristic curves* (PCC) can be used to estimate additional performance parameters which provide quantitative information about the qualitative analytical method. To do so, the PCC curves assess the method in a dynamic situation instead of a static situation like other methodologies, and consider values around (above and below) the pre-set threshold value. Note that in the literature, PCC curves have also been referred to as performance curves **[26]**, logistic regression **[5]** or the generalized linear model **[27]**.

Table 4. *2x2 contingency table.*

Predictions	Actual	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN
Total analysed samples	TP+FN	FP+TN

TP, true positive; TN, true positive; FP, false positive; FN, false negative

Table 5. *Description of the performance parameters.*

Sensitivity	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
False negative (FN) rate ***(1-sensitiviy)***	$\frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}}$
Specificity	$\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
False positive (FP) rate ***(1-specificity)***	$\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}$
Efficiency	$\frac{\mathrm{TN}+\mathrm{TP}}{\mathrm{TN}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN}}$
Youden's index	$(Sensitivity + Specificity - 1)$
Likelihood ratio	$\frac{Sensitivity}{1-Specificity}$

To illustrate how PCC curves are used, let us consider the case of a contaminant which is regulated by legislation, so it has a threshold value (i.e. maximum content 5mg/L). This value leads to an instrumental *threshold signal*, which is obtained experimentally. Positive output is defined as a contaminated sample when *sample signal > threshold signal* and negative as a non-contaminated sample when *sample signal* $\leq$ *threshold signal*. To establish the PCC curve, samples with concentrations around *5* mg/L (i.e. 3, 4 and 6, 7 mg/L) are considered. Several samples are analyzed for each level concentration. The probability of getting a positive result, *P(x)*, is obtained by the frequency of positive outputs for each concentration studied. The experimental PCC curve is obtained by representing the probabilities of positives, *P(x)*, versus the corresponding concentration. As Fig. 4a shows, the ideal qualitative analysis would be 100% sure to give a positive response, *P(x)*, when the amount of contaminant is > 5 and a negative response, *N(x)*, when the amount of contaminant is $\leq$ 5.

Real behaviour, however, is different from the ideal one (see Fig. 4b). The experimental *P(x)* values are fitted to a sigmoid function, minimizing the root mean square of the residuals, obtaining the PCC curve. To obtain the concentration limits two horizontal lines have to be drawn. The upper line is usually set at *P(x)*=95 % and the lower line at *P(x)*=5%.

From the intersection of those horizontal lines with the PCC curve, the concentration limits are obtained:

- $x_{cc\beta}$ *(detection capability)*: it is obtained from the intersection between the lower horizontal line — which corresponds to the probability of committing an FP error, *P(x) 5%* — and the sigmoid curve.

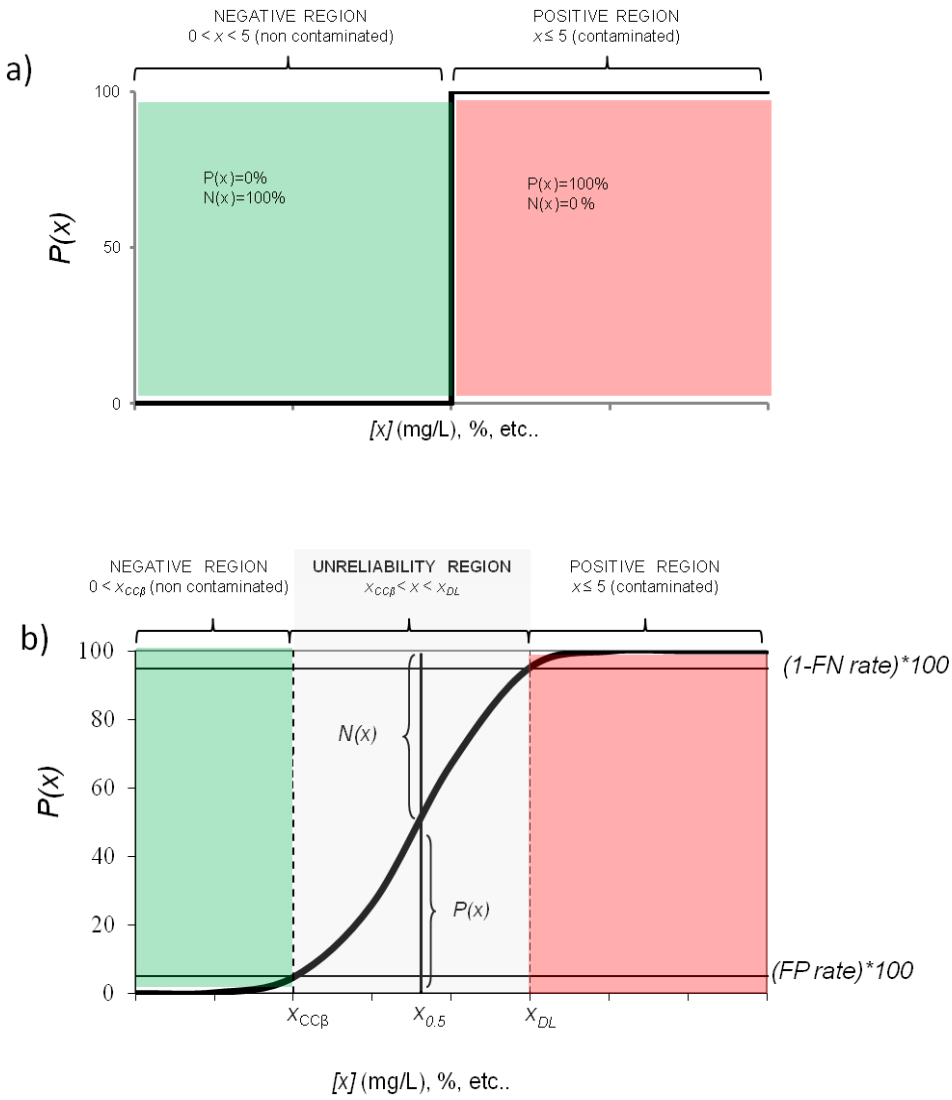

Fig.4. *(a) Ideal graph and (b) real graph for a binary response whose decision criterion corresponds to a threshold value ≠ 0. P(x): probability of getting a positive response, N(x) =100- P(x): probability of getting a negative response, FN rate: false negative rate, FP rate: false positive rate, x_{DL}: Decision Limit, $x_{CC\beta}$: Detection capability and $x_{0.5}$: when P(x) = N(x)= 50%.*

- x_{DL} *(decision limit)*: it is obtained from the intersection between the upper horizontal line — which corresponds to the probability of committing an FN error, *P(x)= 95%=1-FN* — and the sigmoid curve.

- *unreliability region*: it is the region between the two previously defined limits where there is the probability of false compliance.

The region $0 < x < x_{cc\beta}$ is considered the reliable negative region since at concentrations lower than $x_{cc\beta}$ ($x < x_{cc\beta}$) the probability of getting a true negative result is > 95%, and the rate of getting a false negative result decreases (lower than 5%). Similarly, the region $x > x_{DL}$ is considered the reliable positive region since there is a 95% or higher probability of getting a true positive result at concentrations higher than x_{DL}.

4. Multivariate qualitative analysis

As mentioned before, multivariate analysis is required when the problem at hand cannot be solved by a specific measurement. Thus, multivariate qualitative methods can also be applied to detect the presence of a substance or a group of substances. The analytical problem previously described in the univariate analysis — to detect whether a sample is contaminated or not, in accordance with a threshold value set by legislation (5 mg/L) — can be, hence, solved from a multivariate point of view. Using multivariate analysis, two categories have to be defined: category A (sample is compliant or not contaminated; concentrations $\leq$ 5 mg/L) and category B (sample is not compliant or contaminated; concentrations $>$ 5 mg/L).

However, multivariate approaches are commonly used to solve analytical problems related to an intrinsic property of the samples, which is usually known as categorical property. Thus, the decision criterion — how positive and negative responses are defined — is not directly related to a

threshold concentration but to a sample belonging to a predefined category (this is also known as class assignation). As example, the detection of cancerous tissue, quality control of a manufacturing process or the determination of the origin of a wine among some other examples.

Let us take as example a wine authentication problem (i.e., PDO Priorat wine). Hence, the compliant category (category A) is defined as 'Priorat wine'. Depending on how the non-compliant category (category B) is defined, this authentication problem can be tackled from two different points of view:

- *unspecific category B:* In this case, the analyst has a lack of knowledge of the kind of wine that could be use to commit fraud. Thus, the non-compliant category is not focused on specific type of samples, thus, it is defined as 'not Priorat wine'.

- *specific category B*. The difference compared to previous case is that the analyst is awarded of the kind of fraud that can be committed. So, the analyst knows that another similar and usually cheaper wine (i.e. Montseny wine) can be dishonestly labelled as a PDO Priorat. In this case, the non-compliant category is well defined by Montseny wine samples, thus, it is defined as 'Montseny wine'.

Regardless the problem to be solved, the process followed to perform multivariate qualitative analysis implies three steps: (1) sampling and analysis, (2) classification rule and (3) validation. Mention to wine examples will be done in order to facilitate understanding.

1) sampling and analysis.

The sampling must be representative of the total population to ensure accuracy in the result. Thus, samples representative of both category A (i.e., Priorat wine) and category B (i.e., Montseny wine) must be collected. When the category B is unspecific (i.e., 'not Priorat wine'), it is

hardly representatively sampled because no set can include all the different sorts of wine from around the world that are not Priorat. Thus, 'not Priorat wine' will be always under representative.

To guarantee representativeness in the category, several factors should be taken into consideration during the sampling (i.e., harvests, cellars, wine-ageing, among others). A minimum number of samples (i.e, 20 - 30) must be selected for each category, depending on the availability of samples, the cost of analysis, the factors considered, etc. The higher the number of samples, the better the total population is represented, thus, the better are the conclusion obtained from the analysis.

The analysis of all samples is carried out after the sampling process, getting a data vector for each sample. Ranging the vectors, an *n-by-p* data matrix is obtained, where *n* is the sample size, and *p* is the number of variables measured.

2) classification rule

By applying a classification technique, a class for each category is mathematically defined. It is beyond the scope of this tutorial to discuss in detail all the classification techniques available, and detailed information can be found elsewhere **[28–30]**. Unlike the univariate case, the decision criterion is not related to a *threshold signal* but to a *mathematical function* (classification rule) which allow assigning the samples to the predefined categories: sample belongs to category A (positive) or to category B (negative). Fig. 5 depicts the different type of classification rules obtained depending on the classification techniques used.

When a discriminant classification technique is used, the decision criterion is called *delimiter* (Fig. 5a). Both categories A and B have to be defined, obtaining a unique delimiter for sample assignation. When

unknown samples are predicted (☆, Fig. 5a) they are always assigned to one of the predefined categories, 'Priorat' or 'Montseny', even if the analysed wine correspond to neither them. Positive/negative outputs can be defined according to Eq. 1, being *x* a sample:

$$Positive: x \in A\ (sample\ belongs\ to\ category\ A - Priorat\ wine) \qquad Eq.(1)$$

$$Negative: x \in B (sample\ belongs\ to\ category\ B - Montseny\ wine)$$

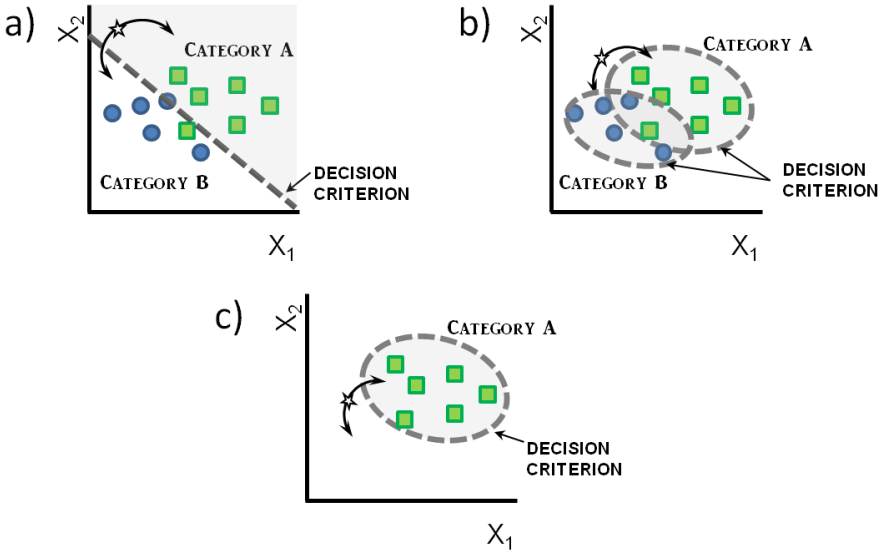

Fig. 5. *Example of decision criterion obtained depending on the type of classification technique used. Discriminate technique (a). Class-modelling technique: two classes (b) and one class (c). Samples belonging to Category A (■), samples belonging to Category B (●) and unknown sample used in prediction (☆).*

When class-modelling techniques are applied, the decision criterion is called *model boundary* (Fig. 5b). Two categories must be defined although a model is built for each category individually by using only the samples belonging to the category. At the end, two models are characterized, each one with different model boundary. When unknown samples are predicted (☆, Fig. 5b), they can be assigned to one

category, to both categories or to neither of them. This kind of classification techniques can be interesting since the analyst can detect a wine which is neither Priorat nor Montseny. If a sample belongs to both categories or to neither of them, the result is considered to be inconclusive. Then, the sample might be submitted to a confirmatory analysis to check which type of wine is, if this information is required. Positive, negative and inconclusive outputs are defined according to Eq. 2, being *x* a sample:

$$Positive: x \in A\ (sample\ belongs\ to\ category\ A - Priorat\ wine) \qquad Eq.(2)$$

$$Negative: x \in B(sample\ belongs\ to\ category\ B - Montseny\ wine)$$

$$Inconclusive: x \in (A \cap B)\ \ or\ \ x \notin (A \cup B)$$

A particular case of class-modelling techniques is when only one class is modelled, either because the analyst is interested in characterizing only one category or because only samples from one of the categories can be collected. This could be a good option to tackle the authentication PDO wine problem when dealing with unspecific category B ('not Priorat wine').

Fig. 5c shows an example in which only one category is modelled ('Priorat wine'). When unknown samples are predicted (☆, Fig. 5c), they can be recognised by the model (compliant) or not (non compliant) Thus, positive and negative outputs are defined according to Eq. 3, being *x* a sample:

$$Positive: x \in A\ (sample\ belongs\ to\ category\ A - Priorat - compliant) \qquad Eq.(3)$$

$$Negative: x \notin A(sample\ does\ not\ belong\ to\ category\ A - not\ Priorat - non - compliant)$$

Although the model is built only using samples belonging to the category under study, the specificity must be checked by submitting non-compliant samples to the model.

Note that classification techniques can characterize two or more categories; however, we had focused on those cases in which only two possible outputs (binary response) are considered.

3) validation

In the field of classification techniques, the term *ability* is used to assess the quality of the class assignation. The ability is calculated by dividing the number of samples correctly classified in the category by the total amount of samples of the category. Similarly, the assignation error is obtained by dividing the number of samples misclassified by the total amount of samples of the category. This ability — and error — can be calculated individually for each defined category (i.e., A and B), but also for all samples independently of the category (global ability).

If those abilities are computed from the set of samples used to build the classification rule (training set), the performance parameters are named: classification ability of category A, category B and global. These parameters are often optimistic since they are obtained from the training set (autopredictive). For this reason, predictions of new well-categorised samples (test set) are generally more realistic, being a key step in assessing model success. Depending on the sample size, two strategies can be used to assess predictions:

- If the initial dataset is large enough, it is randomly split into *training and test sets*: the training set is used to build the classification rule and the test set to assess prediction ability. Similarly to the classification ability, three prediction abilities are obtained: prediction ability of category A, category B and global.

- If the initial dataset is not large enough, an alternative is to follow the *cross-validation* strategy to assess the prediction ability. It requires a single dataset — which is the training set. One (or more) samples are removed from the dataset and the model is built with the remaining samples. Then, the prediction ability of the model is tested with the removed samples. This procedure is repeated until all the samples have been left out of the dataset. Cross-validation can be carried out through several strategies: contiguous blocks, leave-one-out, random subsets, etc. **[28]**. Thus, three prediction abilities are obtained: prediction ability of category A, category B and global.

The relationship between the performance parameters terminology used in univariate (Table 5) and multivariate qualitative analysis is the following:

- Sensitivity: It is the classification and/or prediction ability obtained from the category defined as positive (in our example category A, hence, 'Priorat wine'). The key point is that sensitivity is computed from positive samples (positive output).

- Specificity: It is the classification and/or prediction ability obtained from the category defined as negative (in our example category B, hence, 'Montseny wine'). Again, the key point is that specificity is defined from negative samples (negative output).

- Efficiency: It corresponds to the global classification and/or prediction ability.

Note that when dealing with class-modelling techniques in which two classes are modelled (case corresponding to Fig. 5b), this relationship could not be strictly correct if there are samples classified to any of the categories.

On the other hand, if only one category is modelled (case corresponding to Fig. 5c), it is necessary to analyse non-compliant (negative) samples to assess the specificity and the global ability.

Efficiency (Table 5) is not the only term used among the scientific community since it is commonly known as accuracy [31]. Also, some authors use the term efficiency but it is calculated geometrically [32] (Eq.4):

$$\sqrt{\frac{TN \cdot TP}{(TN + FP) \cdot (TP + FN)}} \qquad \text{Eq. (4)}$$

Finally, instead of expressing parameters as abilities (usually in %), other authors prefer to express the parameters as % of error, which is just 1 minus the corresponding ability values [33].

Once the mandatory performance parameters (both autopredictive and predictive) satisfy the requirements, PCC curves can also be used when the problem under study is related to a quantitative value (i.e., adulteration problem). In practice, this has rarely been done because multivariate qualitative analysis is primarily used for the authentication of samples, for which PCC curves cannot be used.

5. Cases study

The first case study deals with a univariate qualitative analysis based on a visual signal obtained from a test kit designed to detect a regulated compound in nuts [17]. Other examples of univariate methods based on visual and instrumental signals can be found elsewhere [2,34–36].

The second case study considers a multivariate qualitative methodology based on spectroscopic signals designed to detect adulterants in nuts

[37]. Other examples of this kind of method can be found in the following references **[33,38–44]**.

5.1. Example 1: Univariate qualitative analysis

The analytical problem to be solve is the determination of Aflatoxin B_1 in fried ready-salted peanuts **[17]**. A commercial test kit was used that had been specially designed for this kind of compound and which gave a visual response. According to the European Union, the maximum concentration of this compound permitted in nuts is 2.0 ng/g. Thus, commercial kit has been developed to indicate a negative result by displaying colour when concentrations are < 2.0 ng/g, and a positive result by not displaying colour when concentrations are ≥ 2 ng/g. Positive samples are submitted to confirmatory analysis.

Several performance parameters were established. From contingency tables: false positive and negative rates, sensitivity and specificity were obtained at the maximum permitted concentration (static situation at 2.0 ng/g). Results were successful, both sensitivity and specificity value were 100%.

To build the PPC curve (Fig. 6), a total of 84 samples containing Aflatoxin B_1 with concentrations ranging from 0.6 to 2.6 ng/g were analysed. The obtained probabilities of positives, *P(x)*, at each studied concentration are fitted to a sigmoid function and the performance parameters are estimated. The detection capability ($x_{cc\beta}$) is set at 0.8 ng/g, indicating that at lower concentrations the probability of truly negative output is equal to or higher than 95%. Similarly, the decision limit (x_{DL}) is set at 1.6 ng/g, indicating that at higher concentrations the probability of truly positive output is equal to or higher than 95%. Therefore, the unreliability region is between 0.8 and 1.6 ng/g. It should be pointed out that the unreliability region is far away from the maximum allowed by law.

In practice, the probability of a real false positive is almost zero since the decision limit is at 1.6 ng/g, much lower than the value claimed by the kit's manufacturer. Since we are dealing with a contaminant food product, it is desirable to control and minimize the error of considering a contaminated sample as non-contaminated. Thus, the bias in the decision limit can be regarded as an advantage. This fact is in accordance with screening method definition which as it has been stated previously (section 2.2), they are specifically designed to avoid false compliant results.

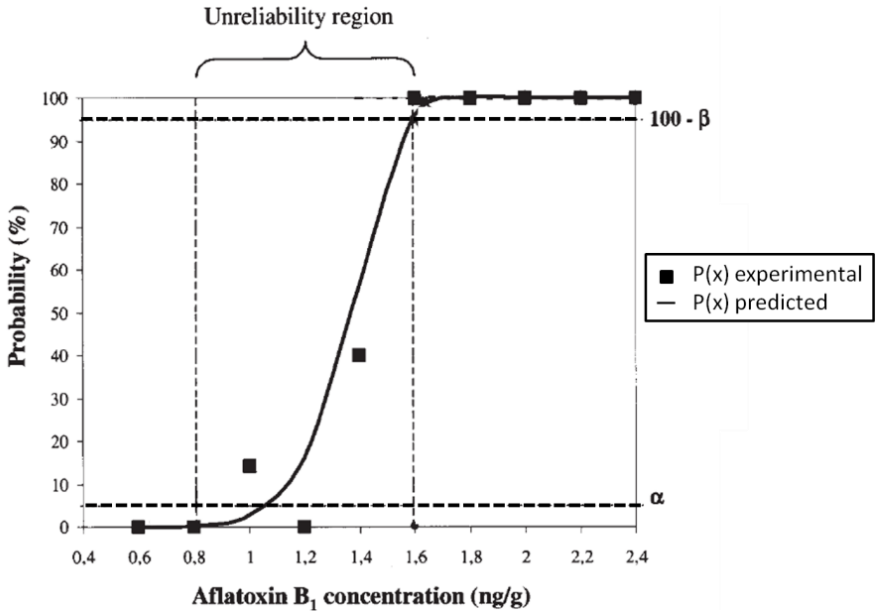

Fig. 6. *Adaptation from [17] with permission. Probability of positive responses, P(x), is plotted versus the concentration levels tested. Upper limit: x_{DL}; Lower limit: $x_{CC\beta}$.*

Considering that the Aflatoxin B_1 contamination has to be a minor fact in the global context of commercialized samples, the majority of the analysed samples should have Aflatoxin B_1 levels below 2.0 ng/g, hence, negative samples. Therefore, using this screening method, the number of samples submitted to confirmatory analysis is highly reduced.

5.2. Example 2: Multivariate qualitative analysis

The analytical problem to be solve deals with the adulteration of hazelnut pastes with other substance that are not contaminants but which are used mainly for economical reasons to reduce the cost **[37]**. The price of hazelnuts depends on the market and it can be kept down by adding such ingredients as almond paste or flour, since it is very similar to hazelnut but usually much cheaper. Another cheap adulterant that can be used is chickpea flour. Although this adulterant is more unexpected, it can be used because its physical properties are similar to those of the hazelnut. Experience shows that the most common percentage of adulteration is around 7%.

The qualitative method combines infrared spectroscopy with a classification technique named soft independent modelling of class analogies (SIMCA). The one-class modelling strategy was used (as described in Fig. 5c), thus, the output is positive when a sample is recognised as model compliant and negative when a sample is recognised as model non-compliant.

To build the model, 28 hazelnut samples from different geographic origins were used. The sensitivity and false negative rate were assessed by cross-validation. The specificity, false positive rate and global parameters were assessed by predicting adulterated samples at 7% (28 adulterated samples with almond and 28 adulterated samples with chickpea). Table 6 shows the performance parameters values. Results show successful sensitivity and specificity results, even when considering almond adulterant, which has similar properties to hazelnut.

Since obtained performance parameters showed successful values, the PCC curves were attempted to be established since this kind of problem allows it. To do so, different percentages of adulteration were studied (1-8% in intervals of one). Finally, around 13 samples were studied at each

adulteration level and for each adulterant, being about 104 the total amount of samples.

Table 6. *Adapted from [37] with permission. Performance parameters obtained for each adulterant, expressed in %. For simplicity, false positive (1-sensitivity) and false negative (1-specificity) rates are omitted.*

	Sensitivity	Specificity	Efficiency	Youden's Index
Hazelnut	93	-	-	-
Almond	-	100	97	93
Chickpea	-	98	93	91

In this particular case, the additional performance parameters — unreliability region and the limits — were not able to be estimated. Although the percentages of adulteration were from 1 to 8% (in intervals of one), all sample were correctly recognised as model non-compliant (true negatives). Thus, from 0% to 1% of adulterant, there was a sudden drop in the probability of positives (from $P(x)$=93% to close to 0%), making no sense trying to fit a PCC curve. To be able to fit the curve, additional experimentation at percentages in between 0 and 1 should be required. In practice, and again in this particular case, it makes no sense to experiment at very low percentages of adulterant and in a narrow range, since the economic impact might not be significant.

6. Concluding remarks

In this tutorial, univariate and multivariate qualitative method validation is discussed in the context of international and official legislation. Qualitative performance parameters and the methodologies used to estimate them are also discussed. It should be borne in mind that there is no consensus about the terminology used, and we have attempted to reflect the most common terms.

Researchers must be aware of the performance parameters that can be established for each analytical problem. The methodology used to estimate them should be chosen for its suitability to the problem at hand.

This tutorial aims to encourage researchers who work in multivariate qualitative analysis to fulfil the validation by establishing performance parameters that involve quantitative information (unreliability region and concentration limits).

7. Acknowledgements

This work was supported by project 2009 SGR 549 of the Catalan Government. M.I López thanks the Rovira i Virgili University for providing her doctoral fellowship (2011BRDI-06-09).

8. References

[1] T. Jagric, S. Potrc, T. Jagric, Dig. Dis. Sci. 55 (2010) 3252–61.

[2] N.N. Chidumayo, K. Yoshii, N. Saasa, M. Sakai, H. Kariwa, Diagn. Microbiol. Infect. Dis. 78 (2014) 373–8.

[3] S. Tomatsu, T. Fujii, M. Fukushi, T. Oguma, T. Shimada, M. Maeda, et al., Mol. Genet. Metab. 110 (2013) 42–53.

[4] .M. Reisfield, B.A. Goldberger, R.L. Bertholf, Bioanalysis. 1 (2009) 937–52.

[5] S.L.R. Ellison, T. Fearn, TrAC - Trends Anal. Chem. 24 (2005) 468–476.

[6] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, et al., Accredit. Qual. Assur. 8 (2003) 68–77.

[7] E. Trullols, I. Ruisánchez, F.X. Rius, TrAC - Trends Anal. Chem. 23 (2004) 137–145.

[8] IUPAC Project 2005-024-2-600, Establishment of guidelines for the validation of qualitative and semi-quantitative (screening) methods by collaborative trial: a harmonized protocol.

[9] B. Magnusson and U. Örnemark, Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (2nd ed.), 2005.

[10] B. Hibbert, Quality Assurance in the Analytical Chemistry Laboratory, Oxford University Press, 2007.

[11] J.K. Taylor, H. V. Oppermann, Handbook for the Quality Assurance of Metrological Measurements, Lewis Publ, 1988.

[12] International. standard. ISO/IEC. 17025 - General requirements for the competence of testing and calibration laboratories, 2005.

[13] M. Thompson, S.L.R. Ellison, R. Wood, Pure Appl. Chem. 74 (2002) 835–855.

[14] E. Hund, D.L. Massart, J. Smeyers-Verbeke, Anal. Chim. Acta. 423 (2000) 145–165.

[15] Commission decision (EC) No. 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, Off. J. Eur. Communities. (2002) 8–36.

[16] M. Belter, A. Sajnóg, D. Barałkiewicz, Talanta. 129 (2014) 606–16.

[17] E. Trullols, I. Ruisánchez, F.X. Rius, M. Òdena, M.T. Feliu, J. AOAC Int. 87 (2004) 417–423.

[18] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, TrAC Trends Anal. Chem. 22 (2003) 647–654.

[19] L.A. Currie, Pure Appl. Chem. 67 (1995) 1699–1723.

[20] S. Brunelle, R. Labudde, M. Nelson, P. Wehling, AOAC INTERNATIONAL Methods Committee Guidelines for Validation of Microbiological Methods for Food and Environmental Surfaces, (2012).

[21] U.S. Food and Drug Administration, FDA Foods Program Guidelines for Chemical Methods - Guidelines for the Validation of Chemical Methods for the FDA Foods Program, 2012.

[22] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong,P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, 1998.

[23] S.L.R. Ellison, S. Gregory, W.A. Hardcastle, Perspective Quantifying uncertainty in qualitative analysis, 123 (1998) 1155–1161.

[24] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Anal. Chim. Acta. 674 (2010) 123–142.

[25] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Anal. Chim. Acta. 455 (2002) 267–275.

[26] R. Song, P.C. Schlecht, K. Ashley, J. Hazard. Mater. 83 (2001) 29–39.

[27] R. Macarthur, C. von Holst, Anal. Methods. 4 (2012) 2744.

[28] R.G. Brereton, Chemometrics for Pattern Recognition, Wiley, 2009.

[29] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Wiley, 2003.

[30] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong,P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, 1998.

[31] S.A. Drivelos, K. Higgins, J.H. Kalivas, S.A. Haroutounian, C.A. Georgiou, Food Chem. 165 (2014) 316–22.

[32] P. Oliveri, G. Downey, TrAC - Trends Anal. Chem. 35 (2012) 74–86.

[33] E. Borràs, J.M. Amigo, F. van den Berg, R. Boqué, O. Busto, Food Chem. 153 (2014) 15–9.

[34] C. de S. Gondim, O.A.M. Coelho, R.L. Alvarenga, R.G. Junqueira, S.V.C. de Souza, Anal. Chim. Acta. 830 (2014) 11–22.

[35] M. Zougagh, H. Téllez, A. Sánchez, M. Chicharro, A. Ríos, Anal. Bioanal. Chem. 391 (2008) 709–14.

[36] S. Vogliardi, D. Favretto, G. Frison, S. Maietti, G. Viel, R. Seraglia, et al., Anal. Bioanal. Chem. 396 (2010) 2435–40.

[37] M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, Anal. Chim. Acta. 827 (2014) 28–33.

[38] S.J. Mazivila, F.B. de Santana, H. Mitsutake, L.C. Gontijo, D.Q. Santos, W.B. Neto, Fuel. 142 (2015) 222–226.

[39] P. Oliveri, M. Casale, M.C. Casolino, M.A. Baldo, F. Nizzi Grifi, M. Forina, Anal. Bioanal. Chem. 399 (2011) 2105–2113.

[40] D.-H. Deng, L. Xu, Z.-H. Ye, H.-F. Cui, C.-B. Cai, X.-P. Yu, J. Am. Oil Chem. Soc. 89 (2012) 1003–1009.

[41] C. V Di Anibal, M. Odena, I. Ruisánchez, M.P. Callao, Talanta. 79 (2009) 887–92.

[42] C. V. Di Anibal, I. Ruisánchez, M.P. Callao, Food Chem. 124 (2011) 1139–1145.

[43] P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, et al., Anal. Chim. Acta. 851 (2014) 30–36.

[44] M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, Food Chem. 147 (2014) 177–181

4. 2. Results

VALIDATION OF MULTIVARIATE SCREENING METHODOLOGY. CASE STUDY: DETECTION OF FOOD FRAUD

M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao

Analytica Chimica Acta, 827 (**2014**) 28–33.

Analytica Chimica Acta, 827 (2014) 28–33

Validation of multivariate screening methodology. Case study: Detection of food fraud

M. Isabel López, Núria Colomer, Itziar Ruisánchez, M. Pilar Callao

Chemometrics, Qualimetric and Nanosensors Grup, Department of Analytical and Organic Chemistry, Rovia i Virgili University, Marcel·lí Domingo s/n, 43007 Tarragona, Spain

ABSTRACT

Multivariate screening methods are increasingly being implemented but there is no worldwide harmonized criterion for their validation. This study contributes to establish protocols for validating these methodologies. We propose the following strategy: (1) Establish the multivariate classification model and use receiver operating characteristic (ROC) curves to optimize the significance level (a) for setting the model's boundaries. (2) Evaluate the performance parameter from the contingency table results and performance characteristic curves (PCC curves). The adulteration of hazelnut paste with almond paste and chickpea flour has been used as a case study. Samples were analyzed by infrared (IR) spectroscopy and the multivariate classification technique used was soft independent modeling of class analogies (SIMCA). The ROC study showed that the optimal a value for setting the SIMCA boundaries was 0.03 in both cases. The sensitivity value was 93%, specificity 100% for almond and 98% for chickpea, and efficiency 97% for almond and 93% for chickpea.

***Keywords**: Multivariate screening validation, ROC curves, Performance characteristic curves, Food fraud, Performance parameters*

1. Introduction

Screening methods have been used with increasing success in routine analysis, thanks to their ability to identify the properties of samples at considerably reduced costs and times. They are characterized by their binary output – presence/absence, yes/no, etc. – according to a pre-set

threshold. Screening methods were first used in univariate analysis, which usually require specific measurement (i.e., test kits) [1]. More recently, screening methods have been developed for multiple measurements (i.e., analysis of different properties) or nonspecific signals (i.e., spectroscopic data). In these cases, proper multivariate data treatment is required if the output is to be binary.

In the field of food, multivariate screening methodologies have increased in importance ever since it became important to detect anomalous samples as well as ensure quality and safety [2,3].

According to the literature, analytical and classification techniques have been successfully combined in food fraud in both adulteration [4–8] and authentication [9–12]. We recently reported a multivariate screening methodology based on two classification approaches for a food adulteration problem [13].

As any analytical method, multivariate screening has to be validated to be implemented as routine methods in control laboratories. This involves establishing performance parameters. However, validation protocols for qualitative methods are not as developed as quantitative methods. For qualitative methods, the main guide is established in the Commission Decision CD/657/EC 2002 [14]. Nevertheless, this CD/657/EC 2002 has been interpreted ambiguously, which has led to a confusion in the terminology [15].

Multivariate validation is not as well established as univariate validation. There is considerable consensus about the definition of such performance parameters as sensitivity and specificity [16] but no agreement has been reached about other related indexes [2,17,18]. In addition, while a positive output in univariate analysis means the ‘presence of the analyte under study’, a positive output in multivariate analysis means that the ‘sample belongs to the pre-established model’.

Therefore, some performance parameters might have to be re-defined according to the compliance definition.

The goal of the present study is to establish a strategy for validating a multivariate screening methodology based on a parametric classification technique. In the proposed strategy, the steps were the following: (1) Establish the classification model. We propose using receiver operating characteristic (ROC) curves to optimize the significance level (α) for setting the model's boundaries. (2) Evaluate the performance parameters, some of which can be obtained directly from the output of the screening model. In this study, we propose using the performance characteristic curves (PCC curves) to obtain additional quality parameters such as the unreliability region and limits related to concentration (decision limit and detection capability). A hazelnut adulteration problem is considered as a case study. The price of hazelnuts depends on the market and can be reduced by adding such other ingredients as almond, because of its similarity, but other unexpected products might be added (for example, chickpea).

ROC curves have been extensively described by Fawcett in 2006 **[19]**. They have mainly been used in such fields as biomedicine, clinical analysis and biometrics **[20,21]** to set the cut-off value of a test or to compare the performance of different tests. However, their application in the field of food is less extensive, and they are usually used to select variables in multivariate classification techniques **[22–24]**.

PCC curves have mainly been used to obtain performance parameters other than the ones obtained by contingency tables in the quality characterization of univariate screening methodologies **[18,25–27]**. No references have been found in which PCC curves are used in multivariate screening validation.

This paper takes a further step in several aspects. Firstly, we use ROC curves to optimize the significance level (α) used to establish the boundaries of the model. This α value is usually set to 0.05 by default. Secondly, we propose to adapt PCC curves to a multivariate methodology, for the first time, when dealing with modelling classification techniques.

2. Experimental

2.1 Samples

The unadulterated set is composed of 28 hazelnut pastes. Experience shows that the most common percentage of adulteration is around 7%. So, representative samples were selected and spiked with almond paste or chickpea flour at different levels (1– 8%) so that the PCC curves could be established. In total, there were around 13 adulterated samples at each adulteration level and for each adulterant studied.

Additional details about hazelnuts, adulterants and sample preparation can be found in our previous study **[13]**.

2.2 Instrumentation and software

The spectra data was acquired by an infrared (IR) spectrophotometer (FTIR 680 Plus JASCO) equipped with a diamond crystal in the spectrophotometer ATR cell, which was continuously purged with N_2. Spectra were the average of 32 scans, recorded in the spectral range of 4000–600 cm^{-1} every 2 cm^{-1}. The CO_2 and H_2O contributions were removed with the control software Spectra Manager before the spectra were exported to Matlab **[28]** and treated with PLS Toolbox **[29]**.

Analytica Chimica Acta, 827 (2014) 28–33

3. Strategy

We propose a multivariate screening strategy based on one class approach, so only the unadulterated samples are modeled. Once the model is established, the performance parameters are evaluated. The algorithm used in the present work is soft independent modeling of class analogies (SIMCA), which has been widely applied for classification problems and it is of great interest whenever dealing with spectroscopic data since it is not influenced by working with collinear variables (highly correlated).

SIMCA was introduced by Wold in 1976 **[30]**. It is a classification technique based on principal component analysis (PCA) that characterizes each sample in relation to the build model by calculating two scalar statistics, *Q* and the Hotelling T^2. The *Q*-statistic is related to the amount of original information of each sample not included in the model whereas Hotelling T^2 measures the information of each sample considered by the model. Like any parametric classification technique, some limits, Q_{lim} and T^2_{lim}, must be set to delimit SIMCA boundaries, which depend on the significance level (α) **[31]**.

3.1 Screening output

In a food adulteration problem such as the one being studied, the proposed multivariate screening methodology is used to establish a one-class model since the main interest is to describe the compliant samples. Therefore, the one-class model was built with the unadulterated samples. From prediction step a binary output is obtained for each sample: positive (sample belongs to the model) or negative (sample does not belong to the model). Whether the output is positive or negative depends on the boundaries of the model, which depends on the significance level (α). For a given sample i, the decision criterion, (D), is defined according to:

$$D(\alpha) = \begin{cases} if & T^2_{lim,\alpha} \geq T_i \quad AND \quad Q_{lim,\alpha} \geq Q_i \quad then \quad + \ output \\ if & T^2_{lim,\alpha} < T_i \quad OR \quad Q_{lim,\alpha} < Q_i \quad then \quad - \ output \end{cases}$$

Comparing actual data and output results, four responses are obtained which are organized in a contingency table (see Table 1). In the case under study, a true positive (TP) result is when a positive output is obtained from an actual unadulterated sample and a false positive (FP) result is when a positive output is obtained from an actual adulterated sample. In like manner, similar reasoning is followed for TN and FN.

Table 2 shows some of the performance parameters that can be calculated from the contingency table results. As can be seen, sensitivity is defined as the ability of the model to recognize its own samples and specificity as the ability of the model to distinguish external samples. Both have a maximum value of 1 and a minimum of 0. False negative (FN) and false positive (FP) rates are related to sensitivity and specificity, respectively. Other parameters such as efficiency and Youden's index assess the overall suitability of the model. Once more, there is some terminological confusion as, according to some authors, efficiency is referred to as accuracy [2,18].

Table 1. *2x2 Contingence Table.*

Predictions	**Actual**	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

TP, true positive; TN, true positive;
FP, false positive; FN, false negative

Table 2. *Description of the performance parameters.*

Sensitivity	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
False negative (FN) rate (1-sensitiviy)	$\frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}}$
Specificity	$\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
False positive (FP) rate (1-specificity)	$\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}$
Efficiency	$\frac{\mathrm{TN}+\mathrm{TP}}{\mathrm{TN}+\mathrm{FP}+\mathrm{TP}+\mathrm{FN}}$
Youden's index	$(Sensitivity + Specificity - 1)$

3.2 Screening strategy

The strategy for establishing the screening model is summarized on the left-hand side of Fig. 1. It consists of several steps: (1) Establish the SIMCA model and define its boundaries considering different significance levels (α); (2) obtain a contingency table at each α, according to the screening outputs; (3) calculate the specificity and the sensitivity at each α; (4) use ROC curves to select the optimal confidence limit (α) of the model that shows the best sensitivity and specificity values.

ROC analysis is a comprehensive visual tool for summarising the relation between sensitivity and specificity. The ROC curves depict sensitivity against '1-specificity' for each of the thresholds studied **[19]**. The theoretical curves are shown in Fig. 2a. They begin at (0, 0), which corresponds to (sensitivity = 0, specificity = 1), and end at (1, 1) indicating (sensitivity = 1, specificity = 0). The optimal threshold is the one that shows highest values for both sensitivity and specificity. The diagonal line (Fig. 2a(i)) is called the "chance diagonal", which represents

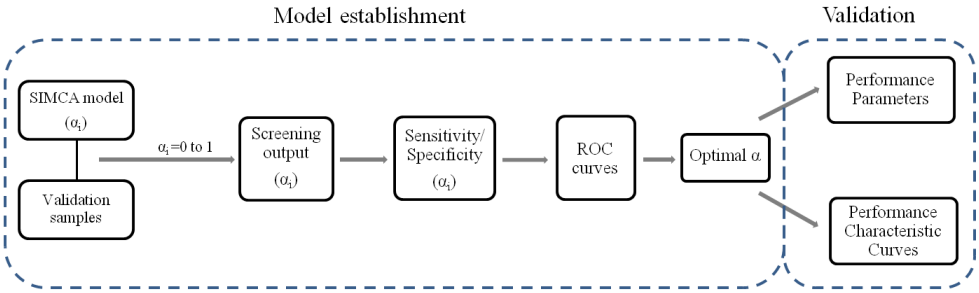

Fig. 1. *Scheme of the multivariate screening strategy*

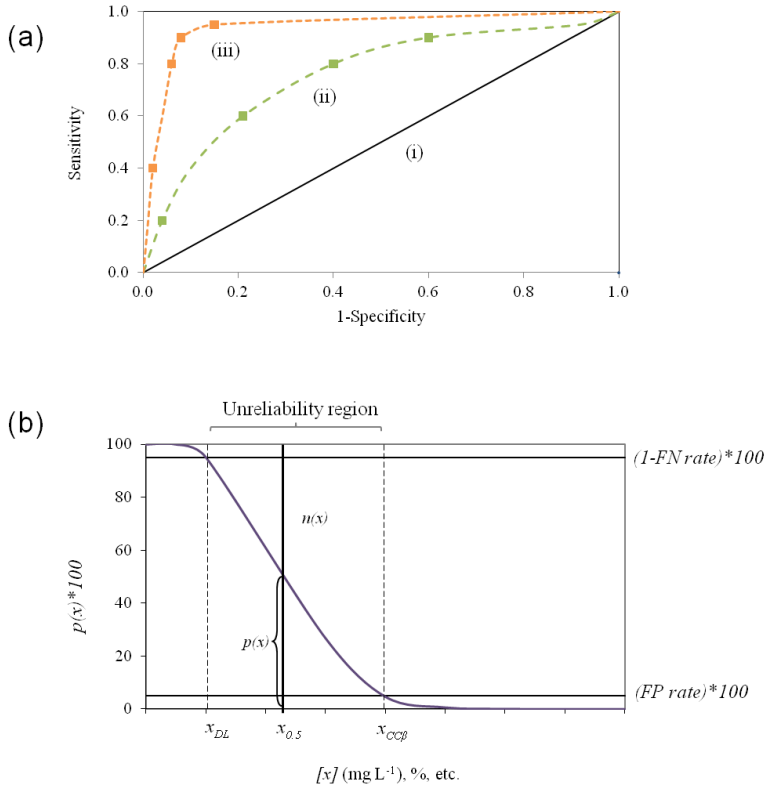

Fig. 2. *(a) Theoretical ROC curves showing increasing goodness (from i to iii), being (i) the worse one called the "diagonal chance line". (b) Theoretical PCC curve with 'Ƨ'shape.*

the ROC curve with the same probability of classifying a sample as true positive and false positive. The area under the curve (AUC) is frequently used to compare the overall performance of different curves. It is close to 1 when it shows maximum classification ability (Fig. 2a(iii)), whereas the diagonal line (Fig. 2a(i)) indicates the minimum AUC = 0.5.

3.3 Validation

Once the model boundaries are established by setting the optimal α value, the final quality parameters are calculated according to Table 2 and PCC curves (right-hand side of Fig. 1).

Quality parameters such as unreliability region, decision limit (also referenced as threshold, cut-off, LOD and CCα **[15]**) and detection capability (CCβ), all of which are concentration parameters, can be estimated from the PCC curves. Their definition depends on the screening criteria (see Section 3.1). In the present study, the unreliability region is a concentration interval in which there is some probability of having false outputs. The decision limit is a concentration value below which the screening system will give a positive output with high probability. The detection capability is a concentration value from which the screening method gives negative outputs with highest reliability or smallest probability of error **[14]**. Finally, another interesting concentration value is the one from which the probability of getting a negative output (being true negative, TN) is higher than getting a positive output (being false positive, FP) **[32]**.

To obtain the PCC curve, samples at different concentrations of adulterant have to be predicted by the SIMCA model. Then the rates of positive, *p(x)*, and negative, *n(x)*, outputs are calculated (both rates are related: *n(x) = 1 - p(x))*. The *p(x)*100* values are plotted versus the concentration of adulterant, and the experimental points are fitted to a

sigmoid function, according to Eq. 1, minimizing the root mean square of the residuals (RMSE) [33].

$$p(x) = \frac{1}{1 + e^{b(x-a)}} \qquad Eq.1$$

This PCC curve has an 'Ƨ'-shape from 1 to 0, where *p(x)* is the rate of having a positive output, *x* is the adulterant concentration, and *a* (amplitude of the curve) and *b* (slope) are the regression coefficients which are fitted to minimize the RMSE.

Fig. 2b shows a theoretical PCC curve that corresponds to a screening system such as the one defined in Section 3.1. Once the curve has been plotted, to obtain the concentration limits two horizontal lines have to be drawn. The upper line is usually set at $p(x) = 0.95$ which corresponds to a FN rate of 0.05 ($p(x) = 1 - FN\ (0.05) = 0.95$). The intersection between the upper horizontal line and the sigmoid curve sets a limit on the concentration ($x_{0.95}$). This is the decision limit ($x_{DL} = x_{0.95}$) since for concentration values lower than the x_{DL} ($x \leq x_{DL}$) the rate of having a TP is ≥ 0.95, and the rate of having a FN decreases (lower than 0.05).

Similar reasoning was used to set the bottom horizontal line at $p(x) = 0.05$, which corresponds to a FP rate of 0.05 ($p(x) = 0.05$). Once more, the intersection between the bottom horizontal line and the sigmoid curve sets another concentration limit ($x_{0.05}$) which is the detection capacity ($x_{CC\beta} = x_{0.05}$) since for concentration values higher than the $x_{CC\beta}$ ($x \geq x_{CC\beta}$), the rate of positive outputs is lower than 0.05 (being FP) and the rate of negative outputs is higher than 0.95 (being TN).

The concentration region in between these two limits ($x_{DL} < x < x_{CC\beta}$) is the unreliability region and is the central part of the sigmoid curve where most of the screening errors occur. In Fig. 2b, a third concentration value ($x_{0.5}$) has been marked because it is highly informative. This value corresponds to the concentration at which there is the same rate (0.50) of

having a positive output (being FP) than a negative one (TN). This value is in accordance with some of the definitions of limit of detection for qualitative methods [32].

Indeed, all three concentration limits can be calculated from Eq. 2, just by setting the value of *p(x)*.

$$x_{\text{limit}} = \frac{\ln\left(\frac{1}{p(x)} - 1\right)}{b} + a \qquad Eq.2$$

where x_{limit} is the concentration limit (either decision, detection capability, etc.), *p(x)* is the rate of obtaining a positive output (usually set at 0.95, 0.05 or 0.50) and *a* and *b* are the estimated sigmoid regression coefficients.

Although the rates of committing errors are usually set by default to 0.05, they can be set according to other criteria. For instance, take into account the best sensitivity while having specificity above a predefined threshold; consider FN and FP misclassifications, among others. We propose to set them by taking into account the errors obtained from the model at optimal α, being, therefore, the upper horizontal line plotted at *p(x)* = 100 - FN_α and the bottom horizontal line at *p(x)* = FP_α.

4. Results and Discussion

4.1. Pre-processing

A vertical offset correction was applied to original IR spectra to eliminate any vertical shift. It involves subtracting the absorbance at 3822 cm^{-1}, where no peak is observed, from the absorbance at each wavenumber. Then, a Savitzsky–Golay smoothing [34] was applied to all data using a 29 datapoint window and a second-order polynomial to suppress the instrumental noise.

Fig. 3 shows the corrected IR spectral data of whole samples. For clarity, the spectra of adulterated data have been shifted in the y-axis direction. As can be seen, some regions of the spectra show no peak. Keeping redundant spectral information can negatively affect the performance of the modeling step, so these variables were removed. The IR spectra, then, were reduced to the ranges 600–1864 cm^{-1} and 2630–3137 cm^{-1} through visual variable selection, which involves 3673 final variables.

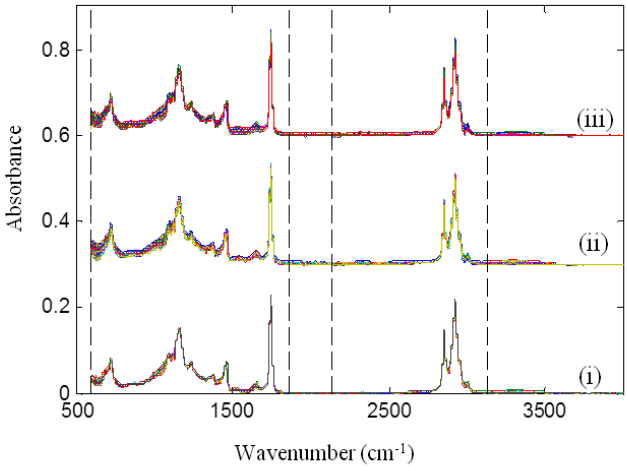

Fig. 3. *Corrected IR spectra data of unadulterated (i) and adulterated samples with almond (ii) and chickpea (iii). Vertical discontinuous lines mark out the ranges of the spectra that were removed.*

4.2. Screening strategy

The model was built with unadulterated data and validated by leave-one-out cross-validation (LOO-CV). The number of principal components needed to describe the model was chosen by plotting the percentage of correct classification as a function of the number of PCs. In our case, the model was developed on the basis of the first two PCs.

Analytica Chimica Acta, 827 (2014) 28–33

To obtain the contingency table data, the unadulterated samples and the samples adulterated with almond paste and chickpea flour were predicted. Since the expected adulteration level is around 7%, the first part of this study was performed using samples with adulteration levels of 6, 7 and 8%, giving a total number of 38 and 41 samples adulterated with almond and chickpea, respectively.

ROC curves were established for each adulterant (plot not shown). Experimental data are obtained by setting α value from 0.90 to 0.10 in tenth and from 0.09 to 0.01 in hundredth. The AUCs of the ROC curves were close to 1 in both cases (0.9736 and 0.9728 for almond and chickpea, respectively), which is almost the ideal behavior.

Table 3 summarized the performance parameters obtained for α value between 0.09 and 0.01, as this is the region of interest. Generally speaking, maximum values for both sensitivity and specificity are sought but, in practice, a compromise has to be reached as it is not usually possible to obtain an α value with a maximum value for both parameters. When dealing with a hazardous contaminant, specificity is obviously the key point since the main goal is not to consider an adulterated sample to be unadulterated. In our case, as we are dealing with an adulterant that only has economic implications, the global performance parameters could be useful for the final decision.

For almond, the performance parameters were best, with very slight differences, with α = 0.03 or α = 0.02 (indicated in bold). For chickpea, the performance parameters were best with α = 0.09–0.06 or α = 0.03 (indicated in bold). So, taking into account the four parameters considered, α = 0.03 is our choice in both cases. Note that the value of the performance parameters is slightly different and the end user could use other criteria to set the α value.

Table 3. *Performance values obtained at different α values. Bold values correspond to the discussed values being the underlined the ones chosen.*

Threshold (α value)	Almond			
	Sensitivity	Specificity	Efficiency	Youden's Index
0.10	0.82	1.00	0.92	0.82
0.09-0.06	0.86	1.00	0.94	0.86
0.05-0.04	0.86	1.00	0.94	0.86
0.03	**0.93**	**1.00**	**0.97**	**0.93**
0.02	**0.96**	**0.95**	**0.95**	**0.91**
0.01	0.96	0.92	0.94	0.88
Threshold (α value)	**Chickpea**			
	Sensitivity	Specificity	Efficiency	Youden's Index
0.10	0.82	1.00	0.93	0.82
0.09-0.06	**0.86**	**1.00**	**0.94**	**0.86**
0.05-0.04	0.86	0.98	0.93	0.84
0.03	**0.93**	**0.98**	**0.93**	**0.91**
0.02	0.96	0.90	0.88	0.86
0.01	0.96	0.80	0.83	0.76

4.3. Validation

Once the boundaries of the model had been set to α = 0.03, the final performance parameters were the ones underlined in Table 3. Obviously,

sensitivity is the same for both adulterants as it depends on the type of sample used to build the model (unadulterated). The other performance parameters, which depend on the external samples (adulterated), have similar values, indicating that the adulterant type does not matter.

Considering that quality performance ability of the model is the goal, we try to fit the PCC curve with both adulterant outputs. We have proposed setting the p(x) values, at which the horizontal lines on the PCC curve should be drawn (Fig. 2b), from the estimated parameters (Table 3). In our case, the FN rate (1 - sensitivity) was 0.07, so the upper horizontal line should be drawn at $p(x)$ = 1 - FN (0.07) = 0.93. The FP rate obtained, however, was slightly different for each adulterant and we propose to take the less restrictive value, so the horizontal line should be drawn $p(x)$ = 0.02.

In the case under study, the percentage of positive outputs was 93% for the unadulterated samples (0% of adulterant) and 0% for samples at the lower level of adulterant studied (1% of adulterant). The same tendency is observed for higher percentages of adulterant (from 2 to 8%) present in the hazelnuts samples ($p(x)$ values close to 0). With that experimental data, it makes no sense trying to fit a PCC curve, since $p(x)$ should gradually decrease from 100% to 0% as the percentage of adulterant increases. Several points in between are required, therefore, additional points should be analyzed for a proper definition of the PCC curve. From a practical point of view, again in that particular case, it makes no sense to experiment at very low percentages of adulterant and in a narrow interval, since the economic impact might not be significant. This fact agrees with the characteristics of the developed multivariate screening model which has very high values for both, sensibility and specificity (Table 3). As it has been shown, as far as the percentage of adulterant decreases, the specificity still has very high values.

5. Conclusions

In the validation of a multivariate screening methodology based on one-class modeling, it has to be taken into account that the positive (or negative) output might have to be re-defined according to the compliance definition. Usually, in a univariate approach, positive output always means “presence of analyte” while in the multivariate approach it depends on the class modeled. For instance, if the class modeled corresponds to an unadulterated situation, positive output means “no presence of adulterant”.

ROC curves have proved to be a useful tool for setting the significance level (α) required for establishing the model boundaries. They give the end user the chance to optimize the model boundaries in accordance with the best performance parameters thus giving more importance to sensitivity or specificity depending on the problem under study.

PCC curves (well known in univariate approaches) have been used to calculate some performance parameters in a multivariate methodology. This allows calculating the performance parameters of concentration (decision limit, detection capability and unreliability region). There is no other way of doing so. Two points should be mentioned: there is no agreement on the definition of those concentration limits and the shape of the PCC curve might change according to the positive output definition.

The proposed screening strategy to determine adulteration in hazelnut paste allows choosing the α value to set the model’s boundaries. In our case, the traditional α value set to 0.05 was not the best option. Instead, higher sensitivity and specificity were obtained by reducing the significance level to α = 0.03. This underlines the importance of this step. From a practical point of view it has not been possible to fit a PCC curve

because the specificity is very high for any studied percentage of adulterant.

6. Acknowledgments

This work was supported by project 2009 SGR 549 of the Catalan Government. M.I López thanks the Rovira i Virgili University for providing her doctoral fellowship (2011BRDI-06-09). We would like to thank LaMorellaNuts S:A.U. for providing the samples.

7. References

[1] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, K. Klemm, B. Lendl, B. Milman, B. Neidhart, R.W. Stephany, A. Townshend, A. Zschunke, M. Valcárcel, Accredit, *Quality Assurance* 8 (2003) 68–77.

[2] P. Oliveri, G. Downey, *Trends in Analytical Chemistry* 35 (2012) 74–86.

[3] R. Karoui, J. De Baerdemaeker, *Food Chemistry* 102 (2007) 621–640.

[4] C.V. Di Anibal, M. Odena, I. Ruisánchez, M.P. Callao, *Talanta* 79 (2009) 887–892.

[5] G. Gurdeniz, B. Ozen, *Food Chemistry* 116 (2009) 519–525.

[6] M. Maia, A.I.R.N.A. Barros, F.M. Nunes, *Talanta* 107 (2013) 74–80.

[7] A. Oussama, F. Elabadi, S. Platikanov, F. Kzaiber, R. Tauler, *Journal of the American Oil Chemists Society* 89 (2012) 1807–1812.

[8] C.V. Di Anibal, I. Ruisánchez, M.P. Callao, *Food Chemistry* 124 (2011) 1139–1145.

[9] Q. Zhang, C. Liu, Z. Sun, X. Hu, Q. Shen, J. Wu, *Food Chemistry* 132 (2012) 1607–1613.

[10] D. Cozzolino, *Applied Spectroscopy Reviews* 47 (2012) 518–530.

[11] M. Bevilacqua, R. Bucci, A.D. Magrì, A.L. Magrì, F. Marini, *Analytica Chimica Acta* 717 (2012) 39–51.

[12] E. Ferrari, G. Foca, M. Vignali, L. Tassi, A. Ulrici, *Analytica Chimica Acta* 701 (2011) 139–151.

[13] M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, *Food Chemistry* 147 (2014) 177–181.

[14] Commission Regulation (EC) No. 2002/657/EC, Official Journal of the European Communities, (2002) 8–36.

[15] R. Galarini, R. Buratti, L. Fioroni, L. Contiero, F. Lega, *Analytica Chimica Acta* 700 (2011) 2–10.

[16] P. Feldsine, C. Abeyta, W.H. Andrews, *Journal of AOAC International* 5 (2002) 1187–1200.

[17] N. Rodríguez, M.C. Ortiz, L.A. Sarabia, A. Herrero, *Analytica Chimica Acta* 657 (2010) 136–146.

[18] S.L.R. Ellison, T. Fearn, *Trends in Analytical Chemistry* 24 (2005) 468–476.

[19] T. Fawcett, *Pattern Recognition Letters* 27 (2006) 861–874.

[20] L. Ohno-Machado, T.A. Lasko, J.G. Bhagwat, K.H. Zou, *Journal of Biomedical Informatics* 38 (2005) 404–415.

[21] C.H. Lee, J. Kim, Y. Park, Y.C. Park, Y. Kim, K.J. Yoon, Y. Uh, K.-A. Lee, *Clinica Chimica Acta* 427 (2014) 11–14.

[22] R. Moscetti, R.P. Haff, B. Aernouts, W. Saeys, D. Monarca, M. Cecchini, R. Massantini, *Journal of Food Engineering* 118 (2013) 1–7.

[23] D. Lorente, J. Blasco, J. a. Serrano, E. Soria-Olivas, N. Aleixos, J. Gómez-Sanchis, *Food and Bioprocess Technology* 6 (2013) 3613–3619.

[24] X. Luo, T. Takahashi, K. Kyo, S. Zhang, *Journal of Food Engineering* 109 (2012) 457–466.

[25] R. Song, P.C. Schlecht, K. Ashley, *Journal of Hazardous Materials* 83 (2001) 29–39.

[26] E. Trullols, I. Ruisánchez, F.X. Rius, *Trends in Analytical Chemistry* 23 (2004) 137–145.

[27] E. Trullols, I. Ruisánchez, F.X. Rius, M. Òdena, M.T. Feliu, *Journal of AOAC International* 87 (2004) 417–423.

[28] T. Mathworks, MATLAB Version 7. 0, Natick, MA, 2014.

[29] Eigenvector Research Incorporate (PLS_Toolbox Version 7.0.2).

[30] S. Wold, *Pattern Recognition* 8 (1976) 127–139.

[31] A. Rius, M.P. Callao, F.X. Rius, *Analyst* 122 (1997) 737–741.

[32] K. Ashley, T.J. Fischbach, R. Song, *American Industrial Hygiene Association Journal* 57 (1996) 161–165.

[33] L.M. McDowall, R.A.L. Dampney, *American Journal of Physiology – Heart and Circulatory Physiology* 291 (2006) H2003–H2007.

[34] A. Savitzky, M.J.E. Golay, *Analytical Chemistry* 36 (1964) 1627–1639

4. 3. References

[1] C. B. R. dos Santos, J. B. Vieira, A. da Silva Formigosa, E. V. M. da Costa, M. T. Pinheiro, J. O. da Silva, W. J. de Cruz Macêdo, J. C. T. Carvalho, *J. Comput. Theor. Nanosci.* **2014**, *11*, 553.

[2] S. Naz, A. Garcia, M. Rusak, C. Barbas, *Anal. Bioanal. Chem.* **2013**, *405*, 4849.

CHAPTER 5

GENERAL CONCLUSIONS

This chapter summarises the main conclusions of the work developed in this doctoral thesis. Specific conclusions drawn from each individual study are presented at the end of the corresponding research paper.

This thesis deals with multivariate qualitative methodologies to detect food fraud. The general goals and therefore its conclusions, cover the different steps in a multivariate qualitative methodology: (1) *data source*: studies working with different spectroscopic techniques have been presented: NIR, IR and SERS spectroscopies; (2) *classification models*: studies working with different classification techniques (SIMCA, UNEQ, OCPLS and PLS-DM) as well as different strategies (targeted and untargeted) have been assessed; (3) *method validation*: a validation proposal for multivariate qualitative methodologies has been presented indicating the way to estimate the performance parameters.

The specific objectives described in Chapter 1 and the conclusions associated to them are the following:

1. To evaluate Raman signal in the Surface-enhanced Raman spectroscopy (SERS) modality for its use in multivariate approaches.

The characterisation of the SERS support shows the high potential of SERS technique in the qualitative and quantitative analysis, meaning that it can be regarded as a good alternative or as a complement to other spectroscopic techniques. In our work, the complete figures of merit have been established, which it is rarely performed when working with SERS technique. This constitutes just a mid-step for the final goal, which aims for multivariate analysis.

Regarding its application as multivariate qualitative analysis to check adulteration of Sudan in spices (i.e., mild and hot paprika), some

reproducibility problems were found when dealing with real samples. Our group is working to solve them, considering SERS hyperspectral imaging modality as data source for the multivariate qualitative purposes.

2. To develop multivariate qualitative methods based on untargeted modelling.

The classification models developed in this thesis have given suitable results, making possible to detect food fraud, either authentication or adulteration. The new developed classification technique (PLS-DM) has shown to be successful when dealing with complex class distribution (class heterogeneity).

Untargeted (or one-class) approach has shown to be a useful approach either if the analyst is focused on characterizing only one category or if only samples from one of the categories can be representatively collected. Samples that are non-compliant with the modelled requirements (adulterated) are suitably detected regardless the adulteration source.

By contrast, targeted (or multi class) modelling provides information that otherwise would not be available. If an adulteration source is known as probable, it can be controlled by modelling this class whether a non-compliant sample contains this specific adulterant.

3. To establish validation protocols for multivariate qualitative analysis in accordance with the indications of the European directives and to calculate the performance parameters associated with them.

An attempt to clarify the terminology regarding qualitative method validation in accordance with international and official definitions has

been presented, considering both univariate and multivariate approaches.

The key steps in a multivariate method validation have been described, introducing the performance parameters that involve quantitative information (unreliability region and concentration limits). Our proposal has addressed the estimation of those parameters by means of the performance characteristic curves, which represent a novelty in multivariate analysis. In this regard, it brings multivariate qualitative analysis closer to the performance parameters that international directives propose to be set.

As a final conclusion, this thesis has presented several methodologies that can be extrapolated to other analytical applications (environmental, biological, etc.) and can be implemented by using other analytical techniques, instrumental modalities and classification techniques.

APPENDIX

APPENDIX A: SUMMARY OF MAIN ABBREVIATIONS

Chemometrics

CV	Cross-Validation
LDA	Linear Discriminant Analysis
LOO-CV	Leave-One-Out Cross-Validation
LV	Latent Variable
OCPLS	One-Class Partial Least Squares
PC	Principal Component
PCA	Principal Component Analysis
PFM	Potential Functions Method
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PLS-DM	Partial Least Squares Density Modelling
SIMCA	Soft Independent Modelling Of Class Analogies
UNEQ	Unequal Dispersed Classes

Instrumentation

FTIR-ATR	Fourier Transform Infrared Spectroscopy Attenuated Total Reflection
FT-NIR	Fourier Transform Near-Infrared Spectroscopy
IR	Infrared Spectroscopy
NIR	Near-Infrared Spectroscopy
SERS	Surface-Enhanced Raman Spectroscopy

Quality Performance

CCβ	Detection Capability
FN	False Negative
FP	False Positive
LOD	Limit of Detection
N(x)	Negative Response Rate
P(x)	Positive Response Rate
PCC	Performance Characteristic Curves

ROC	Receiver Operating Characteristic
TN	True Negative
TP	True Positive
DL	Decision Limit

APPENDIX B: LIST OF PAPERS

✓ Papers presented by the author in this thesis

M.I. López, I. Ruisánchez, M.P. Callao, **FIGURES OF MERIT OF A SERS METHOD FOR SUDAN I DETERMINATION AT TRACES LEVELS**, *Spectrochim. Acta. A.*, 111 (2013) 237–241.

M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, **MULTIVARIATE SCREENING IN FOOD ADULTERATION: UNTARGETED VERSUS TARGETED MODELLING**, *Food Chemistry*, 147 (2014) 177-181.

P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, S. Lanteri **PARTIAL LEAST SQUARES DENSITY MODELING (PLS-DM) – A NEW CLASS-MODELING STRATEGY APPLIED FOR THE AUTHENTICATION OF OLIVES IN BRINE BY NEAR-INFRARED SPECTROSCOPY**, *Anal. Chim. Acta,* 851 (2014) 30-36.

M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, **VALIDATION OF MULTIVARIATE SCREENING METHODOLOGY. CASE STUDY: DETECTION OF FOOD FRAUD**, *Anal. Chim. Acta*, 824 (2014) 28–33.

M.I. López, M.P. Callao, I. Ruisánchez, **A TUTORIAL ON THE VALIDATION OF QUALITATIVE METHODS: FROM THE UNIVARIATE TO THE MULTIVARIATE APPROACH**, *Anal. Chim. Acta*, Submitted.

APPENDIX C: MEETING CONTRIBUTIONS

✓ Oral Communications

M.I. López, M.P. Callao, I. Ruisánchez
Untargeted versus targeted modelling in a food adulteration problem. **VIII COLLOQUIUM CHEMIOMETRICUM MEDITERRANEUM**, Bevagna (Italy, 2013)

M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao
Validation of screening multivariate methodology for testing adulteration. **V WORKSHOP QUIMIOMETRÍA**, Badajoz (Spain, 2013)

P. Oliveri, M.I. López, M.C. Casolino, S. Lanteri, L. Medini
Authentication of Taggiasca olives in brine by application of a novel PLS-based class-modelling method on NIR spectra. **XXV CONGRESSO NAZIONALE DELLA SOCIETÀ CHIMICA ITALIANA**, Calabria (Italy, 2014)

✓ Poster Communications

M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao
Food adulteration study by a multivariate screening approach. **XVIII REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA**, Jaén (Spain, 2013)

P. Oliveri, M.I. López, M.C. Casolino, L. Bagnasco, S. Lanteri, L. Medini
Near infrared characterisation of Taggiasca olives in brine by a PLS-based class-modelling method. **VI SIMPOSIO ITALIANO DI SPETTROSCOPIA NIR**, Modena (Italy, 2014)

APPENDIX D: RESEARCH STAY AND TRAINING COURSES

✓ Research stay

Genoa (January - April 2014)

Objective: *Development and application of new chemometric tools for class-modelling problems*. Supervised by Dr. Paolo Oliveri

University of Genoa, Department of Drug and Food Chemistry and Technology, Via Brigata Salerno, 13, I-16147 Genoa, Italy.

✓ Training courses

Herramientas quimiométricas para PAT, June 2012.

University of Barcelona, Chemistry Faculty, Diagonal, 645. 08028 Barcelona, Spain.

Scuola di Chemiometria, January 2014.

University of Genoa, Department of Drug and Food Chemistry and Technology, Via Brigata Salerno, 13, I-16147 Genoa, Italy.

UNIVERSITAT ROVIRA I VIRGILI
DEVELOPMENT AND VALIDATION OF MULTIVARIATE QUALITATIVE METHODS IN THE FOOD FIELD
Maria Isabel López Vilardell
Dipòsit Legal: T 893-2015

<< It was a wonderful tree that promised exquisite and generous fruits. Certainly, it gave the best of itself. >>

Everyone plays a part