Tutorial

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

HIGHLIGHTS

- Strategies for qualitative method validation are considered.
- Univariate and multivariate qualitative methods are described.
- Performance parameters for qualitative methods are discussed.
- Cases study of qualitative methods validation is illustrated.

GRAPHICAL ABSTRACT
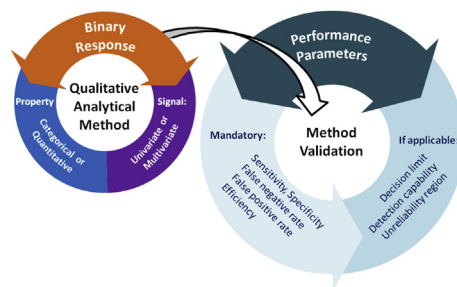
ARTICLE INFO

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.

© 2015 Elsevier B.V. All rights reserved.

Contents

---

* Corresponding author.
  E-mail address: mariapilar.callao@urv.cat (M.P. Callao).

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, focussing 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.

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.
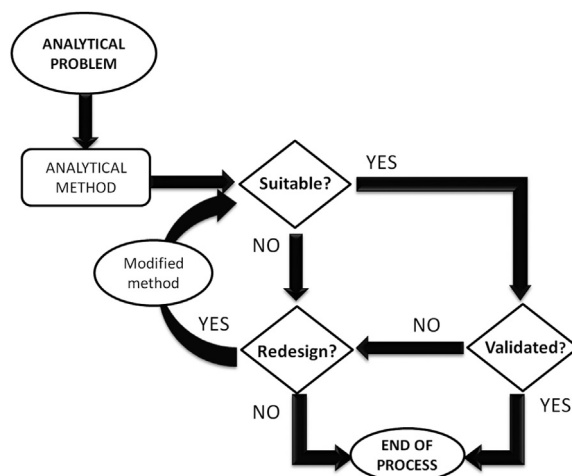

Fig. 1. Diagram of method development and validation (adapted from EUROCHEM The Fitness for Purpose of Analytical Methods [9]).

2.1.1. Validation level

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

2.1.1.1. 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.

2.1.1.2. 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.

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.

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

Table 1
Quality performance parameters.

Quantitative	Qualitative
✓ Accuracy: **trueness,** precision	✓ **Trueness**
✓ Uncertainty	✓ False positive (FP) and false negative (FN) rates
✓ <u>Sensitivity and specificity</u>	✓ <u>Sensitivity and specificity</u>
✓ 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**

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].

It should be pointed out that in the literature the term limit of detection is used to refer to both CCβ and CCα. 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, α and β errors are both evaluated from experimental data (compliant and noncompliant samples) and usually expressed as frequency of two false decisions (or errors), false non-compliant and false compliant. 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 non-compliant) or to restrain the errors of saying that a sample does not contain a substance when in fact it does (false compliant).

The unreliability region is defined by the two limits CCα 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 analysed variable. This variable should provide enough information to solve the analytical problem at hand (Fig. 2a) [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 (in the web version) 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 analysed variables (Fig. 2b) [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.

Regardless of the number of variables analysed, 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 assignation as compliant/non-compliant 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.

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 $\geq$ 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
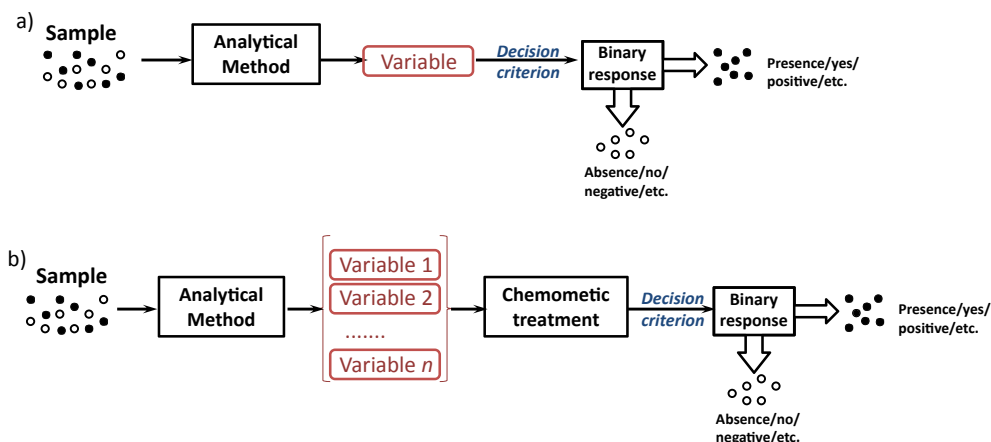
Fig. 2. Scheme of a qualitative analysis using (a) univariate and (b) multivariate approaches.

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 $\neq$ 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)

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

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

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.

Several methodologies can be used to estimate the mandatory performance parameters: Bayesian decision theory, 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 Bayesian decision theory [23] and statistical hypothesis tests [24,25] can be found elsewhere.

Contingency tables are based on frequency evaluation of the responses (positive and negative) 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 2 × 2 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

Table 4
2 × 2 contingence table.

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

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

Table 5
Description of the performance parameters.

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

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], the generalized linear model [27], probability of detection [28,29] and probability of identification [30], among other names, or just used without specific name [6].

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 5 mg/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 ≤ 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 analysed 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. 3a 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 ≤5.

Real behaviour, however, is different from the ideal one (see Fig. 3b). To obtain the PCC curve, the experimental $P(x)$ values are fitted to a sigmoid function (Eq. (1) corresponds to the expression of one possible function), minimizing the root mean square of the residuals.

$$p(x) = 1 - e^{-\left(\frac{x}{a}\right)^b} \tag{1}$$

where $p(x)$ is the rate of having a positive output, x is the concentration of an analyte and a (amplitude of the curve) and b (slope) are the regression coefficients which are fitted to minimize the root mean square of the residuals.

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:

- CC$_\alpha$ *(decision limit)*: it is obtained from the intersection between the lower horizontal line — which corresponds to the probability of committing an FP error (false non-compliant) of 5% ($P(x) = 5\%$) — and the sigmoid curve.
- CC$_\beta$ *(detection capability)*: it is obtained from the intersection between the upper horizontal line — which corresponds to the probability of committing an FN error (false compliant) of 5% ($P(x) = 95\% = 1 − FN$) — 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 $0 < x < CC_\alpha$ is considered the reliable negative region since at concentrations lower than CC$_\alpha$ ($x < CC_\alpha$) the probability of getting a true negative result is >95%, and the rate of getting a false positive result decreases (lower than 5%). Similarly, the region $x > CC_\beta$ is considered the reliable positive region since there is a 95% or higher probability of getting a true positive result at concentrations higher than CC$_\beta$.

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 ≤ 5 mg/L) and category B (sample is non-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
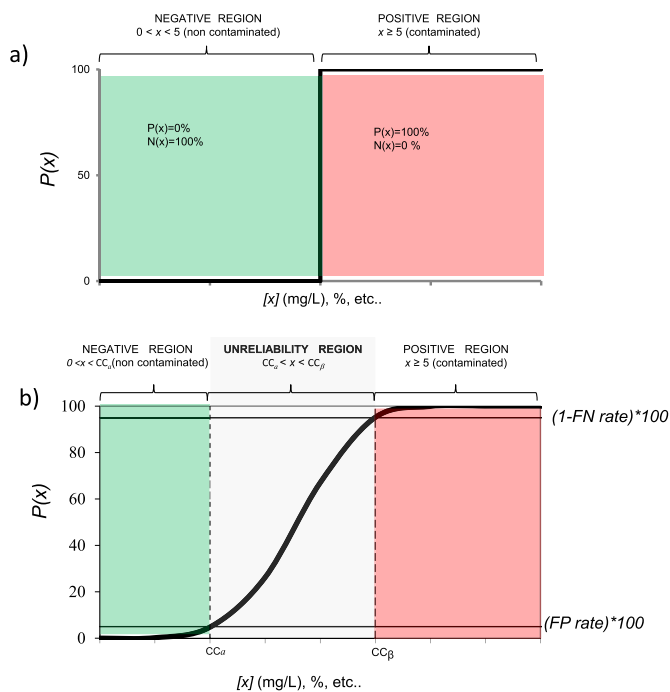
Fig. 3. (a) Ideal graph and (b) real graph for a binary response, the decision criterion of which corresponds to a threshold value ≠ 0. $P(x)$: probability of getting a positive response. $N(x) = 100 − P(x)$: probability of getting a negative response. FN rate: false negative rate. FP rate: false positive rate. CC$_\alpha$: decision limit. CC$_\beta$: detection capability.

(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 used 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.

4.1. 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.

4.2. 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 [31−33]. 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. 4 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. 4a). Both categories A and B have to be defined, obtaining a unique delimiter for sample assignation. When unknown samples are predicted (☆, Fig. 4a) 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. (2), being *x* a sample:

$$\begin{aligned} Positive &: sample \ belongs \ to \ category \ A(Priorat \ wine) \\ Negative &: sample \ belongs \ to \ category \ B(Montseny \ wine) \end{aligned} \quad (2)$$

When class-modelling techniques are applied, the decision criterion is called *model boundary* (Fig. 4b). 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. 4b), 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, thus it is a different wine to the modelled ones. If a sample belongs to both categories, 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. (3), being *x* a sample:

$$\begin{aligned} Positive &: sample \ belongs \ to \ category \ A(Priorat \ wine) \\ Negative &: sample \ belongs \ to \ category \ B(Montseny \ wine) \\ Inconclusive &: sample \ belongs \ to \ both \ categories \ A \ and \ B \\ Different \ from \ modelled &: sample \ is \ neither \ assigned \ to \\ category \ A \ nor \ B \end{aligned} \quad (3)$$

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. 4c shows an example in which only one category is modelled ('Priorat wine'). When unknown samples are predicted
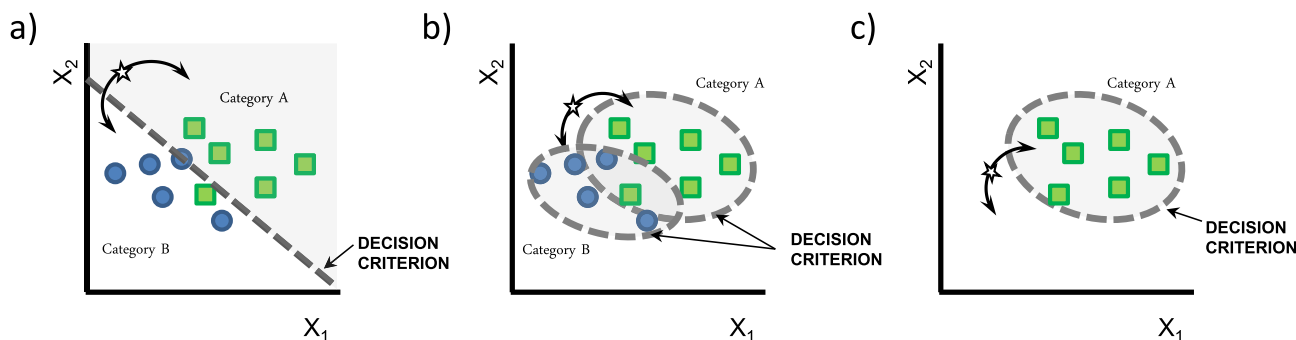

Fig. 4. Example of decision criterion obtained depending on the type of classification technique used. Discriminate technique (a). Class-modelling technique: two classes (b) and one class (c). Samples belonging to Category A (▣), samples belonging to Category B (●) and unknown sample used in prediction (☆).

(★, Fig. 4c), they can be recognised by the model (compliant) or not (non-compliant) Thus, positive and negative outputs are defined according to Eq. (4), being *x* a sample:

Positive : *sample belongs to category A*(*Priorat*; *compliant*)
Negative : *sample does not belong to category*   (4)
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.

4.3. 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 almost always optimistic and sometimes seriously misleading 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 split into *training and test sets*: the training set is used to build the classification rule and the test set to assess prediction ability. There are different ways to split the initial dataset; however, it must be done ensuring the representativeness for each category. 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. [31]. 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. 4b), 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. 4c), 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 [34]. Also, some authors use the term efficiency but it is calculated by means of the geometric mean [35] (Eq. (5)):

$$\sqrt{\frac{TN \cdot TP}{(TN + FP) \cdot (TP + FN)}} \qquad (5)$$

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 [36].

Once the mandatory performance parameters (both autopredictive and predictive) satisfy the requirements, PPC 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 PPC 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 [2,37–39].

The second case study considers a multivariate qualitative methodology based on spectroscopic signals designed to detect adulterants in nuts [40]. Other examples of this kind of method can be found in the following references [36,41–47].

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 in static situation, at the maximum permitted concentration (2.0 ng/g) and without contaminant (0.0 ng/g). Results were successful, both sensitivity and specificity value were 100%.

To build the PPC curve (Fig. 5), 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
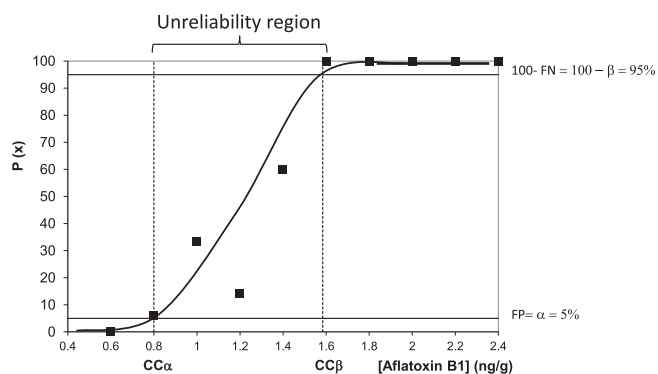
Fig. 5. Adaptation from Ref. [17] with permission. Probability of positive responses, $P(x)$, is plotted versus the concentration levels tested. Upper limit: CC_β; lower limit: CC_α.

performance parameters are estimated. The decision limit (CC_α) 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% and the probability of FP or false non-compliant is lower than 5%. Similarly, the detection capability (CC_β) 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 false negative is almost zero since the detection capability 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.

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 [40]. 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). Since the main interest of the case under study is to describe the compliant samples, the one-class modelling strategy was used (as described in Fig. 4c), therefore the model was just built with the unadulterated samples. From prediction step a binary output is obtained for each sample: the output is positive when a sample is recognised as model compliant and negative when a sample is recognised as model non-compliant.

Table 6
Adapted from Ref. [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

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 obtain the PCC curve, samples at different concentrations of adulterant have to be predicted by the SIMCA model. 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. Then the rates of positive outputs, $p(x)$, are calculated and fitted to a sigmoid function, according to Eq. (6), minimizing the root mean square of the residuals (RMSE).

$$p(x) = \frac{1}{1 + e^{b(x-a)}} \tag{6}$$

Due to the fact that the output is positive when a sample is recognised as model compliant (non adulterated), this curve has an '$\mathcal{Z}$'-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.

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 some definitions are context specific (depend on how positive and negative outputs are defined), so 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 dealing with a binary output to fulfil the validation by establishing performance parameters that involve quantitative information (unreliability region and concentration limits). Nonetheless, multivariate qualitative analysis is usually developed to solve an analytical problem that involves more than two outputs. In such cases, there is still any guideline that summarizes the validation procedure and the performance parameters that should be estimated.

Finally, there is still a lot to be done regarding the transference of advances achieved in qualitative method validation from the research world to the routine laboratories. One reason could be precisely the need to interpret the definitions to each particular case which is not always an easy task. Other reason is that most of the current protocols are thought to detect or identify a compound. In this tutorial, however, we have shown that qualitative methods can be used to assign samples according to a categorical propriety. At the end, routine laboratories are resistant to change their own established protocols. For reasons of time and cost, they prefer its updating instead of including new ones.

Acknowledgements

References

[1] T. Jagric, S. Potrc, T. Jagric, Prediction of liver metastases after gastric cancer resection with the use of learning vector quantization neural networks. Dig. Dis. Sci. 55 (2010) 3252–3261.

[2] N.N. Chidumayo, K. Yoshii, N. Saasa, M. Sakai, H. Kariwa, Development of a tick-borne encephalitis serodiagnostic ELISA using recombinant Fc-antigen fusion proteins. Diagn. Microbiol. Infect. Dis. 78 (2014) 373–378.

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

[4] G.M. Reisfield, B.A. Goldberger, R.L. Bertholf, "False-positive" and "false-negative" test results in clinical urine drug testing, Bioanalysis 1 (2009) 937–952.

[5] S.L.R. Ellison, T. Fearn, Characterising the performance of qualitative analytical methods: statistics and terminology, TrAC – Trends Anal. Chem. 24 (2005) 468–476.

[6] A. Ríos, D. Barceló, L. Buydens, S. Cárdenas, K. Heydorn, B. Karlberg, et al., Quality assurance of qualitative analysis in the framework of the European project 'MEQUALAN', Accredit. Qual. Assur. 8 (2003) 68–77.

[7] E. Trullols, I. Ruisánchez, F.X. Rius, Validation of qualitative analytical methods, 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, U. Örnemark, Eurachem Guide: the Fitness for Purpose of Analytical Methods – a Laboratory Guide to Method Validation and Related Topics, second ed., 2014.

[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, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC technical report), Pure Appl. Chem. 74 (2002) 835–855.

[14] E. Hund, D.L. Massart, J. Smeyers-Verbeke, Inter-laboratory studies in analytical chemistry, 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, Over a century of detection and quantification capabilities in analytical chemistry–historical overview and trends, Talanta 129 (2014) 606–616.

[17] E. Trullols, I. Ruisánchez, F.X. Rius, M. Òdena, M.T. Feliu, Qualitative method for determination of aflatoxin B1 in nuts, J. AOAC Int. 87 (2004) 417–423.

[18] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Uncertainty of results in routine qualitative analysis, TrAC – Trends Anal. Chem. 22 (2003) 647–654.

[19] L.A. Currie, Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995), 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, Perspect. Quantifying Uncertain. Qual. Anal. 123 (1998) 1155–1161.

[24] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Tutorial on evaluation of type I and type II errors in chemical analyses: from the analytical detection to authentication of products and process control, Anal. Chim. Acta 674 (2010) 123–142.

[25] A. Pulido, I. Ruisánchez, R. Boqué, F.X. Rius, Estimating the uncertainty of binary test results to assess their compliance with regulatory limits, Anal. Chim. Acta 455 (2002) 267–275.

[26] R. Song, P.C. Schlecht, K. Ashley, Field screening test methods: performance criteria and performance characteristics, J. Hazard. Mater 83 (2001) 29–39.

[27] R. Macarthur, C. von Holst, A protocol for the validation of qualitative methods of detection, Anal. Methods 4 (2012) 2744.

[28] P. Wehling, R.A. LaBudde, S.L. Brunelle, M.T. Nelson, Probability of detection (POD) as a statistical model for the validation of qualitative methods, J. AOAC Int. 94 (2011) 335–347.

[29] AOAC International Stakeholder Panel on Alternative Methods, Guidelines for Validation of Qualitative Binary Chemistry Methods, Version 12.3, approved at March 2013.

[30] R.A. LaBudde, J.M. Harnly, Probability of identification: a statistical model for the validation of qualitative botanical identification methods, J. AOAC Int. 95 (2012) 273–285.

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

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

[33] 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.

[34] S.A. Drivelos, K. Higgins, J.H. Kalivas, S.A. Haroutounian, C.A. Georgiou, Data fusion for food authentication. Combining rare earth elements and trace metals to discriminate "Fava Santorinis" from other yellow split peas using chemometric tools, Food Chem. 165 (2014) 316–322.

[35] P. Oliveri, G. Downey, Multivariate class modeling for the verification of food-authenticity claims, TrAC – Trends Anal. Chem. 35 (2012) 74–86.

[36] E. Borràs, J.M. Amigo, F. van den Berg, R. Boqué, O. Busto, Fast and robust discrimination of almonds (Prunus amygdalus) with respect to their bitterness by using near infrared and partial least squares-discriminant analysis, Food Chem. 153 (2014) 15–19.

[37] C. de S. Gondim, O.A.M. Coelho, R.L. Alvarenga, R.G. Junqueira, S.V.C. de Souza, An appropriate and systematized procedure for validating qualitative methods: its application in the detection of sulfonamide residues in raw milk, Anal. Chim. Acta 830 (2014) 11–22.

[38] M. Zougagh, H. Téllez, A. Sánchez, M. Chicharro, A. Ríos, Validation of a screening method for rapid control of macrocyclic lactone mycotoxins in maize flour samples, Anal. Bioanal. Chem. 391 (2008) 709–714.

[39] S. Vogliardi, D. Favretto, G. Frison, S. Maietti, G. Viel, R. Seraglia, et al., Validation of a fast screening method for the detection of cocaine in hair by MALDI-MS, Anal. Bioanal. Chem. 396 (2010) 2435–2440.

[40] M.I. López, N. Colomer, I. Ruisánchez, M.P. Callao, Validation of multivariate screening methodology. Case study: detection of food fraud, Anal. Chim. Acta 827 (2014) 28–33.

[41] S.J. Mazivila, F.B. de Santana, H. Mitsutake, L.C. Gontijo, D.Q. Santos, W.B. Neto, Discrimination of the type of biodiesel/diesel blend (B5) using mid-infrared spectroscopy and PLS-DA, Fuel 142 (2015) 222–226.

[42] P. Oliveri, M. Casale, M.C. Casolino, M.A. Baldo, F. Nizzi Grifi, M. Forina, Comparison between classical and innovative class-modelling techniques for the characterisation of a PDO olive oil, Anal. Bioanal. Chem. 399 (2011) 2105–2113.

[43] D.-H. Deng, L. Xu, Z.-H. Ye, H.-F. Cui, C.-B. Cai, X.-P. Yu, FTIR spectroscopy and chemometric class modeling techniques for authentication of Chinese sesame oil, J. Am. Oil Chem. Soc. 89 (2012) 1003–1009.

[44] C. V Di Anibal, M. Odena, I. Ruisánchez, M.P. Callao, Determining the adulteration of spices with Sudan I-II-III-IV dyes by UV-visible spectroscopy and multivariate classification techniques, Talanta 79 (2009) 887–892.

[45] C.V. Di Anibal, I. Ruisánchez, M.P. Callao, High-resolution 1H nuclear magnetic resonance spectrometry combined with chemometric treatment to identify adulteration of culinary spices with Sudan dyes, Food Chem. 124 (2011) 1139–1145.

[46] P. Oliveri, M.I. López, M.C. Casolino, I. Ruisánchez, M.P. Callao, L. Medini, et al., Partial least squares density modeling (PLS-DM) – a new class-modeling strategy applied to the authentication of olives in brine by near-infrared

spectroscopy, Anal. Chim. Acta 851 (2014) 30–36.

[47] M.I. López, E. Trullols, M.P. Callao, I. Ruisánchez, Multivariate screening in food adulteration: untargeted versus targeted modelling, Food Chem. 147 (2014) 177–181.

Maria Isabel López conducted her PhD research in the lab of Assistant professor Itziar Ruisánched and Prof. Maria Pilar Callao at Rovira i Virgili University in Spain. During her research, she worked in chemometrics for pattern recognition, focussing on the use of classification techniques. She has work in the detection of food fraud, considering authentication as well as adulteration problems.

M. Pilar Callao obtained her Ph.D. in chemistry in 1986 at the University of Barcelona. At present she is a full professor in analytical chemistry in the Department of Analytical and Organic Chemistry of the Faculty of Chemistry at the Rovira i Virgili University (Tarragona). She belongs to the group of Chemometrics (currently Chemometrics, qualimetry and nanosensors) since its founding in the late 80s. So, her research is focused on developing new analytical methods than include some chemometric techniques, such as: first-order calibration and related techniques (standardization of instruments); second-order calibration; statistical techniques for monitoring and control; design of experiments and multivariate qualitative analysis and related aspects (selection of variables, data fusion and validation). The areas in which she has applied her work are industrial chemistry, environmental and food chemistry. She has supervised 8 Ph.D. thesis and she has published about 70 articles in international peer reviewed journals.

Itziar Ruisánchez received her BS and Ph.D. (1994) in Chemistry at Universitat Rovira I Virgili, Tarragona (Spain). Then she performed postdoctoral studies and since 1997 she has been working as Associate Professor at the Analytical and Organic Chemistry department of the mentioned university. She is member of the Qualimetrics and Chemometrics group were she is focused on the development of univariate and multivariate qualitative methodologies and their validation. Recent contributions are: one-class modelling, data fusion techniques, variable selection by means of wavelet, qualitative multivariate validation protocol, etc.