



Contents lists available at ScienceDirect

## Trends in Analytical Chemistry

journal homepage: [www.elsevier.com/locate/trac](http://www.elsevier.com/locate/trac)

# Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis



Adriano de Araújo Gomes <sup>a, d, \*\*, \*</sup>, Silvana M. Azcarate <sup>b, c</sup>, Ivan Špánik <sup>d, \*\*\*, \*</sup>,  
Liudmyla Khvalbota <sup>d</sup>, Héctor C. Goicoechea <sup>c, e, \*</sup>

<sup>a</sup> Universidade Federal do Rio Grande do Sul, Instituto de Química, 90650-001, Porto Alegre, RS, Brazil

<sup>b</sup> Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa, Instituto de Ciencias de la Tierra y Ambientales de La Pampa (INCITAP), Av. Uruguay 151, 630 0, Santa Rosa, La Pampa, Argentina

<sup>c</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, CABA, (C1425FQB), Argentina

<sup>d</sup> Institute of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Radlinského 9, 812 37, Bratislava, Slovakia

<sup>e</sup> Laboratorio de Desarrollo Analítico y Quimiometría, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, 3000, Santa Fe, Argentina

## ARTICLE INFO

### Article history:

Received 18 November 2022

Received in revised form

17 May 2023

Accepted 18 May 2023

Available online 26 May 2023

### Keywords:

Food integrity

Safety

Quality

Chemometrics

Classification

## ABSTRACT

According to the Food and Agriculture Organization of the United Nations, food safety is defined as “assurance that food will not cause harm to the consumer which is prepared and/or eaten according to its intended use”. Regulatory authorities, food producers, and consumers are very interested in food authenticity certification, imposing the need to establish new approaches for identifying and assessing food quality markers. In this review, we present a description of the main analytical techniques and data acquisition used in food analysis, the mostly employed pattern recognition chemometric tools, a survey of applications of different methodologies that have been developed up to date, practical examples, and a critical analysis of advantages and disadvantages of the processing multivariate data in food analysis.

© 2023 Elsevier B.V. All rights reserved.

## 1. Introduction

Food quality and authenticity is a permanent problem of food production. Authentic foods must have a defined origin and quality, and they must originate from precisely defined sources. An essential requirement of authentic food is that all food components must be in direct connection to vegetable or animal metabolism from which they originate [1]. On the contrary, there are significant efforts of food producers or sellers to reduce production costs, in other words, to use cheaper, frequently less-valuable materials, while they are offered to the consumers as “full-valuable” under the

same name or brand. One outcome of using these methods is that products may be altered in terms of their character, taste, nutritional value, or overall quality due to the intentional addition of undisclosed compounds. This can also result in the partial or complete restriction of valuable components or their replacement with less nutritionally valuable ones. Other common types of fraud include changes to traditional recipes, declared technological procedures, source materials, or products, as well as the labeling of products with the wrong geographical region or misuse of well-known brands or protected marks.

Additionally, food safety and defense including food crime also must be considered in order to get complex information about particular food commodity. The difference between those terms is intention, whether the act was intentionally (defense) or unintentionally (safety) caused by a simple accident. If changes in food commodity or its composition were done intentionally on purpose to make economical profit or harm particular consumers (act of terror or harm public health) such behavior must be classified as food crime. According to the Food and Agriculture Organization of

\* Corresponding author. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, CABA, (C1425FQB), Argentina.

\*\* Corresponding author. Universidade Federal do Rio Grande do Sul, Instituto de Química, 90650-001, Porto Alegre, RS, Brazil.

\*\*\* Corresponding author.

E-mail addresses: [araujo.gomes@ufrgs.br](mailto:araujo.gomes@ufrgs.br) (A. de Araújo Gomes), [ivan.spanik@stuba.sk](mailto:ivan.spanik@stuba.sk) (I. Špánik), [hgoico@fcb.unl.edu.ar](mailto:hgoico@fcb.unl.edu.ar) (H.C. Goicoechea).

the United Nations, food safety is defined as “the assurance that food will not cause harm to the consumer which is prepared and/or eaten according to its intended use” [2]. Accordingly, World Health Organization [3] aims to enhance at a global and country level the capacity to prevent, detect, and respond to public health threats associated with unsafe food. Even though food safety is related to unintentional contamination of food that makes food harmful to health, intentional acts, such as food fraud, may create food safety issues. Moreover, to prevent harm to consumers, other concepts such as food defense and food quality are interlaced, which intrinsically imply the authenticity of the food, meaning its truthful composition and preparation way [4]. Nowadays, several requirements, such as enhanced nutritional value, satisfactory sensory quality, health-promoting properties, or sustainable and ecological food production, must be met by food industries in order to assure food quality and safety [5]. For this reason, regulatory authorities, food producers, and consumers are very interested in food authenticity certification, imposing the need to establish new approaches for identifying and assessing food quality markers [6].

At the beginning of the recent millennium food authenticity/fraud, food quality and food safety were strictly separated terms that were treated by different approaches. Food authenticity and quality were assessed mostly by methods determining isotopic ratios of stable elements or non-target profiling analytical methods. On the other hand, food safety required precisely defined analytical protocols for target analysis, usually published as internationally recognized standards. Moreover, food safety is focused only on the presence of dangerous chemicals such as allergens, pesticides, pharmaceuticals, polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), dioxins, etc. With the improvement of understanding of food complexity, production processes as well as the role of particular ingredients, it was proven that those terms significantly overlap and some forms of manipulation with food render them unsafe. Thus, the new term food integrity was introduced in 2017 [7] that combines all qualitative, safety and fraud aspects of food commodities. Food integrity refers to the product's adherence to its label in terms of quality, including the presence of the specified constituents in the correct quantities and the absence of any undisclosed substances. Furthermore, when relevant, it fulfills the requirements of the technological process and geographic origin. To put it differently, food integrity is “the product of specification defined such as quality and label claim” [8].

The introduction of the food integrity concept significantly changes requirements on analytical methods, protocols and data evaluation. Traditional methodologies consider a simplest way of evaluation which is the analysis based on qualitative determination followed by quantification only one or some specific compounds. For example, minor organic acids can be used for the authentication of some fruit juices [9]. This simple approach, however, can identify only a specific type of fraud, e.g., decreasing citric/iso-citric acids ratio indicates juice dilution or its adulteration using synthetic citric acid [10].

Therefore, the more frequently used detection of illicit manipulation is based on a comparison of selected criteria with a database. A food authenticity database is a collection of data that have been gathered using established methods taking into account a number of authentic samples to ensure sufficient representativeness. Its purpose is to define the limits of natural variability of defined parameters of a food commodity. One of the most comprehensive and well-known examples of these databases is the EU-Wine databank. In the context of a food authenticity database, an authentic product is required to meet specific parameters that have been established and recorded for each type of product. These parameters include a given range of acceptable values, as well as comments on how to interpret the data. Typically, the permitted

values are based on a particular standard that has been established for the given product [11]. Such databases must be created by independent laboratories and have to be continuously updated. The quality and integrity assessment for a particular food commodity is always a sum of strictly defined parameters and associated with analytical protocols for the determination of selected compounds (authentication markers), which are characteristic of a particular food commodity, and simultaneously different for similar food commodities. The utilization of this way requires as wide as possible knowledge about the physical properties and chemical composition of particular food commodities [12].

Various compounds are used as authentication markers, such as amino acids, volatile organic compounds, major and trace elements, lanthanoids, fatty free acids, organic acids, carbohydrates, anthocyanins, flavonoids, proteins, etc. The development of instrumental methods in past years allowed researchers to also utilize information about isotopic ratios of stable elements, however, regardless of very useful information [13], expensive instrumentation and especially very expensive access to available databases make this method still unattractive for food producers and sellers. Generally, this approach is time-consuming and requires the utilization of many analytical methods and still can give answers only on specific adulteration or manipulation with food products. The broadening scope of the food integrity term also calls for new approaches to its assessment. It is clear, that such a wide term cannot be solved by the determination of one or several parameters. Considering the latter remarks, a plethora of analytical methodologies aimed at food integrity have recently been reported in the scientific literature. In the last decade, the development of analytical techniques, methods and improvement of analytical protocols has substantially improved the capability to reveal and understand the complexity of food systems [5,14]. The main analytical platform used for the detection of a wide range of molecules in different types of foods has been based on separation techniques such as gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), capillary electrophoresis-mass spectrometry (CE-MS). Additionally, several *spectralprint* techniques such as vibrational spectroscopy (mid-infrared (MIR), near-infrared (NIR), and Raman), fluorescence spectroscopy (FS), ultraviolet-visible spectroscopy (UV-Vis), nuclear magnetic resonance (NMR) are among the most reported [15,16].

Considering that foods are challenging heterogeneous systems, which include complex mixtures of different molecular families with a wide range of molecules present at very different concentration levels, it is necessary to use various modern analytical methods that produce a large set of different data. Moreover, multivariate data generation and modeling can be a smart way to gain a better understanding of food integrity. Notwithstanding, extracting essential information from these datasets in their raw form often is too difficult and their assessment and interpretation are usually not straightforward. For that, it should be kept in mind that owing to the acquired data being made up of thousands of variables, it is almost impossible to interpret them without the help of the appropriate statistical mathematical tools, when applied in the context of chemical and/or physical-chemical information modeling is called chemometric.

This review provides a brief description of the main analytical techniques and data acquisition used in food analysis, as well as their characteristics from the perspective of data processing in the context of pattern recognition (PR). Additionally, a roadmap is presented from exploratory analysis to the final model, which includes sampling and partitioning of sample sets, preprocessing, and the most commonly used PR chemometric tools. The review also presents a survey of applications of different methodologies that

have been developed up to date, along with practical examples, and a critical analysis of the advantages and disadvantages of processing multivariate data in food analysis.

## 2. Instrumentation and data acquisition

Food commodities are complex mixtures consisting of numerous chemical compounds present at various concentration levels. To assess food quality and safety, diverse instruments, measurement techniques, experimental designs, and data sources must be employed. However, each type of analytical method, protocol, or instrumental setup utilizes various measurement principles that reflect specific food properties and provides different measured signals converted to data sets followed up treatment by different approaches. One of the most successful treatments consists of the application of a PR algorithm that is often used to classify food samples based on selected criteria [17]. The reliability of classification models depends on a deep understanding of the type, structure, nature, and dimensionality of acquired data to select the most effective chemometric algorithms to solve a classification problem. The obtained data can be classified in different manners as displayed in Fig. 1. The most useful way for understanding the final data structure is to consider the outputs generated by a specific implemented analytical technique.

When final data contain only a single value (e.g., measurement of temperature, pH, or the concentration of a single compound), the data are considered as “zeroth-order” data (scalar). Thus, for  $M$  samples, they can be grouped to form a column vector. In case of multiple pieces of information per sample are recorded (like spectra, chromatogram, or even concentration of different compounds (as depicted in Fig. 1a and b), such data structure corresponds to a row vector of size  $1 \times N$ . While in the first case (Fig. 1a) the total ion chromatogram (TIC) is a continuous profile with a characteristic shape; in the second one (Fig. 1b), the areas are discrete data. Fig. 1d (continuous data vector) and 1e (discrete data vector) are similar scenarios. So, the final data structure becomes a set of independent variables ( $N$ ) measured for  $M$  samples, and a matrix  $\mathbf{X}$  ( $M \times N$ ) is arranged. Depending on the purpose of the analysis, the row vector (row in obtained data matrix) can be treated as “first-order” data. Additionally, also zeroth-order information (column vector) can be extracted from the first-order data matrix (e.g., the area under the curve between specific variables or the peak intensity of a particular variable). In turns, instruments providing two-dimensional data (a full matrix  $\mathbf{X}_{K \times N}$  like depicted in Fig. 1c, f, 1g, 1h and 1k) such as GC-MS, excitation-emission matrix in spectroscopy, etc.) form a set of  $M$  matrices obtained for each analyzed sample which can be arranged at least in three different manners: (i) three-dimensional data tensor  $\mathbf{X}$  ( $M \times K \times N$ ) or bidimensional (ii) unfolded or (iii) augmented matrix. Such data are called “Second-order” data. On this basis, an alternative nomenclature has been suggested in which the data are named according to the number of ways (modes) of an array for a sample set. Then, when more than one sample is analyzed, the recorded zeroth-order data is one-way data; first-order data is two-way data, second-order data is three-way, and so on. Thus, it is often the case that having measurements of a completely different nature, a similar matrix  $\mathbf{X}$  ( $M \times N$ ) can be generated [18–20].

Multidimensional chromatography adds a new temporal dimension to the data by introducing retention times from the second column, resulting in a one-unit increase in data order. Data produced by multidimensional chromatographic systems will have a  $t_1$  (retention times in column 1)  $\times$   $t_2$  (retention times in column 2)  $\times$  detection structure. If the detection involves a single channel that generates only intensities, such as GC  $\times$  GC – FID or LC  $\times$  LC – FLD (with excitation  $-\lambda_{\text{ex}}$  and emission  $-\lambda_{\text{em}}$  at a specific

wavelength) the resulting data will resemble Fig. 1g. However, if the detection records an entire profile, such as GC  $\times$  GC – MS ( $m/z$  fragments) or LC  $\times$  LC – FLD (with excitation at a specific wavelength  $\lambda_{\text{ex}}$  and recording an entire emission spectrum), the data can be organized into a three-way array per sample, which allows a set of samples to be arranged in a four-way array. Other experimental systems can also produce higher-order data, such as EEM recorded at different pH, kinetics of a repeated action at different pH monitored by a spectroscopic technique (time  $\times$  wavelength  $\times$  pH) are examples [19,20].

Data used in food characterization are, in general, multivariate in nature since they consist of a list or array of values according to the complex composition of food matrices. Depending on the analytical technique applied for food analysis, the generated data for a sample could have two, three or more modes (dimensional structure). The structure and arrangement of the datasets are also relevant for the choice of the suitable chemometric tool to be applied to a food issue.

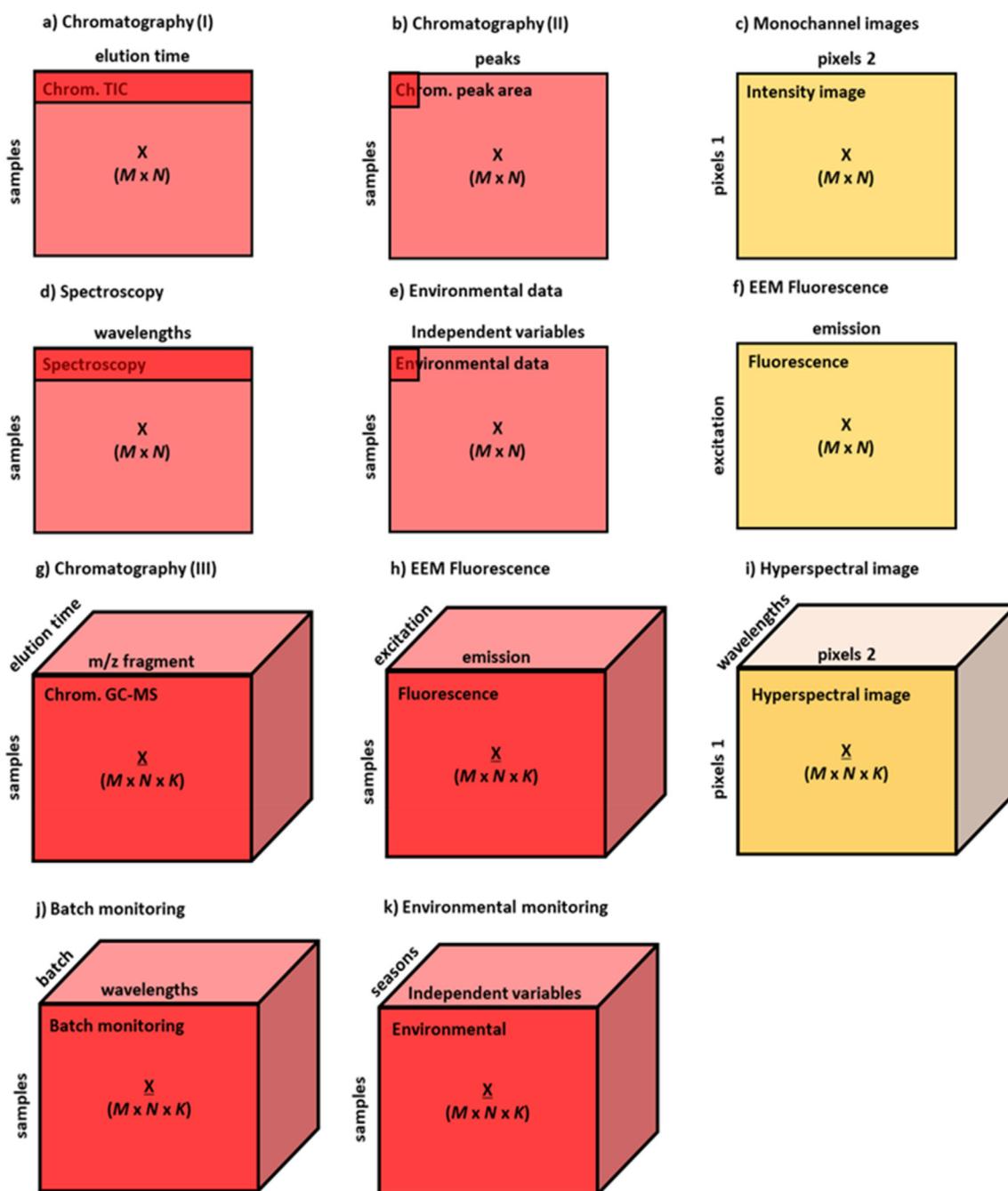
### 2.1. Spectroscopies

A wide variety of spectroscopic techniques such as ultraviolet–visible spectroscopy (UV–Vis), vibrational spectroscopy (infrared (IR), near-infrared (NIR), Raman, fluorescence spectroscopy, nuclear magnetic resonance (NMR), electron spin resonance (ESR), and X-ray spectroscopy have been used in spectral print studies performed on food matrices [18–21]. Most of these techniques require a relatively small amount of sample and performed analyses are non-destructive, fast, easy, direct (solvent-free) and they allow obtaining information about several compounds simultaneously. It is well known that a single wavelength does not provide enough information to cover most authenticity issues. However, these spectroscopic techniques could deliver hundreds of other pieces of information (from single wavelength measurement through recording continuous spectra at many wavelengths) that could be used for both, inferring individual contribution of different chemical components present in a food sample, as well as their interactions. For this reason, statistical analysis of the spectral shape in a sample (obtained in defined instrumental and sample conditions) can be interpreted as a fingerprint itself [22]. In this sense, all above mentioned spectroscopic techniques are able to provide spectral data recorded for a single sample that correspond at least to first-order data. Additionally, some techniques such as multidimensional fluorescence spectroscopy, 2D NMR, etc., are capable to provide second-order data.

### 2.2. Separation techniques

In the last 10 years, there has been a significant increase in the application of PR techniques applied to data obtained from separation techniques coupled to spectroscopic detectors, especially applied to solve food integrity problems [18,23]. The final data outputs from separation methods are chromatograms/electropherograms, which contain separated (fully/partially) peaks. The measured chromatographic/electrophoretic profiles gathered for several samples generate a matrix. If the peaks are integrated and the areas obtained, they become the new variables, consequently, another kind of matrix can be arranged. Both separation profile and peaks areas involve the use of first-order data that can be organized in a matrix containing continuous and discrete variables, respectively.

Another fact that should be considered is similar analytical information presented in both matrices should be treated and visualized as completely different data structures since the natural



**Fig. 1.** Examples of how a matrix  $X$  ( $M \times N$ ) can be obtained by different instruments: a) represents chromatographic profiles, while b) represents the integrated areas of peaks in chromatographic profiles; c) represents a mono channel image; d) a common spectroscopy case; e) arrangement of independent variables in environmental data, and f) the matrix obtained in excitation–emission fluorescence. Also, several three-way structures generated by different instruments: g) hyphenated chromatographic device; h) excitation–emission fluorescence; i) hyperspectral image; j) batch monitoring with spectral devices, and k) environmental monitoring [21]. The data for a sample set is displayed in red, while structured data for a single sample is displayed in yellow. In parts a/d and b/e, the rectangle and square at the top represent a vector and a scalar, respectively.

connection between variables in the first matrix is lost in the second matrix [24]. These approaches are not usually applied in food integrity assessment since foods are complex and dynamic samples, and their multiple constituents generally overlap in the elution/migration time mode. Consequently, several fingerprint approaches have been developed throughout the food chain to monitor food integrity issues. In this context, hyphenation from separation techniques besides spectroscopic techniques such as diode array detection, fast scanning fluorescence detection, or mass

detector have provided many possible analytical strategies for PR studies in foods. Thus, second-order data are generated and modeled; either unfolded (two-way structure) or retained as a three-way structure when several samples are analyzed.

Another fact that should be considered when recording and modeling first- or second-order data obtained from some separation technique is that measurements usually involve some lack of reproducibility in the peak profiles in terms of their shape and elution time. Considering that the subsequent data modeling is

highly dependent on their structure, implementation of adequate data pre-processing is often mandatory prior to modeling. Moreover, considering the complexity of the chromatographic data structure, it is also important to know the linkage between the variables to understand the nature of the data. For instance, data obtained by hyphenated chromatographic devices such as GC with mass spectroscopic detection can be arranged in a three-way structure where the variables are dependent on the elution times and the mass fragments. In LC coupled with DAD detection, for example, variables can be considered independent of each other. It is important to note that the dependence or independence between variables in a data set is a critical point to define a suitable algorithm to model them [25,26].

### 2.3. Sensors

In the last decade, sensor technology was sufficiently developed and became often used to investigate food integrity. Such methods emulate human senses to predict sensory scores of foods. The most common devices for instrumental sensory analyses are electronic noses, electronic tongues and colorimetric techniques that transform some form of input signal acquired from the sample into electric, magnetic, chemical, thermal, or radiation energy. The responses are then correlated to aroma, taste and visual attributes or parameters. The response sensor provides output data in the shape of the vectors. Data obtained from sensors are usually related to food aroma and taste [27].

### 2.4. Image-based methods

Digital images (DI) were not developed with the purpose to be used as an analytical signal, but in the last two decades, there has been witnessed exponential growth of this type of application, including in food authenticity [28]. Probably due to the low cost, simplicity, and possibility of developing out-of-the-lab analytical methodologies. Briefly, a digital image is the capture of a real scene by a device. The process of capturing and storing one involves compressing and discretizing the real scene. This is done using mathematical models to describe color. Color plays a very important role since it is directly related to concentration. Basically, the use of digital images in food authenticity and other areas consists of exploring numerical information from color models as input for multivariate approaches. Probably the best-known color model is the RGB (Red, Green, Blue) space, but others such as grayscale, International Commission on Illumination (CIE), hue-oriented color space (hue-H, saturation-S and value -V) among others have been explored for evaluation of food integrity. A complete description of the currently used color model and mathematical description can be found elsewhere [29]. In order to measure and process digital images for food authenticity analysis, first of all, it is important to make sure that the desired information is related to color change within a set of samples. In this way, a source of variation can be captured by a multivariate approach, otherwise, the attempt to create a new methodology may fail. Supposing that the first requirement is fulfilled and there is no commercial manifold to capture DI for a specific analysis goal, this must be developed. On the other hand, to ensure robustness and reproducibility of recorded DI the lighting, device type (smartphone, camera or scanner), distance from the camera to sample, image resolution and the image format must be taken into account. Recently some home-made manifolds for image acquisition were summarized [30] and a common feature is that almost all of them are closed systems to avoid spurious radiation, with a controlled lighting source. Once the images have been registered and saved, an image processing tool package for the purpose of building multivariate models is

required. Otherwise, own programming routines in some programming language like MatLab®, R or any other can be also developed. A list of free software that can be implemented for these purposes is displayed in Table S1.

### 2.5. Hyperspectral imaging

Although digital color systems have found widespread use in food quality control, the information that can be obtained from a color camera is limited to the vision range of the three-color channels. To overcome this problem, hyperspectral imaging systems have recently emerged as a process analytical tool for food quality assessment, to collect spectral information of one surface in a relatively short time offering a huge amount of spectral and spatial information from a sample. Thus, hyperspectral imaging encompasses the capability of traditional imaging techniques by obtaining spatial images of a sample at a series of contiguous wavelengths attaining both spatial and spectral information and providing reliable and accurate food information [31]. Spectral measurements can be acquired from several spectroscopic techniques, while visible, near or middle infrared (NIR or MIR), or Raman spectroscopies are the most popular. Consequently, hyperspectral data consist of multicomponent systems since the pixels measured seldom contain selective wavelengths for a specific component since they usually contain mixed information of more than one component. A hyperspectral image represents a particular case with respect to its structure. Thus, the data defined by three dimensions, two spatial ( $X$  and  $Y$ ) and one spectral ( $\lambda$ ), can be arranged as a three-dimensional array  $\mathbf{X}$  ( $X \times Y \times \lambda$ ), which is obtained for each sample containing all the chemical information related to the surface measured. Like other sorts of data, hyperspectral data present some artifacts such as spectral noise, spatial interferences, and redundant information, which need to be properly eliminated prior to the chemometric modeling. On the other hand, in the mathematical decomposition concerns, the three-dimensional array is normally unfolded, converting it into a matrix  $\mathbf{X}$  with dimensions ( $MN \times \lambda$ ) with a consequent loss of the spatial relationship. In order to cope with this drawback, some constraints can be added to the mathematical calculations to preserve the spatial relationship. Further information to deepen these relevant points has been widely reported in the literature [32,33]. The final target of hyperspectral analysis is to correlate some features of the individual objects in an image with some properties of interest. In this context, hyperspectral devices and chemometric methods have been successfully implemented in many different areas of food sciences.

## 3. Pattern recognition chemometric tools

Traditionally, analytical methods are implemented to achieve a qualitative/quantitative response related to the presence of macro and/or micronutrients, toxic compounds, and adulterants, among others. Often, methods are developed and applied to target species that act as markers of food integrity [34]. In many cases, food authentication issues can be properly addressed without the need for methods that focus on target compounds; they are the so-called screening methods [35]. Implementing these approaches, issues such as the geographical origin of food, brand, type of technological process used in manufacturing, presence, or absence of undeclared ingredients, etc., can be resolved. PR techniques provide the analyst with a qualitative or semi-quantitative dichotomous answer, based on a previous stage of triangulation and adjustment of a model (linear or not) on a data set containing one or more target classes. This type of strategy does not involve the quantification of a specific chemical species, but the modeling of measured data that represent

part of the composition of the food analyzed. To develop a supervised PR-based methodology with the aim of food integrity analysis, the steps below can be considered as a good way forward:

1. To ask what the best analytical technique is to deal with the problem to be solved. Aspects such as the detection limit of the technique, speed of measurement, robustness, portability and type of information provided should be considered. An example would be the creation of a methodology to detect gluten in foods that are mainly gluten-free but could be contaminated during the production process. While a digital imaging enthusiast might consider utilizing this method, it might not be the most effective strategy due to the target species' nature and the potential for very low concentration levels. Spectroscopic fluorescence may be more suitable.
2. In order to develop a sampling strategy that is sufficiently representative to account for all possible sources of variability, one must consider all factors carefully. This will undoubtedly increase the final model's predictive capability. When planning the sampling process, the analyst should be clear about the method's ultimate objective. Including extraneous sources of variability may decrease the fitness quality.
3. After collecting the necessary samples to cover the desired variability, pretreatment of them is carried out if necessary and the measurement of the chosen instrumental signal is taken. At this point, the analyst has obtained a digital file with a data set that in PR context it would be a matrix  $\mathbf{X}$  (first-order data or several matrices in second- or third-order data), and a vector  $\mathbf{y}$  (class index). The following is performed on these data: (i) initial exploratory analyses to obtain knowledge of the data, (ii) pre-processing in order to remove sources of variability arising from non-idealities of the measurement systems of the analytical signal and (iii) partition of the dataset into, at least, two subsets (training and testing), although in some cases, it can be partitioned into three sets: training, testing, and external validation.
4. The development of data modeling through a suitable classification algorithm. It is crucial to consider certain aspects, such as whether the problem at hand requires discrimination or a modeling approach, and if the data meet the mathematical and statistical requirements of the chosen modeling method.
5. After adjusting the model and properly configuring all target optimization parameters, it is necessary to evaluate the model's predictive capacity against a set of samples that were not used during the fitting and optimization steps.

### 3.1. Sampling and sample set partition

Unlike quantitative methodologies based on multivariate calibration, where the concentration of one or more target species are obtained using a reference method, in the context of PR a representative set of authentic samples must be obtained (authentic here refers that the samples must actually belong to the class to which they were included in the modeling step). This is a critical and challenging point, because in many cases the acquisition of samples with guaranteed authenticity is not a trivial task or even possible. When a qualitative PR method is developed it can be of two types. The first one is based on the detection/quantification of one or more target species. The second way is the analysis of the sample as a whole. In this context, the possibilities of ensuring the veracity of training and validation samples can be categorized into three groups:

- (i) To collect samples from a secure source: in scenarios involving authentication, brand, process, technology and/or

geographical origin, among other similar cases, authentic samples can be collected directly from manufacturers when possible or even in partnership with regulatory agencies in each country.

- (ii) Reference method: when the methodology and development although qualitative involves the concentration of one or more target species, an official reference method can be used to ensure the authenticity of samples as well as to indicate samples that do not meet the quantity requirements.
- (iii) To mimic of non-authentic (fraudulent) samples: a common approach in cases involving tampering issues, where authentic samples are acquired from a secure source and potential tampering is generated in the laboratory.

Once a representative sample set has been collected, there is the task of their partitioning. From the mathematical point of view, it is a finite set partition problem. Briefly, a sample set  $K$  with  $k$  samples should be split into two subsets  $P$  and  $Q$  containing  $p$  and  $q$  non-overlapping samples respectively, which means that  $P \cap Q$  is null and  $p + q = k$ . [36]. The most evident manner to make this partition would be in a random way. However, it is important to note that extreme samples must be included in the training set to avoid extrapolation in the prediction step of a set. The word extreme is being used here in the context of the highest and smallest signals recorded for every sample of a full set. In this context, some methods have been proposed for sample set partitioning as it is shown in Table 1. Most of the sample set partitioning methods were implemented in the multivariate calibration context, considering that many users of these tools are not familiar with the programming. It is necessary to keep in mind that the correct way of arranging and partitioning the dataset must relate to the data modeling step. A schematic representation of the different data partitioning approaches is shown in Fig. 2, where it is possible to note that, unlike multivariate calibration, in the context of PR, the chosen partition method must be used separately for each class. Fig. 2a illustrates the scenario where a set of samples (represented by the yellow square) is considered for building a model taking into account only a single target class in all optimization steps. This model will not involve the use of an alternative class (rigorous way). As can be seen, the sample set partitioning method is applied only to the matrix of samples that make up the target class (represented by the green rectangle) and these are divided into three subgroups: training (light blue rectangle), optimization (violet rectangle) and evaluation set (orange rectangle). The first two subsets of samples are used for the modeling and optimization stages. Then finally the non-target samples are only included in the prediction step once the model is already fitted to check its ability to accept and reject the samples according to whether or not they belong to the target class.

Fig. 2b, in turn, shows a sample set partitioning scheme when the purpose is to model one or more target classes, but the optimization process will involve the use of one or more alternative classes (compliant). Note that in this case both sets of samples (target and non-target) will be submitted to partition as desired, but this procedure is done individually and the subsets of resulting samples are gathered at the end of every partition process. Finally, Fig. 2c shows a case involving modeling via methods where samples of different classes must be included in the training stage (discriminant methods, see Section 3.3).

### 3.2. Pre-processing

Instrumental data recorded from some techniques such as spectroscopy, chromatography etc., usually present sources of physical variation due to non-idealities of measurement systems,

**Table 1**  
Summary of sample selection techniques based on different partitioning of the dataset.

Techniques	Background principle	Remarks	References
Random selection	Randomly.	All samples have the same probability to be selected for the training set. There is no guarantee of the absence of extrapolation in the prediction step.	[36,37]
K&S algorithm	It selects the subset that covers the overall spectral experimental domain based on their Euclidean distance from each other.	When selecting the outermost samples for training the test set results can be very optimistic.	[36,38]
Kernnard-Stone Duplex	It is based on Euclidean distances. It is a variant of K&S algorithm.	When selecting the outermost samples for the training set, the results can be very optimistic. However, this effect is greatly reduced compared to K&S.	[36,39]
PCA	Manual selection based on sample scores distribution in the score plot.	Manual strategy not optimized by defined criteria.	[36,40]
Federov algorithm	Maximizes the determinant matrix of $\mathbf{X}^T\mathbf{X}$ with $\mathbf{X}$ centered a prior.	According to comparative studies, it generates good results for using D-optimal strategy.	[36,41]

also called artifacts, or it is just a dataset in different scales of magnitudes. Anyway, these inconveniences must be properly corrected before data modeling to improve the fit and make the models less complex. Table 2 displays a summary of the pre-processing approaches most often reported in the literature. In many situations, due to the complexity of the data, the use of a single pre-processing technique is not enough for a total correction and/or adequacy of the data. In order of making the pre-processing step more effective, some ensemble approaches have been reported in the literature in the last years; a complete description of that can be found elsewhere [36].

### 3.3. Data modeling

The pre-processed training set undergoes a modeling approach in this stage. It is worth noting that PR techniques can be categorized into two subgroups: unsupervised and supervised methods. Supervised methods, on the other hand, can be further classified as discriminating or modeling. Modeling approaches can be focused on a single class (one class) or multiple classes, and in both cases, the approach can be rigorous or compliant, as shown in Fig. 3 (more details will be provided later). The first subgroup will be discussed here in the context of initials or exploratory analysis as a previous step in the development of a supervised model. The second subgroup will be presented as predictive models in their different forms and recurrent approaches in the food analysis literature. In addition, Table S1 shows a list of several chemometric toolboxes to perform models for PR.

#### 3.3.1. Exploratory algorithms

At the start of developing a PR method, it is crucial to determine whether the recorded data can effectively differentiate samples from various target classes. This action involves assessing the impact of pre-processing and identifying outlier samples. It is also important to examine the degree of overlap between different classes to determine whether linear or nonlinear classifiers are more suitable. Principal component analysis (PCA) is the most widely used method for this purpose, as it can transform high-dimensional data and provide results that can be visualized in two- or three-dimensional spaces (score plots). This allows analysts to identify clusters and patterns within the data. In contrast, hierarchical cluster analysis (HCA) does not require prior data transformation and is based on calculating distances and connections between clusters. However, a variety of distance and cluster connection techniques can be combined in different ways, leading to different results. Certain distances, such as the Mahalanobis distance based on the inverse of the covariance matrix, are only appropriate for low-dimensional data without high multicollinearity [[48],49]. In addition to PCA and HCA, other less exploited tools can also be used for exploratory analysis purposes as

multidimensional scaling (MDS) and different similarity coefficients, where both are based on the distance that is a similarity/dissimilarity between samples calculated by approaches like Euclidean, Mahalanobis, and Manhattan approaches, for example [50].

#### 3.3.2. Classification models

After splitting and pre-processing the dataset, the analyst typically gains some understanding of the overall complexity of the problem through exploratory analysis. The next step is to determine whether a discriminating or modeling approach should be employed. In the realm of classification models, significant progress have been made in distinguishing discriminant from modeling methods to address food integrity issues [51]. Discriminating methods, which use multiclass (at least two) during training step [52,53], are generally more effective than rigorous one-class models in producing higher specificity and sensitivity models. Many researchers prefer to use discriminating models over modeling ones because of the apparent superiority of the former [51,55]. But it is important to keep in mind that the possibilities of non-authentic samples are endless, making adequate sampling impractical. In specific situations where both authentic and false classes can be accurately sampled, discriminant approaches may be used to solve authentication issues. In all other cases, it is recommended to use a one-class classifier, even though lower specificity and sensitivity may be achieved, as it provides more realistic results.

**3.3.2.1. Class-modelling.** In class modeling approaches, each target group is modeled separately, making it possible to create one-class or multiclass models. For instance, when verifying the authenticity of a food sample, the target class is defined as authentic samples and must be sampled representatively. These samples are then used in the training, optimization, and evaluation stages, with the model parameters calculated solely on samples from the target class. Non-authentic or false samples may be included as a non-target class in the optimization step. The resulting models are referred to as compliant, with both sensitivity and specificity available to guide the selection of latent variables and p-values. If only authentic samples are used in the optimization step, the model is considered rigorous, with only sensitivity available to make the decision. In most cases, the modeling parameters representing the class are estimated and used to define an acceptance area at a given degree of statistical significance, as summarized in Table 3 [56–58]. The parameters presented in Table 3 form the acceptance area border, as depicted in Fig. 4.

In Fig. 4, the targeted and untargeted classes are represented by cylinders and cubes, respectively. Fig. 4a depicts a scenario where the training step does not include samples from an untargeted class. This leads to two possibilities: (i) samples belonging to the

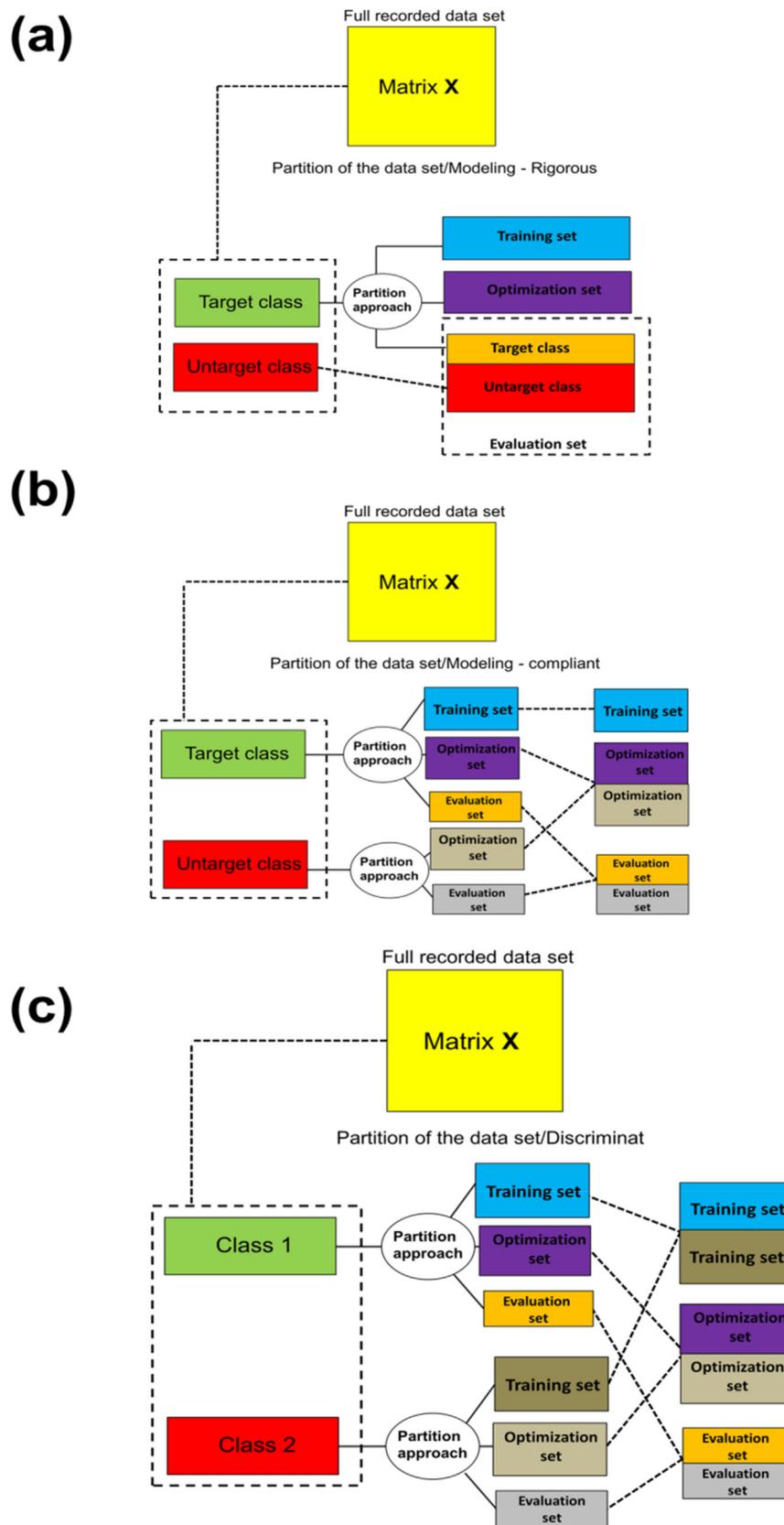
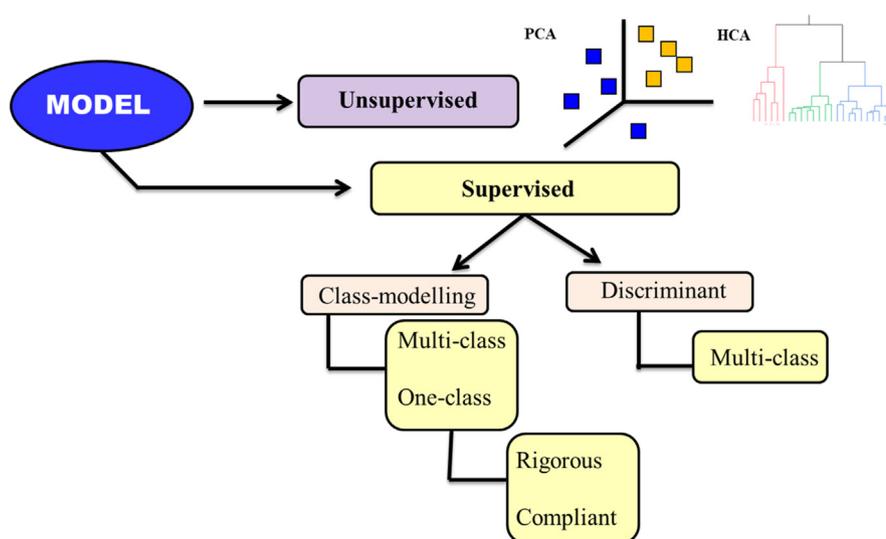


Fig. 2. Schematic diagram of the different forms of arrangement and partitioning of the dataset considering each type of approach that will be adopted: (a) class modeling in rigorous mode; (b) class modeling in compliant mode and (c) modeling by discriminant methods.

**Table 2**  
Summary of pre-processing approaches that have been described in the literature for investigating food integrity.

Issue	Data type	Remarks	Reference
Data scale	All kinds of data.	Autoscaling and normalization.	[36]
Baseline	Infrared, near-infrared, Raman, chromatographic, electrophoresis, nuclear magnetic resonance, and voltammetry.	Spectral derivatives, baseline correction (offset and linear), asymmetric least-squares.	[36]
Multiplicative effects	Infrared, near-infrared, and Raman.	Standard normal variate (SNV), multiplicative scatter correction (MSC), spectral derivatives, and normalization.	[36,42]
Peak shift	Chromatographic, electrophoresis, nuclear magnetic resonance, voltammetry.	Dynamic time warping (DTW) and correlation optimized warping (COW), correlation shifting ( <i>coshift</i> ), intervals correlation shifting ( <i>icoshift</i> ),	[43]
Noise reduction	Infrared, near-infrared, laser-induced breakdown spectroscopy, Raman, nuclear magnetic resonance.	Fourier transform, wavelet transform, smoothing filters, variable selection.	[44]
Missing data	Fluorescence, chromatographic, electrophoresis.	Maximum likelihood PCA-based data imputation.	[45]
Rayleigh (Ry) and Raman (Rm) scatter	Excitation emission fluorescence data.	Asymmetric least-squares, blank subtraction, interpolation-based method, gaussian fit-based method.	[46,47]



**Fig. 3.** Schematic diagram of different types of pattern recognition approaches.

**Table 3**  
Summary of the ways in which acceptance areas are defined in various classification methods, along with the corresponding statistical assumptions.

Model	Parameters	Remark	Reference
<sup>a</sup> SIMCA	Score and orthogonal distance.	Assumes normal and $F$ distribution respectively.	[59]
<sup>b</sup> DD-SIMCA	Total distance	Assume $\chi^2$ distribution.	[60]
<sup>c</sup> PLS-DM	Normalized Q and normalized log PF (probability density).	Q statistics and normal distribution respectively.	[61]
<sup>d</sup> OC-PLS	Absolute centered residual and score distance.	Assumes normal and $F$ distribution respectively.	[62]
<sup>e</sup> M-A CAIMN	Leverage in score space.	Assumes $F$ distribution.	[63]

<sup>a</sup> Soft Independent Model of Class Analogy.

<sup>b</sup> Data Driven – SIMCA.

<sup>c</sup> Partial Least-Squares – Density Modeling.

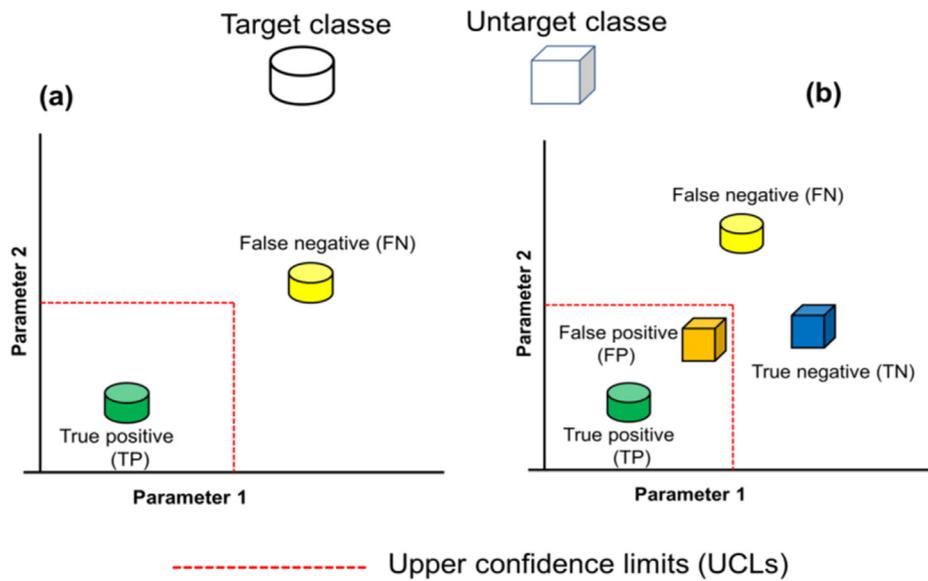
<sup>d</sup> One Class – Partial Least-Squares.

<sup>e</sup> Modeling - Asymmetric (equivalent to a class) Classification and Influence Matrix Analysis method.

targeted class are correctly recognized as such, representing a true positive; or (ii) a sample belonging to the targeted class is not properly recognized, resulting in a false negative as it falls outside the acceptance area.

It is worth noting that modeling approaches rely on a certain level of statistical significance, which can be assessed by the analyst or estimated using data-driven methods. When a significance level of  $\alpha = 0.05$  is selected for the upper confidence limits (UCLs), we can expect 5 false negatives for every 100 samples. This agreement between the calculated and a priori sensitivity can serve as a tool

for model optimization and evaluating overfitting. In the evaluation step, as shown in Fig. 4b, the optimized model is used to predict samples from both targeted and untargeted classes. This can result in four possible outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Refer to Section 3.3.3 for a complete description of the figures of merit and their references. By estimating both sensitivity and specificity, the multiply acceptance area can be calculated if the problem involves multiclass food integrity. The compatibility of the samples to be predicted can then be checked against all available acceptance areas, leading to



**Fig. 4.** Schematic of a typical area of acceptance of modeling methods in a rigorous way: (a) training and (b) evaluation stages respectively. Target and non-target samples are represented as cylinders and cubes in different colors (green cylinders-true positive; yellow cylinders - false negative; blue cube true negative and orange cube false positive).

unambiguous classification in some cases, where a sample belongs to a single class. In other cases, a sample may be recognized to belong to at least two classes or not be assigned to any class. This flexibility of modeling methods makes them suitable for addressing problems where the non-targeted class does not form a well-defined group, as samples may come from different sources and origins, making adequate sampling and modeling challenging or impossible [59–63].

**3.3.2.2. Discriminant models.** The discriminant models differ from class modeling methods as they create a boundary between two or more modeled classes. Therefore, these methods are effective in solving problems where all involved classes are well-defined and have been sampled with sufficient and balanced representation between classes. A relevant example of the use of discriminant models in the field of food integrity is the development of a qualitative analytical method to screen the shelf life of a given product, with the classes being “expired” and “not expired.” It is noteworthy that there is no third possibility for a sample; it is either expired or not. Since there are two clearly defined classes, it is reasonable to consider drawing a boundary between them using discriminant methods. Fig. 5 depicts a visual representation of various discriminant algorithm boundaries.

The literature describes various methods for obtaining boundaries, with linear, quadratic, and nonlinear being the most commonly used [64]. However, all of these methods face the same challenge when a different sample that does not belong to any of the previously modeled classes appears in the prediction set. This results in an incorrect assignment of the sample based on its proximity to one of the modeled classes. To address this issue, modified versions of discriminating methods were developed, such as the soft-PLS-DA method, which defines interval confidence and class boundaries as shown in Fig. 5d. Table 4 lists some of the most widely used discriminant methods, with their routines available (refer to Table S1).

**3.3.2.3. Data fusion approaches.** Nowadays, the concept of data fusion has become a trending topic, especially in the field of food

analysis, mostly to evaluate integrity issues. Indeed, some problems could be solved just by a combination of data sets coming from different instrumental techniques that provide synergistic and/or complementary information. Such an approach is called data fusion and generally is described as an effort to obtain a single result from more than one data source. Currently, three data fusion strategies namely: low-, mid- and high-level have been extensively described in the literature [21,69]. In this context, the combination of two or more data sets obtained from multiple sources provided more accurate knowledge about a sample composition. Consequently, more precise classification and prediction with less uncertainty have been observed compared to results obtained from a single data set [70]. However, there are no standard protocols described in the literature on how to process multisets of data to achieve the final response using data fusion methodologies. Thus, several studies [71–74] have explored the measurements recorded by multiple analytical techniques to increase the possibilities for a better understanding of food integrity. In this context, the data fusion task is not only a direct union of two or more datasets, but the main challenge is a meaningful combination of them followed by suitable pre-processing and finally validation of designed models [75]. Moreover, it is difficult to know at the beginning, which data fusion level is required to solve a given food integrity problem without testing. Usually, starting at a low level (if the data structure allows doing that) and going up to higher levels, if the results achieved are not sufficient [76]. Fig. 6 shows the schematic representation of the data analysis workflow presenting the different data fusion strategies and the main steps of each, depending on the combination in which data occurs: low-level, if the raw data is used as an input in the data fusion procedure; mid-level when the extracted features of the data are used; and high-level when the data are combined at the classification/prediction decision level. However, a wide range of different strategies can be adopted within each level depending mainly on the dimensionality and structure of data, and even multiple levels can be combined.

Low-level data comprises the coupling of datasets in the samples or variable direction depending on the common mode. Since this level retains all the original information from different sources

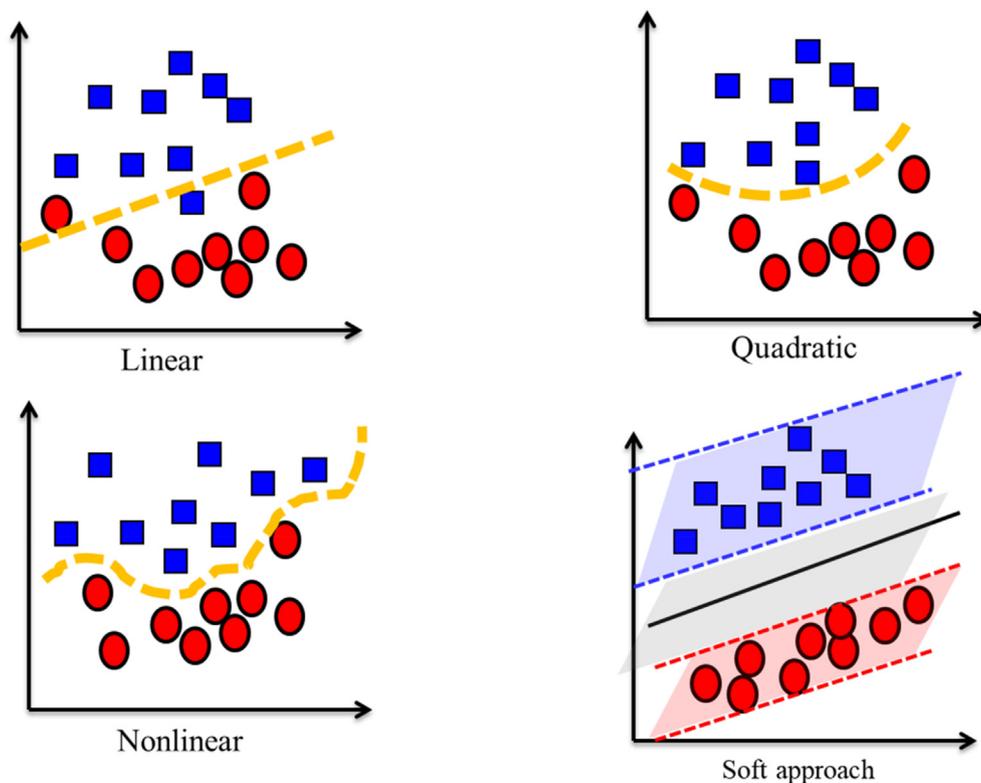


Fig. 5. The sketch of the different forms of boundaries of the discriminant methods.

**Table 4**  
Summary of features in discriminant methods.

Model	Type	Remark	References
LDA <sup>a</sup>	Linear	Requires low-dimensional data or must be combined with variable selection. Homoscedasticity assumption.	[65]
QDA <sup>b</sup>	Quadratic	It assumes that each class follows a Gaussian distribution. It does not assume an identical covariance matrix.	[65]
PLS-DA <sup>c</sup>	Linear	Suitable to large data sets even in presence of multicollinearity. Assumes multivariate normal distribution.	[66]
Soft-PLS-DA <sup>d</sup>	Linear	Suitable to large data sets even in presence of multicollinearity. Assumes multivariate normal distribution.	[66]
PCA-LDA <sup>e</sup>	Linear	Suitable to large data sets even in presence of multicollinearity. Homoscedasticity assumption.	[67]
kNN <sup>f</sup>	Does not require linearity	Non-parametric approach.	[64]
SVMC <sup>g</sup>	Nonlinear	A nonlinear approach to classification problems.	[68]

<sup>a</sup> Linear Discriminant Analysis.

<sup>b</sup> Quadratic Discriminant Analysis.

<sup>c</sup> Partial Least-Squares Discriminant Analysis.

<sup>d</sup> Soft Partial Least-Squares Discriminant Analysis.

<sup>e</sup> Principal Component Analysis - Linear Discriminant Analysis.

<sup>f</sup> K-nearest neighbors' algorithm.

<sup>g</sup> Support-Vector Machines Classification.

also it normally contains noise and redundant information that can affect the proper modeling and calculation, for this reason, data preprocessing steps are usually required. After the data is merged, it is paramount importance that data from different sources have comparable scales, as well as to consider unbalanced number of variables, which should be compensated in order to prevent one block from being dominant in the subsequent data analysis. In this sense, a post preprocessing or normalization can be applied according to the nature of data matrices. This level is mainly applied in zeroth- and first-order data due to the facility in concatenation. For second-order data an unfolding step is usually performed before the concatenation implying losing the variables meaning for the performed models, making the interpretation of results hard. For this reason, other levels are mostly preferred to apply to this kind of data.

Mid-level data fusion is based on a two-step: the extraction of relevant features from individual dataset and the subsequent

concatenation of the output to build a single matrix to be processed (Fig. 6). The first step normally can be performed by calculating latent variables from well-known algorithms such as PCA. However, a straightforward model interpretation in terms of original variables can be accomplished from variable selection methodologies. Noticeably, this level solves the issue of concatenating data blocks of different order. Moreover, it poses more challenges for second-order data since multiple strategies for data extraction, compression, or reduction have been explored for improving outcomes for food analysis. Compared to low-level strategy, improved results have been reported for mid-level data fusion, which could be associated with the features reduction step since the non-informative variance is removed. Nevertheless, it is relevant to consider that this way performs more exhausting methodologies because it applies as many selection algorithms or building as many models as data sets. Furthermore, a more demanding evaluation of the results in terms of the raw variables is required to determine

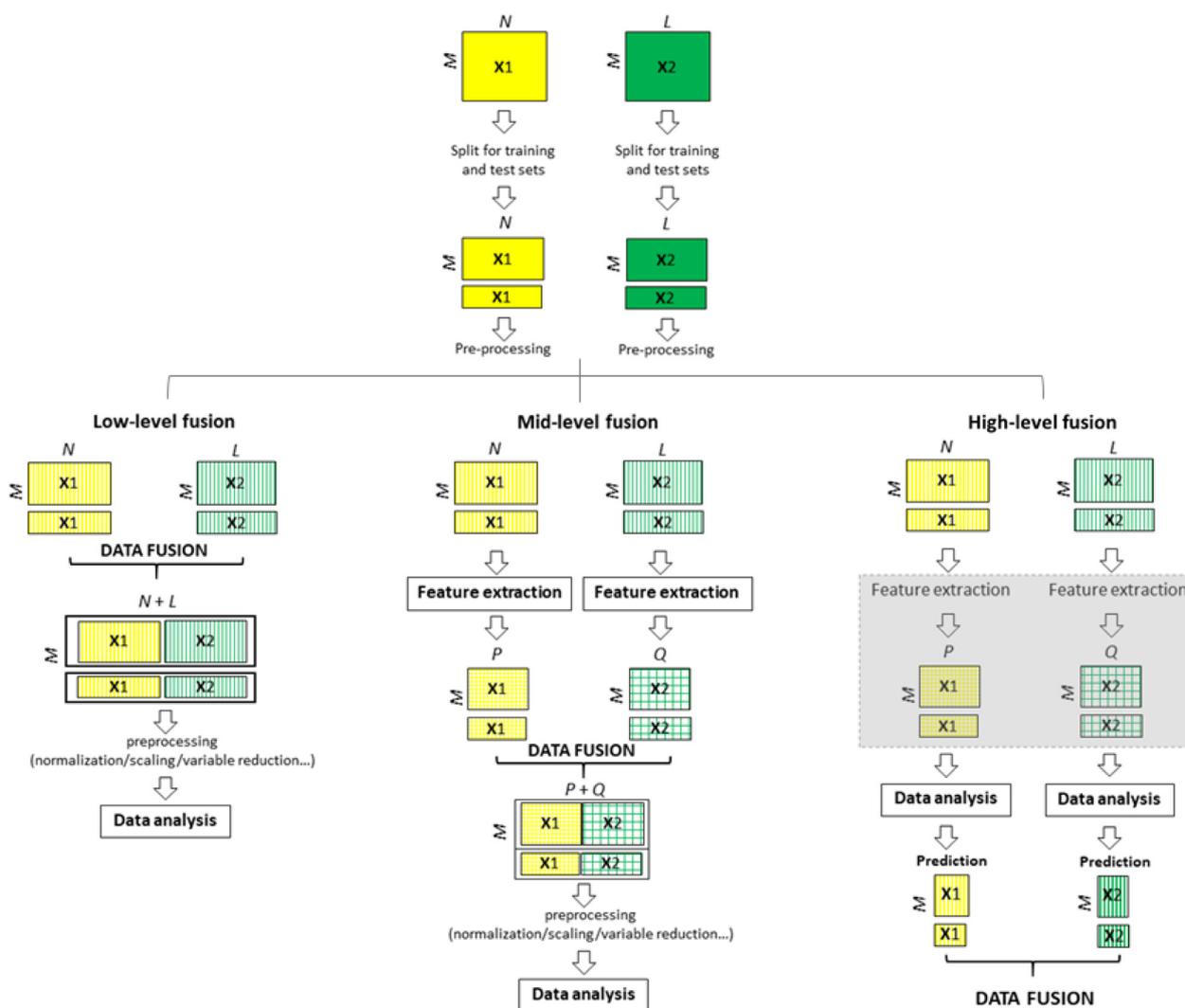


Fig. 6. Schematic representation of data fusion workflow for low-, mid-, and high-levels.

the relationship between the importance of each feature in the final model and its corresponding pattern in terms of the original variables.

High-level data fusion is the most suitable strategy to integrate heterogeneous information, in such a way that each data block is treated independently and then, the outcomes (predictions) are fused to analyze them as a single block. At this level, a new matrix, containing as many rows as samples and as many columns as needed to fulfil the outcomes from the different analyses of the independent data blocks, is created. This data matrix is then analyzed to get a fused final prediction. In this level, to extract relevant features is a possible but not mandatory step (grey box in Fig. 6), which strictly depends on the data structure and data modeling. The new matrix contains as many rows as samples and as many columns as needed to fulfil the outcomes from the different analysis of the independent data. Therefore, this level does not need to adjust an adequate scaling due to each model being fitted independently with its best scaling. However, the order of combining the predictions obtained affects the final decisions.

Most of the data fusion applications in food integrity problems address authentication while a minor extent focused on their adulteration. These applications involve the fusion of data blocks from two or more analytical techniques that provide distinct

information. Different data fusion strategies reported in the literature result in more efficient classification models than the models constructed using the information obtained from each technique separately [77]. In almost all cases, data fusion application is performed using the methods listed in Tables 3 and 4, which were developed in a single block concept. The main drawback of this approach is that common and distinct information in each block might not be explored properly which could affect the final classification. In recent years, increasing attention is devoted to the development of new trends to explore how to efficiently combine data from multiple sources to extract the much information as possible. These new approaches are called “multi-block” methods. During the last 10 years, tools for diverse tasks such as exploratory data analysis, predictive modeling, variable selection, pre-processing optimization, and calibration transfer were sufficiently introduced [78–81]. The main advantage of multi-block data analysis compared to the standard chemometric methods is that they allow a comprehensive understanding of different and common information present in diverse data generated from multiple sources. A summary of multi-block methods available for multi-source data integration in chemometrics has been recently reported in the literature [82,83].

**3.3.2.4. Classification algorithms coupled to variable selection.** The initial step in developing an analytical methodology to control food integrity using PR approaches involves the use of instrumental techniques such as UV–Vis or FTIR, etc., whose signal acquisition is fast [84]. Typically, the analytical signal is collected and evaluated throughout the entire working range of the equipment. Subsequently, regions of excessive noise or those with a complete absence of useful information are often eliminated by simple visual inspection. However, the remaining data typically consists of a large volume, much of which is not chemically linked to the target study (e.g., discriminate or authentic samples). In such scenarios, variable selection (VS) algorithms can be particularly useful for obtaining more robust and interpretable models [74–85]. It should be noted that several reports can be found in the literature regarding variable selection in discriminating models, including routines for VS-LDA, VS-PLS-DA, VS-QDA, and VS-SVM. However, to the best of our knowledge, no variable selection methods are currently coupled with authentication models. A recent review of variable selection in food analysis is available elsewhere [86,87].

### 3.3.3. Model validation

Classification results can be represented in a so-called confusion matrix, which is a square matrix with dimensions ( $G \times G$ ), being  $G$  the number of classes. The confusion matrix contains information about actual and predicted classifications based on a classifier that only assigns each sample to one of the available classes. The matrix element  $nGk$  represents the number of samples belonging to class  $G$  and assigned to class  $k$ . The diagonal values (for example  $nGG$ ) correspond to the number of correct predictions, whereas the values outside the diagonal resemble misclassified samples. From this matrix, several well-established class indices can be derived, such as sensitivity, specificity, and precision. These measures describe the classification results achieved on every single class and should not be taken into consideration, singularly to assess the predictive ability of a model. In fact, since each measure considers different aspects of the overall classification performance, the classification quality should be evaluated considering all of them at the same time [88].

Several measures for the global estimation of classification performances have been proposed in the literature, accuracy being the most known and used [89]. These metrics have been introduced for different applications and may encode different aspects of the classification results. Most classification measures were initially proposed to face binary (two-class) classification tasks (that is, the discrimination of samples in two classes) and were later adapted to deal also with multiclass problems, in which samples are divided into more than two classes. Despite of the confusion matrix containing information on the outcome of the classification modeling, these are generally encoded into one or more classification measures. Three well-known class-based measures (sensitivity, specificity, and precision) can be used to estimate the classification performance achieved in each class. They are calculated on each class separately and encode different aspects of the classification. The sensitivity of a class represents the ability of a given classifier to correctly identify. The precision of a class is defined as the purity of a class, that is, the classifier's ability to avoid wrong predictions in that class. The specificity of a class represents the ability of a classifier to reject samples of other classes. The two-class case is the most found type of classification, as several problems can be represented as binary (yes/no type). When dealing with binary classifications, samples are usually labeled as positive or negative and the confusion matrix is reduced to a  $2 \times 2$  numerical table (see Table 5) [89,90].

Then, the classification measures can be derived from the confusion matrix. All these measures indirectly refer to the positive

class. Sensitivity ( $S_n$ ) is defined as the ratio between  $TP$  and the total number of positive samples:

$$S_n = \frac{TP}{TP + FN} \quad (1)$$

Even in the two class cases, sensitivity can be named in different ways, such as true positive rate ( $TPR$ ), hit rate, or recall. Specificity ( $S_p$ , also known as true negative rate,  $TNR$ ) is defined as the ratio between  $TN$  and the total number of negative samples:

$$S_p = \frac{TN}{TN + FP} \quad (2)$$

Precision ( $Pr$ , also known as positive predictive rate,  $PPR$ ) is calculated as the ratio of  $TP$  and the total number of samples predicted as positive:

$$Pr = \frac{TP}{TP + FP} \quad (3)$$

Accuracy is defined as the ratio of the sum of  $TP$  and  $TN$  over the total number of samples:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Moreover, additional indices can be defined, such as negative predictive value ( $NPV$ ), false positive rate ( $FPR$ ) and false negative rate ( $FNR$ ):

$$NPV = \frac{TN}{TN + FN} \quad (5)$$

$$FPR = \frac{FP}{FP + TN} = 1 - S_p \quad (6)$$

$$FNR = \frac{FN}{TP + FN} = 1 - S_n \quad (7)$$

## 4. Practical examples

### 4.1. Discrimination of edible oil according to the expiration date

This practical example describes the near-infrared spectroscopy to classify soybean oil according to the expiration date. Note that in this case, only two classes are possible, expired and unexpired samples. Thus, both classes can be properly sampled with sufficient representativeness. Therefore, a discriminating classifier matches the goals of this study case. One hundred seventy-seven samples were considered in this example, being 87 non-expired and 90 expired. The oxidation state of soybean oil samples was confirmed by the peroxide index. More details of this case study can be found elsewhere [91].

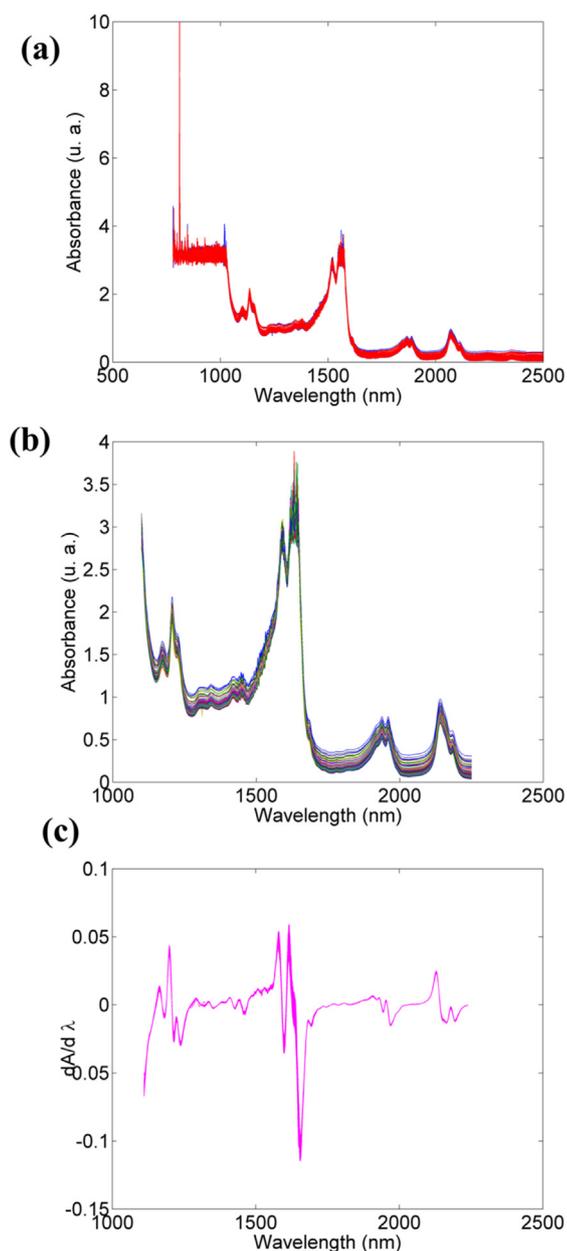
**Table 5**

Typical structure of confusion matrix in a binary classification problem where  $TP$  (true positives) is the number of positive samples correctly predicted as positive,  $TN$  (true negatives) is the number of negative samples correctly predicted as negative,  $FP$  (false positives) is the number of negative samples predicted as positive, and  $FN$  (false negatives) is the number of positive samples predicted as negative.

	Predicted class		
	P	N	
Experimental class	P	TP	FN
	N	FP	TN

As usual, owing to the lack of previous knowledge of the samples, the entire operating range of the equipment was used in the registration of the spectra. Fig. 7a displays the full spectra of the oil samples, which are presented in different colors (expired –yellow- and non-expired –blue-) to emphasize that it is not trivial by simple visual inspection to infer about the kind of samples under investigation, requiring the use of PR models to solve the current issue. However, from the spectra, it is possible to visualize three specific regions related to signal saturation (780–1100 nm), characteristic signal peaks of the samples (1100–2200) and low signal magnitude (2200–2500). Thus, from this analysis, it is possible to note the presence of artifacts that need to be corrected.

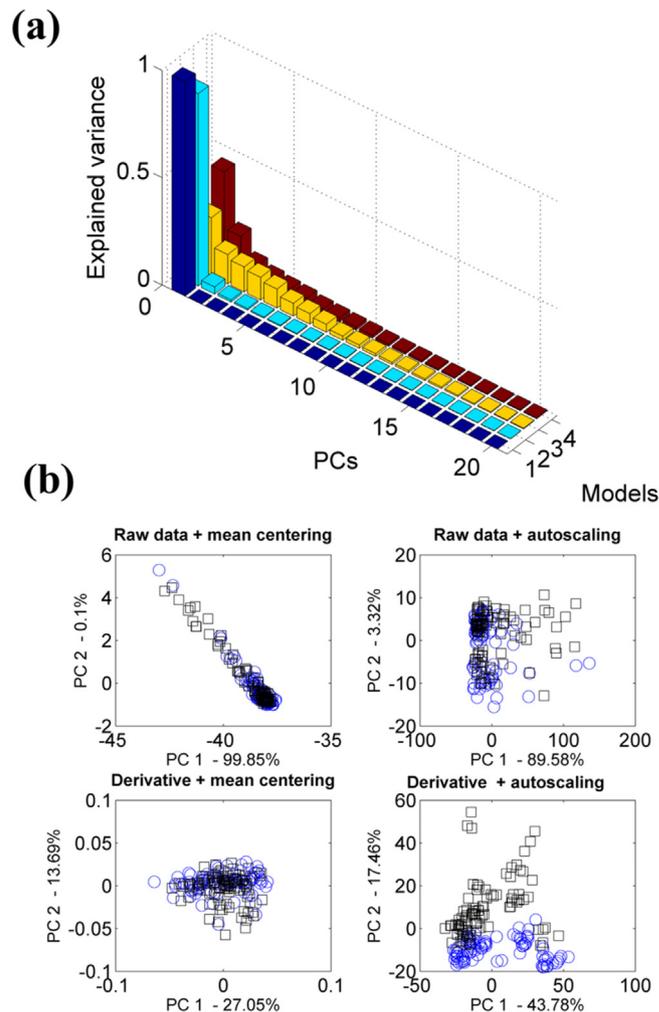
Firstly, the region of interest (ROI) must be defined, and the spectral regions previously identified as non-informative must be



**Fig. 7.** Discrimination of edible oil study case: (a) raw spectra for all samples in blue non-expired and in red expired samples, (b) ROI selected after preliminary inspection, (c) set of all spectra after derivative calculation.

discarded (Fig. 7b). After defining the ROI, the next step is choosing a pre-processing method capable to deal with the artifacts identified in the inspection of Fig. 7a.

Within the options mentioned above in Table 2, the Savitzky-Golay derivative under conditions that represented the best compromise between removing physical sources of variability without resulting in signal deformation/suppression were first derivative, 21 window and second order polynomial) was adopted as a data pre-processing approach and the results are shown in Fig. 7c where it is possible to see that both baseline and multiplicative effects [42] (the latter is generated by some different path-length when radiation interacts with the sample) have been duly corrected. The derivation and smoothing approach of moving filters should be used with caution regarding their parameters to avoid amplification of noise and/or not to deform/suppress some narrower peaks. In an initial work, it is extremely encouraged to evaluate also other possibilities. As it was previously stated, exploratory analysis via PCA (see Fig. 8) is an alternative way to investigate the effect of pre-processing of the data and to check if the data carry the chemical information suitable for distinguishing particular classes.



**Fig. 8.** PCA exploratory step: (a) explained variance for different scenarios (model 1 - top left: raw data + mean centering; model 2 - top right: raw data + mean autoscaling; model 3 - bottom left: derivative data + mean centering; model 4 - bottom right: derivative data + mean autoscaling) and (b) bidimensional PCA score plot (blue circles are unexpired samples and black squares are expired).

Thus, PCA was applied to the data set in four different scenarios: 1) raw +mean centering, 2) raw data autoscaling, 3) derivative +mean centering, and 4) derivative + autoscaling. First, attention should be focused on the fraction of variance explained by each PC. Note that for the first two cases (raw data, see Fig. 8a) the predominant source of variation is the multiplicative effect; this hides the chemical sources of data variability and must be removed. For both cases, the first main component concentrates almost 100% of the variance; it is always good to be suspicious when this occurs. When applying PCA to the derived data it is possible to observe a reasonable distribution of the variance in the other PCs.

Looking at the score plot displayed in Fig. 8b, it is possible to see how harmful the lack of proper pre-processing can be. In the first case (raw data +mean centering) it is possible to observe the multiplicative effect under the spectra, the PC1 versus PC2 scores plot shows practically a straight line. The next case (raw data + auto-scaling) is not useful to solve the problem either due to the significant overlap between studied classes. In the third scenario, it was used derivative +mean centering. Although the derivative approach is efficient for correcting previously identified artifacts, the two classes do not magically separate, as would be expected. On the other hand, a good pattern of separation appears between classes along the axis of PC2 (with about 18% of the explained variance) when combining derivative and autoscaling. These results suggest that the main source of variance does not correspond to the same source of information that distinguishes the classes under study. Therefore, as shown here, it is a matter of back-and-forth work between pre-processing and exploratory analysis to make decisions. Based on those it can be concluded that infrared data are useful to distinguish non-expired from expired samples. Additionally, this chemical information is best reassessed by combining artifact correction (in this case via derivative and autoscaling).

In this stage, it is important to check the presence of outlier samples in the dataset, as they can severely distort the quality of fit. Besides the score graph, PCA also provides another tool such as residual  $Q$  versus Hotelling  $T^2$  that can be useful to identify this sort of samples that must be removed (see Fig. S1). To illustrate the construction of discriminant models, two strategies were chosen: PLS-DA using full spectral information, and LDA combined with Ant of Colony Optimization (ACO) as a variable selection strategy for dimensionality reduction. The dataset was partitioned into three subsets: training (with 50 samples for each class), validation (10 samples for each class) and testing (with 27 unexpired samples and 30 expired samples) using the KS method. The optimal number of latent variables ( $LV$ ) of the PLS-DA model was chosen by cross-validation leave one out (CVLOO) and to minimize an objective function the error rate (ER) was used. The screen plot of ER versus  $LV$  depicted in Fig. S2 shows a well-defined minimum for 5  $LV$ s. This number was selected for the validation step (see Table 6). When the PLS-DA model with 5  $LV$ s was applied to predict the validation set in addition to good training results, similar results were achieved, suggesting that 5 is an adequate number, which gives to the model predictive capacity and apparent absence of both over and under fitting. LDA model coupled to ACO was performed using 300 ants per colony and 500 colonies; the proportion of blind ants per colony was 30%. For more details on variable selection and tune parameters, see Ref. [92]. Variable selection was carried out taking into account the risk minimization of misclassification in cross-validation as defined in Ref. [93]. ACO is a stochastic approach, in this case, it was run 10 times and the best result was selected as the final result. It is important to mention that the ACO parameters were configured based on the authors' previous experience with the dataset. In an initial study, these should be subject to optimization. ACO-LDA final model was based on just five wavelengths

(2102, 1907, 1895, 1400, and 1184 nm, see Fig. S3). That final LDA model was validated; results can be seen in Tables 6 and in Fig. S4.

As can be appreciated in Table 6, both models achieved comparable results. For all cases, the accuracy was always equal to or greater than 90%. In order to compare this difference between the used models, the statistic test  $\chi^2$  [94] was performed showing no significant difference (95% of statistical confidence) between the results obtained by the PLS-DA and ACO-LDA. Thus, both strategies PLS-DA and LDA achieve equivalent results. The first one, based on the variable latent structure can be more robust against noise and small deviation in comparison to LDA. In another hand, the LDA model, based on a few spectral variables can be more suitable to solve food integrity problems in line, even to develop devices based on LEDs and low-cost instrumentation, for instance.

#### 4.2. Classification of selection Slovak Tokaj wine

Tokaj wine is a very famous sweet wine produced in a geographical region with the same name. This small geographical region lies between the territories of Slovakia and Hungary. Slovakia produces different types of Tokaj wine in terms of quality and price ranging from a few euros to a few hundred. The purpose of this case study is to illustrate the use of modeling methods, specifically DD-SIMCA and OC-PLS in the authentication of the Tokaj selection wines (Hungarian equivalent "Aszú") with respect to other Slovak wines and Tokaj wines with lower commercial value, using infrared spectroscopy. A total of 58 samples of Slovakia Tokaj wines from different categories (varietal Tokaj wines, Tokaj cuvée, different Tokaj "putňa" selections and Tokaj essence) and vintages (1959–2017) were selected. The complete description of samples and data acquisition can be found elsewhere [95,96].

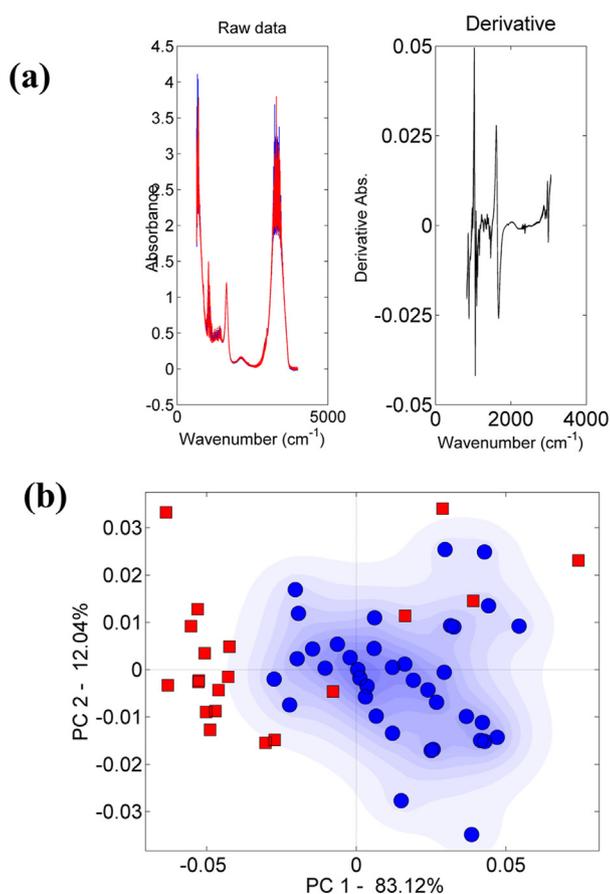
Fig. 9a shows recorded IR spectra of Tokaj wine samples in raw (target class –margent and non-target samples yellow) and derivative spectra (right side). As can be seen in Fig. 9a (left side) both groups of samples have similar spectra in terms of shape and intensity. Moreover, the dataset also presented baseline trends and multiplicative effects. These artifacts were removed using the first derivative and underwent PCA analysis to check the ability of the spectra to distinguish the target samples from the others. The final score plot acquired by PCA is shown in Fig. 9b, where is possible to see slight overlap over the classes suggesting that IR spectra carry the useful chemical information to authenticate Tokaj Selection wines.

Two approaches to one class classification in a rigorous way were selected for this study case: DD-SIMCA and OC-PLS [97]. The sample set was partitioned according to the diagram shown in Fig. S5 in Training (20 samples), Validation (10 samples) and Test sets (30 samples, 10 of the target class and 20 not belonging to the target class) by KS algorithm. The criterion adopted for the selection of the factors in each model was the best compromise between sensitivity and explained variance. The choice for DD-SIMCA was guided by an external validation set containing only samples belonging to the target class in accordance with the rigorous one-class concept (see Fig. S6 and Table S2). It was shown that 4 PC could explain 96.36% of the variance and achieved 87.5% sensitivity in the validation step. At the same time, 100% of sensitivity was achieved in the training set. The acceptance area (see Fig. 10a) was defined by combining score (SD) and orthogonal (OD) distances using different degrees of freedom (SD = 5 and OD = 13) and significance level ( $\alpha = 0.045$ ) estimated via data-driven approach base on  $\chi^2$  distribution. The selection of the suitable number of latent variables in OC-PLS was governed by the standard deviation of residuals (see Fig. S7) calculated via leave-one-out cross-validation (LOOCV). The optimal number of  $LV$  indicated was later validated using external data set. The minimal standard deviation of residuals

**Table 6**  
Figures of merit and confusion matrix for training, validation, and test sets.

		Training			Validation			Test	
		Expired	Non-expired		Expired	Non-expired		Expired	Non-expired
<b>PLS-DA (5)<sup>a</sup></b>	Non-expired (50)	48	2	Non-expired (10)	9	1	Non-expired (30)	25	2
	Expired (50)	0	50	Expired (10)	0	10	Expired (29)	1	29
<b>ACO-LDA (5)**</b>	Non-expired (50)	45	5	Non-expired (10)	10	0	Expired (30)	26	1
	Expired (50)	1	49	Expired (10)	0	10	Non-expired (29)	5	25
	PLS-DA		ACO-LDA	PLS-DA		ACO-LDA	PLS-DA		ACO-LDA
<b>Accuracy</b>	0.98		0.94	0.95		1.00	0.95		0.90
<b>Error rate</b>	0.02		0.06	0.05		0.00	0.05		0.05
<b>Sensitivity</b>	0.96/1.00		0.90/0.98	0.90/1.00		1.0/1.00	0.92/0.97		0.96/0.83
<b>Selectivity</b>	1.00/0.96		0.98/0.90	1.00/0.90		1.00/1.00	0.97/0.92		0.83/0.96

<sup>a</sup> Latent variables and \*\* spectral variables.



**Fig. 9.** Tokaj wine data set, (a) raw (target class –blue and non-target class red) and derivative spectra, (b) PCA score plot (blue circles and red squares are samples of target and non-target class respectively).

was achieved for 3 LVs, while it maintained 100% of sensitivity in both training and validation. Thus, OC-PLS assumes normal distribution for absolute centered residual (ACR), and  $T^2$  Hotelling for SD and acceptance area (see Fig. 10d) is built for a default  $\alpha$  value of 0.05.

As can be seen in the training step for DD-SIMCA and OC-PLS, respectively, all samples fall inside of the acceptance area (see Fig. 10b and e), indicating that the target class is a multimodal group that can be properly modeled by the strategies of a class. When validation samples are predictive using the models, high

sensitivity was achieved in both cases. Only a false negative sample was observed for the DD-SIMCA model. These results agree with over fitting-free model.

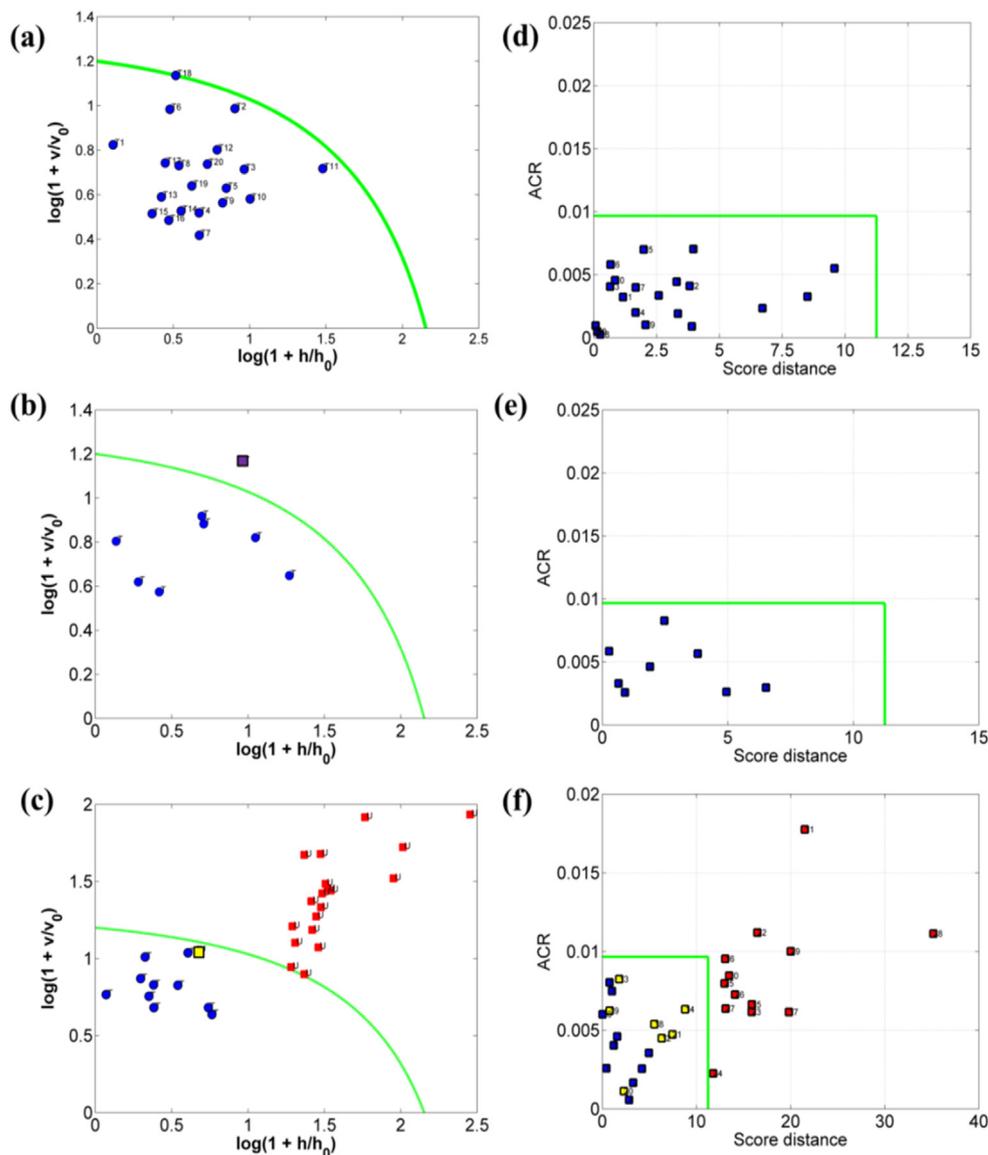
After fitting and validating the models, it is necessary to see their capability to reject samples not belonging to the target class. For that, a completely independent test set was used in this step, containing both target and non-target samples. The statistical summary of the test step can be seen both in Figs. 10c and f also in Table 7.

At this stage, a large difference in performance was observed, although both models exhibited maximum sensibility, only DD-SIMCA also maintained high specificity. This follows from the definition of acceptance areas. DD approach explores the flexibility concept of the SIMCA method, which allows studying important parameters such as degrees of freedom in an empirical way for each dataset. On the other hand, OC-PLS assumes such parameters a priori, and in the presence of deviations, the results can be disastrous.

## 5. Examples from the literature

In addition to the two case studies presented some examples covering more than 190 papers [98–128, 129–150, 151–180, 181–210, 211–240, 241–270, 271–289] published in the last 10 years (from 2013 to 2022) about the application of chemometric tools in the assessment of food integrity that fits in the paper's scope are given in Table 3S. The search parameter and settings applied to Scopus perform included: search term (<Food classification>, <Food discrimination>, <Food authentication>); Search field type (Article title, abstract, keywords); Source type (Journals), Document type (article); Subject areas(Life, health and physical sciences) and Operator used between two search terms (<and>, <or>).

Fig. 11 summarizes the different types of food commodities studied as well as the analytical techniques employed. Most of the PR-based analytical methodologies were developed for wine analysis in the last 10 years [98–110, 111–139] in which the main classification parameters were geographic origin and type of grape variety. Edible vegetable oils [177–200] and milk and cheese [221–237, 239–242] come in second and third place, respectively. Minority products include others such as wheat, barley, soybeans, grain maize, rice, coconut, fruit juice, vinegar, butter, bottled water, hazelnut paste, nectarine, paprika powder, animal fats, peppercorns, pepper powder and rocha pear. From the perspective of the analytical techniques adopted (Fig. 11b) it is possible to note that used the vast majority of cases (72%) they use spectroscopic tools with vibrational spectroscopies that is followed by separation methods (26%). In addition other techniques are reported in smaller



**Fig. 10.** Acceptance area for DD-SIMCA (a, b and c) and OC-PLS (d, e and f) models, for training (a and d), optimization (b and e), and evaluation (c and f). Blue circles (true positive), yellow squares (false positive), red squares (true negative) and violet squares (false negative). The green line is the limit of the acceptance area.

**Table 7**

Statistical summary of test set prediction.

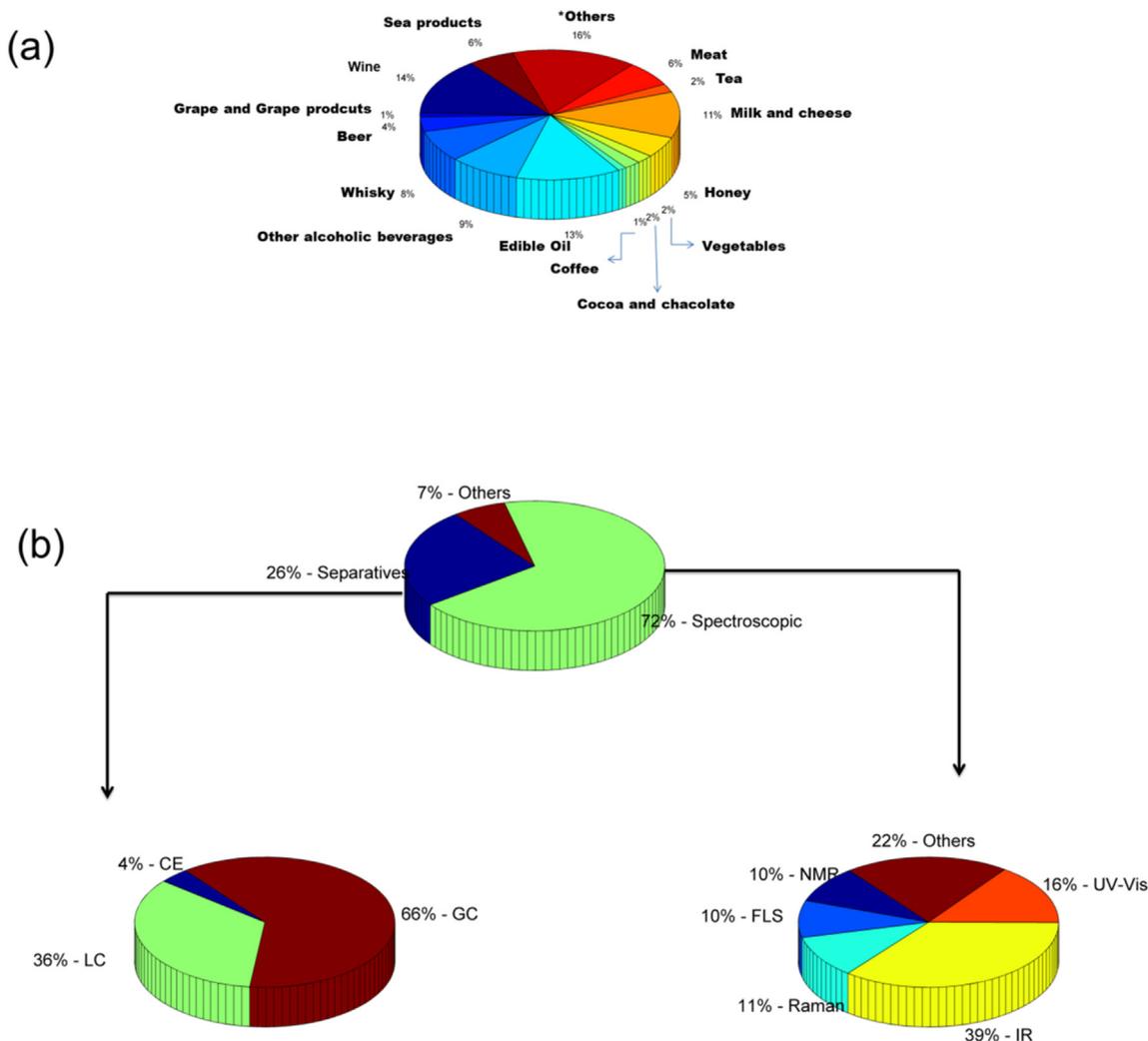
DD-SIMCA <sup>a</sup> (4)			OC-PLS <sup>b</sup> (3)		
Tokaj Selection (10)	Tokaj Selection	Others	Tokaj Selection (10)	Tokaj Selection	Others
Others (20)	1	19	Others (20)	7	13
<b>Sensitivity</b>	1.00			1.00	
<b>Specificity</b>	0.95			0.65	
<b>Efficiency</b>	0.97			0.77	

<sup>a</sup> PC in DD SIMCA model.

<sup>b</sup> LV in OCPLS.

numbers such as: digital image [102,108,251,273], physico-chemical parameters [187], e-sensors (e-eye, e-tongue and e-nose) [121,137], isotope determination [255], radioanalytical method [262], differential scanning calorimetry (DSC) [210] and cyclic voltammetry [217].

After scrutinizing the two major groups of analytical techniques, it is possible to notice that within the separation methods, both liquid and gas chromatography are employed in similar proportions, while capillary electrophoresis [98,120,186] is little explored. In the group of spectroscopic techniques, the vast



**Fig. 11.** Percentage distribution of applications found in the literature in the last 10 years: (a) commodity. \*Others include: Wheat, Barley, Soybeans, Grain Maize, Rice, Coconut, Fruit juice, vinegar, Butter, Bottled water, Hazelnut paste, Nectarine, Paprika powder, Animal fats, Peppercorns, Pepper powder and Rocha pear and (b) Analytical techniques employed. In Fig. 2b, the total percentage can be greater than 100% because in the same article the authors use more than one analytical technique.

majority make use of vibrational spectroscopy (including ATR-FTIR, MID and NIR) the second mostly used technique is molecular absorption spectroscopy in UV-Vis, followed by Raman, molecular fluorescence and NMR. In a reduced number that added up to a total of 22% are included the techniques inductively coupled plasma-mass spectrometry (ICP-MS) [126,158,206], energy dispersive X-ray fluorescence spectrometry (EDXRF) [146], paper spray ionization MS [148,149], desorption atmospheric pressure chemical ionization mass spectrometry [156], microwave plasma-atomic emission spectroscopy (MP-AES) [156] for example. In turn, the chemometric approaches adopted to model those generated in different analytical platforms are summarized in Fig. 12.

Inspection of articles from the last decade indicated that 13% of applications consist only of the initial inspection via exploratory analysis, being the PCA technique the most popular appearing in almost 100% of the cases. On the other hand, 44% of the applications involve fitting a supervised PR model to the data for the purpose of predicting unknown samples. On the other hand, 44% of the applications involve fitting a supervised PR model to the data with the purpose of predicting unknown samples. In 43% of cases, unsupervised PR is adopted as a step prior to supervised PR. When evaluating the type of PR method adopted, it is noted that the

choice of discriminant methods (with PLD-DA being the favorite choice) is much superior to that of modeling methods, even when almost all articles found in the literature are devoted to food authentication. According to the recent literature [35,51–63] such methodologies should be modeling based.

## 6. Conclusion

This review summarizes the use of chemometrics in developing methodologies for food authenticity or quality control, providing a pathway to achieving reliable results that include access to numerous free toolboxes. One significant advantage of classification methods in the context of food authenticity is their ability to implicitly model the partial composition of the chemical constitution of the samples, taking into account the analytical technique used, without the need to identify a group of target compounds for the task. This can result in more comprehensive and efficient methods in comparison to quantitative determinations. In most of the applications described in the literature, the authors frequently note that the food commodity under study has the potential to be fraudulent, but the samples are obtained from the local market without any guarantee of their integrity. Therefore, in the

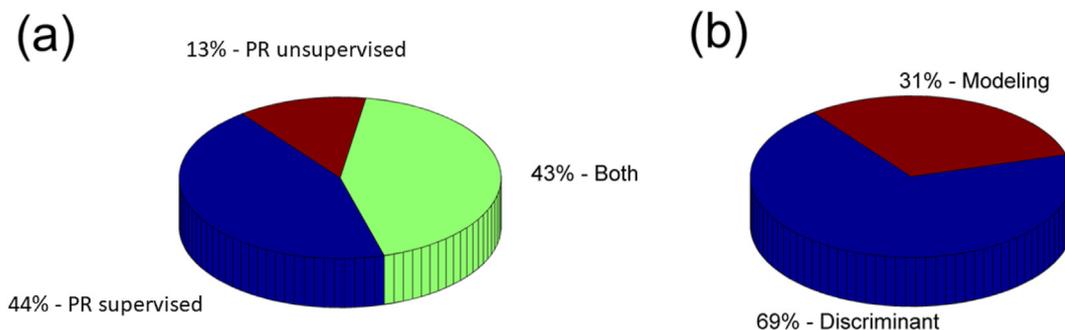


Fig. 12. Chemometric approach adopted: (a) supervised PR versus unsupervised PR and (b) supervised PR approaches.

validation process, it is necessary to use a reference method or samples whose origin and authenticity can be confirmed. However, a reference method or certified samples for classification studies may be rare or nonexistent. Practically unexplored areas in this context include the use of classification employing multiway data, variable selection, and data fusion, particularly with regard to one-class methods.

#### Declaration of competing interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

#### Data availability

No data was used for the research described in the article.

#### Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 945478 and Slovak Research and Development Agency under Contract No.

APVV-21-0211, Universidad Nacional del Litoral (Project CAI + D 2020 - 50620190100020LI), and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project PICT 2020-0105). Dr. Gomes A. A. is thankful to CNPq for his fellowship (313381/2021-6), Universal grants (438734/2018-1) and National Institute of Science and Technology in Molecular Sciences (INCT-CiMol) Project 406804/2022-2/CNPq 58/2022 (National Institutes of Science and Technology Program - INCT).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.trac.2023.117105>.

#### References

- [1] W.S. Alrobaish, L. Jacxsens, P. Spagnoli, P. Vlerick, Assessment of food integrity culture in food businesses through method triangulation, *Food Control* 141 (2022), 109168. <https://doi.org/10.1016/j.foodcont.2022.109168>.
- [2] Food safety - World Health Organization (WHO). <https://www.who.int/health-topics/food-safety>. (Accessed 18 January 2022).
- [3] K. Robson, M. Dean, S. Haughey, C. Elliott, A comprehensive review of food fraud terminologies and food fraud mitigation guides, *Food Control* 120 (2021), 107516. <https://doi.org/10.1016/j.foodcont.2020.107516>.
- [4] F. Ulberth, Tools to combat food fraud – a gap analysis, *Food Chem.* 330 (2020), 127044. <https://doi.org/10.1016/j.foodchem.2020.127044>.
- [5] C.M. Andre, C. Soukoulis, Food quality assessed by chemometrics, *Foods* 9 (2020) 897. <https://doi.org/10.3390/foods9070897>.
- [6] S. Medina, R. Perestrelo, P. Silva, J.A.M. Pereira, J.S. Câmara, Current trends and recent advances on food authenticity technologies and chemometric approaches, *Trends Food Sci. Technol.* 85 (2019) 163–176. <https://doi.org/10.1016/j.tifs.2019.01.017>.
- [7] W.S. Alrobaish, P. Vlerick, N. Steuperaert, L. Jacxsens, An exploratory study on the relation between companies' food integrity climate and employees' food integrity behavior in food businesses, *Foods* 11 (2022) 2657. <https://doi.org/10.3390/foods11172657>.
- [8] W.S. Alrobaish, L. Jacxsens, P. Vlerick, Quantitative study of food integrity climate in Belgian and Saudi Arabian food businesses in view of their organizational characteristics, *Int. J. Food Sci. Technol.* 57 (2022) 4254–4267. <https://doi.org/10.1111/ijfs.15749>.
- [9] W. Simpkins, M. Harrison, The state of the art in authenticity testing, *Trends Food Sci. Technol.* 6 (1995) 321–328. [https://doi.org/10.1016/S0924-2244\(00\)89166-9](https://doi.org/10.1016/S0924-2244(00)89166-9).
- [10] F. Kvasnička, M. Voldřich, P. Pyš, I. Vinš, Determination of isocitric acid in citrus juice—a comparison of HPLC, enzyme set and capillary isotachopheresis methods, *J. Food Compos. Anal.* 15 (2002) 685–691. <https://doi.org/10.1006/jfca.2002.1101>.
- [11] J. Donarski, F. Camin, C. Fahl-Hassek, R. Posey, M. Sudnik, Sampling guidelines for building and curating food authenticity databases, *Trends Food Sci. Technol.* 90 (2019) 187–193. <https://doi.org/10.1016/j.tifs.2019.02.019>.
- [12] E. Anklam, R. Bataglia, Food analysis and consumer protection, *Trends Food Sci. Technol.* 12 (2001) 197–202. [https://doi.org/10.1016/S0924-2244\(01\)00071-1](https://doi.org/10.1016/S0924-2244(01)00071-1).
- [13] L.A. Chessona, D.W. Podlesak, B.R. Erkkila, T.E. Cerling, J.R. Ehleringer, Isotopic consequences of consumer food choice: hydrogen and oxygen stable isotope ratios in foods from fast food restaurants versus supermarkets, *Food Chem.* 119 (2010) 1250–1256. <https://doi.org/10.1016/j.foodchem.2009.07.046>.
- [14] M. Suman, D. Daniele Cavanna, G. Sammarco, F. Lambertini, C. Loffi, Fighting food frauds exploiting chromatography-mass spectrometry technologies: scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities, *TrAC, Trends Anal. Chem.* 142 (2021), 116305. <https://doi.org/10.1016/j.trac.2021.116305>.
- [15] R. Ríos-Reina, J.M. Camiña, Raquel M. Callejón, S.M. Azcarate, Spectralprint techniques for wine and vinegar characterization, authentication and quality control: advances and projections, *TrAC, Trends Anal. Chem.* 134 (2021), 116121. <https://doi.org/10.1016/j.trac.2020.116121>.
- [16] Food and Agriculture Organization (FAO), Food Safety and Quality, 2019. Available online: <http://www.fao.org/food-safety/background/en/>. (Accessed 18 January 2022).
- [17] J.J. Roberts, D. Cozzolino, An overview on the application of chemometrics in food science and technology—an approach to quantitative data analysis, *Food Anal. Methods* 9 (2016) 3258–3267. <https://doi.org/10.1007/s12161-016-0574-7>.
- [18] M.B. Anzardi, J.A. Arancibia, A.C. Olivieri, Processing multi-way chromatographic data for analytical calibration, classification and discrimination: a successful marriage between separation science and chemometrics, *Trends Anal. Chem.* 134 (2021), 116128. <https://doi.org/10.1016/j.trac.2020.116128>.
- [19] G.M. Escandar, H.C. Goicoechea, A. Muñoz de la Peña, A.C. Olivieri, Second- and higher-order data generation and calibration: a tutorial, *Anal. Chim. Acta* 806 (2014) 8–26. <https://doi.org/10.1016/j.aca.2013.11.009>.
- [20] S.M. Azcarate, A. de A. Gomes, A. M. de la Peña, H.C. Goicoechea, Modeling second-order data for classification issues: data characteristics, algorithms, processing procedures and applications, *TrAC, Trends Anal. Chem.* 107 (2018) 151–168. <https://doi.org/10.1016/j.trac.2018.07.022>.

- [21] S.M. Azcarate, R. Ríos-Reina, J.M. Amigo, H.C. Goicoechea, Data handling in data fusion: methodologies and applications, *Trends Anal. Chem.* 116355 (2021). <https://doi.org/10.1016/j.trac.2021.116355>.
- [22] R. Ríos-Reina, J.M. Camiña, R.M. Callejón, S.M. Azcarate, Spectralprint techniques for wine and vinegar characterization, authentication and quality control: advances and projections, *Trends Anal. Chem.* 134 (2021), 116121. <https://doi.org/10.1016/j.trac.2020.116121>.
- [23] L. Toribio, J. Bernal, M.T. Martín, A.M. Ares, Supercritical fluid chromatography coupled to mass spectrometry: a valuable tool in food analysis, *Trends Anal. Chem.* 143 (2021), 116350. <https://doi.org/10.1016/j.trac.2021.116350>.
- [24] J.A. Arancibia, P.C. Damiani, G.M. Escandar, G.A. Ibañez, A.C. Olivieri, A review on second- and third-order multivariate calibration applied to chromatographic data, *J. Chromatogr. B* 910 (2012) 22–30. <https://doi.org/10.1016/j.jchromb.2012.02.004>.
- [25] M.B. Anzardi, J.A. Arancibia, A.C. Olivieri, Processing multi-way chromatographic data for analytical calibration, classification and discrimination: a successful marriage between separation science and chemometrics, *Trends Anal. Chem.* 134 (2021), 116128. <https://doi.org/10.1016/j.trac.2020.116128>.
- [26] F.A. Chiappini, F. Gutierrez, H.C. Goicoechea, A.C. Olivieri, Achieving the analytical second-order advantage with non-bilinear second-order data, *Anal. Chim. Acta* 1181 (2021), 338911. <https://doi.org/10.1016/j.aca.2021.338911>.
- [27] F. Oveissia, L.H. Nguyena, J.E. Giaretta, Z. Shahrbabakia, R.J. Ratha, V.A. Apalanyac, J. Yund, F. Dehghanian, S. Naficya, Sensors for food quality and safety, in: *Food Engineering Innovations across the Food Supply Chain*, 2022. <https://doi.org/10.1016/B978-0-12-821292-9.00010-8> (Chapter 23).
- [28] M. Meenu, C. Kurade, B.C. Neelapu, S. Kalra, H.S. Ramaswamy, Y. Yu, A concise review on food quality assessment using digital image processing, *Trends Food Sci. Technol.* 118 (2021) 106–124. <https://doi.org/10.1016/j.tifs.2021.09.014>.
- [29] L.F. Capitán-Vallvey, N. López-Ruiz, A. Martínez-Olmos, M.M. Erenas, A.J. Palma, Recent developments in computer vision-based analytical chemistry: a tutorial review, *Anal. Chim. Acta* 899 (2015) 23–56. <https://doi.org/10.1016/j.aca.2015.10.009>.
- [30] P.H.G.D. Diniz, Chemometrics-assisted color histogram-based analytical Systems, *J. Chemom.* 34 (2020), e3242. <https://doi.org/10.1002/cem.3242>.
- [31] V. Olmos, L. Benítez, M. Marro, P. Loza-Alvarez, B. Piña, R. Tauler, A. de Juan, Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images, *Trends Anal. Chem.* 94 (2017) 130–140. <https://doi.org/10.1016/j.trac.2017.07.004>.
- [32] C.S. Silva, M.F. Pimentel, J.M. Amigo, R.S. Honorato, C. Pasquini, Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models, *Trends Anal. Chem.* 95 (2017) 23–25. <https://doi.org/10.1016/j.trac.2017.07.026>.
- [33] J.M. Amigo, H. Babamoradi, S. Elcoroaristizabal, Hyperspectral image analysis. A tutorial, *Anal. Chim. Acta* 896 (2015) 31–51. <https://doi.org/10.1016/j.aca.2015.09.030>.
- [34] P. Oliveri, Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues e A tutorial, *Anal. Chim. Acta* 982 (2017) 9–19. <https://doi.org/10.1016/j.aca.2017.05.013>.
- [35] R. Munóz-Olivas, Screening analysis: an overview of methods applied to environmental, clinical and food analyses, *Trends Anal. Chem.* 23 (2004) 203–216. [https://doi.org/10.1016/S0165-9936\(04\)00318-8](https://doi.org/10.1016/S0165-9936(04)00318-8).
- [36] M. Zeaiter, D. Rutledge, Preprocessing methods in comprehensive chemometrics, *Chem. Biochem. Data Anal.* 3 (2009) 123–231.
- [37] E. Bouveresse, C. Hartmann, D.L. Massart, Standardization of near-infrared spectrometric instruments, *Anal. Chem.* 68 (1996) 982–990. <https://doi.org/10.1021/ac9510595>.
- [38] R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, A method for calibration and validation subset partitioning, *Talanta* 67 (2005) 736–740. <https://doi.org/10.1016/j.talanta.2005.03.025>.
- [39] R.D. Sneek, Validation of regression models: methods and examples, *Technometrics* 19 (1977) 415–428. <https://doi.org/10.2307/1267881>.
- [40] P.D. Wentzell, Kompany-Zarehab Giglio, M. Beyond, Principal components: a critical comparison of factor analysis methods for subspace modelling in chemistry, *Anal. Methods* 13 (2021) 4188–4219. <https://doi.org/10.1039/D1AY01124C>.
- [41] F. Pukelsheim, *Optimal Design of Experiments*, John Wiley & Sons, New York, 1993.
- [42] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common preprocessing techniques for near-infrared spectra, *Trends Anal. Chem.* 28 (2009) 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.
- [43] S.J. Mazivila, J.M. Lombardi, R.N.M.J. Páscoa, S.A. Bortolato, J.M.M. Leitão, J.C.G. Esteves da Silva, Three-way calibration using PARAFAC and MCR-ALS with previous synchronization of second-order chromatographic data through a new functional alignment of pure vectors for the quantification in the presence of retention time shifts in peak position and shape, *Anal. Chim. Acta* 1146 (2021) 98–108. <https://doi.org/10.1016/j.aca.2020.12.033>.
- [44] M.F. Wahab, F. Gritti, T.C. O'Haver, Discrete Fourier transform techniques for noise reduction and digital enhancement of analytical signals, *Trends Anal. Chem.* 143 (2021), 116354. <https://doi.org/10.1016/j.trac.2021.116354>.
- [45] F.A. Adnan, K.R. Jamaludin, W.Z.A. Wan Muhamad, S. Miskon, A review of the current publication trends on missing data imputation over three decades: direction and future research, *Neural Comput. Appl.* 34 (2022) 18325–18340. <https://doi.org/10.1007/s00521-022-07702-7> (0123456789(
- ),-volIV)(0123456789.
- [46] V.I. Korepanov, Asymmetric least-squares baseline algorithm with peak screening for automatic processing of the Raman spectra, *J. Raman Spectrosc.* 51 (2020) 2061–2065. <https://doi.org/10.1002/jrs.5952>.
- [47] F.A. Chiappini, M.R. Alcaraz, H.C. Goicoechea, A.C. Olivieri, A graphical user interface as a new tool for scattering correction in fluorescence data, *Chemometr. Intell. Lab. Syst.* 193 (2019), 103810. <https://doi.org/10.1016/j.chemolab.2019.07.009>.
- [48] R. Bro, A.K. Smilde, Principal component analysis, *Anal. Methods* 6 (2014) 2812–2831, 10.1039/c3ay41907.
- [49] R.G. Brereton, The Mahalanobis distance and its relationship to principal component scores, *J. Chemometr.* 29 (2015) 143–145. <https://doi.org/10.1002/cem.2692>.
- [50] P. Zerzucha, B. Walczak, Concept of (dis)similarity in data. Analysis, *Trends Anal. Chem.* 38 (2012) 116–128. <https://doi.org/10.1016/j.trac.2012.05.005>.
- [51] O.Y. Rodionova, A.V. Titova, A.L. Pomerantsev, Discriminant analysis is an inappropriate method of authentication, *Trends Anal. Chem.* 78 (2016) 17–22. <https://doi.org/10.1016/j.trac.2016.01.010>.
- [52] P. Oliveri, G. Downey, Multivariate class modeling for the verification of food-authenticity claims, *Trends Anal. Chem.* 35 (2012), 110.1016/j.trac.2012.02.005.
- [53] P. Oliveri, C. Malegori, E. Mustorgi, M. Casale, Qualitative pattern recognition in chemistry: theoretical background and practical guidelines, *Microchem. J.* 162 (2021), 105725. <https://doi.org/10.1016/j.microc.2020.105725>.
- [54] A.L. Pomerantsev, O.Y. Rodionova, New trends in qualitative analysis: performance, optimization, and validation of multi-class and soft models, *Trends Anal. Chem.* 143 (2021), 116372. <https://doi.org/10.1016/j.trac.2021.116372>.
- [55] O.Y. Rodionova, A.L. Pomerantsev, Chemometric tools for food fraud detection: the role of target class in non-targeted analysis, *Food Chem.* 317 (2020), 126448. <https://doi.org/10.1016/j.foodchem.2020.126448>.
- [56] A.L. Pomerantsev, Acceptance areas for multivariate classification derived by projection methods, *J. Chemom.* 22 (2008) 11–12. <https://doi.org/10.1002/cem.1147>.
- [57] A.L. Pomerantseva, O.Y. Rodionova, Concept and role of extreme objects in PCA/SIMCA, *J. Chemom.* (2012). <https://doi.org/10.1002/cem.2506>.
- [58] A.L. Pomerantsev, O.Y. Rodionova, On the type II error in SIMCA method, *J. Chemom.* 28 (2014) 429–438. <https://doi.org/10.1002/cem.2610>.
- [59] M. Forina, P. Oliveri, S. Lanteri, M. Casale, Class-modeling techniques, classic and new, for old and new problems, *Chemometr. Intell. Lab. Syst.* 93 (2008) 132–148. <https://doi.org/10.1016/j.chemolab.2008.05.003>.
- [60] Y.v. Zontov, O.Y. Rodionova, S.v. Kucheryavskiy, A.L. Pomerantsev, DD-SIMCA – a MATLAB GUI tool for data driven SIMCA approach, *Chemometr. Intell. Lab. Syst.* 167 (2017) 23–28. <https://doi.org/10.1016/j.chemolab.2017.05.010>.
- [61] 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 to the authentication of olives in brine by near-infrared spectroscopy, *Anal. Chim. Acta* 851 (2014) 30–36. <https://doi.org/10.1016/j.aca.2014.09.013>.
- [62] L. Xu, M. Goodarzi, W. Shi, C.B. Cai, J.H. Jiang, A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers, *Chemometr. Intell. Lab. Syst.* 139 (2014) 58–63. <https://doi.org/10.1016/j.chemolab.2014.09.005>.
- [63] M. Forina, M. Casale, P. Oliveri, S. Lanteri, CAIMAN brothers: a family of powerful classification and class modeling techniques, *Chemometr. Intell. Lab. Syst.* 96 (2009) 239–245. <https://doi.org/10.1016/j.chemolab.2009.02.006>.
- [64] R.G. Brereton, Pattern recognition in chemometrics, *Chemometr. Intell. Lab. Syst.* 149 (2015) 90–96. <https://doi.org/10.1016/j.chemolab.2015.06.012>.
- [65] R.G. Brereton, One-class classifiers, *J. Chemom.* 25 (2011) 225–246. <https://doi.org/10.1002/cem.1397>.
- [66] Y.v. Zontov, O.Y. Rodionova, S.v. Kucheryavskiy, A.L. Pomerantsev, PLS-DA – a MATLAB GUI tool for hard and soft approaches to partial least squares discriminant analysis, *Chemometr. Intell. Lab. Syst.* 203 (2020), 104064. <https://doi.org/10.1016/j.chemolab.2020.104064>.
- [67] V. E. de Almeida, D.D. de S. Fernandes, P.H.G. Diniz, A. de Gomes, G. Vêras, R.K.H. Galvão, M.C.U. Araújo, Scores selection via Fisher's discriminant power in PCA-LDA to improve the classification of food data, *Food Chem.* 363 (2021), 130296. <https://doi.org/10.1016/j.foodchem.2021.130296>.
- [68] R.G. Brereton, G.R. Lloyd, Support vector machines for classification and regression, *Analyst* 135 (2010) 230–267. <https://doi.org/10.1039/b918972f>.
- [69] P. Mishra, J.M. Roger, D. Jouan-Rimbaud-Bouveresse, A. Biancolillo, F. Marini, A. Nordon, D.N. Rutledge, Recent trends in multi-block data analysis in chemometrics for multi-source data integration, *Trends Anal. Chem.* 137 (2021), 116206. <https://doi.org/10.1016/j.trac.2021.116206>.
- [70] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Acuña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment – a review, *Anal. Chim. Acta* 891 (2015) 1–14. <https://doi.org/10.1016/j.aca.2015.04.042>.
- [71] A.K. Smilde, I. Måge, T. Næs, T. Hankemeier, M.A. Lips, H.A.L. Kiers, E. Acar, R. Bro, Common and distinct components in data fusion, *J. Chemom.* 31 (2017), e2900. <https://doi.org/10.1002/cem.2900>.
- [72] A. Biancolillo, R. Bucci, A.L. Magri, A.D. Magri, F. Marini, Data-fusion for multiplatform characterization of an Italian craft beer aimed at its

- authentication, *Anal. Chim. Acta* 820 (2014) 23–31. <https://doi.org/10.1016/j.aca.2014.02.024>.
- [73] R. Ríos-Reina, R.M. Callejón, F. Savorani, J.M. Amigo, M. Cocchi, Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars, *Talanta* 198 (2019) 560–572. <https://doi.org/10.1016/j.talanta.2019.01.100>.
- [74] X. Li, M. Cai, M. Li, X. Wei, Z. Liu, J. Wang, K. Jia, Y. Han, Combining Vis-NIR and NIR hyperspectral imaging techniques with a data fusion strategy for the rapid qualitative evaluation of multiple qualities in chicken, *Food Control* 145 (2023), 109416. <https://doi.org/10.1016/j.foodcont.2022.109416>.
- [75] M.P. Campos, M.S. Reis, Data preprocessing for multiblock modelling – a systematization with new methods, *Chemometr. Intell. Lab. Syst.* 199 (2020), 103959. <https://doi.org/10.1016/j.chemolab.2020.103959>.
- [76] P. Mishra, J.M. Roger, D.N. Rutledge, A. Biancolillo, F. Marini, A. Nordon, D. Jouan-Rimbaud-Bouveresse, MBA-GUI: a chemometric graphical user interface for multi-block data visualization, regression, classification, variable selection and automated pre-processing, *Chemometr. Intell. Lab. Syst.* 205 (2020), 104139. <https://doi.org/10.1016/j.chemolab.2020.104139>.
- [77] T. Skotare, D. Nilsson, S. Xiong, P. Geladi, J. Trygg, Joint and unique multiblock analysis for integration and calibration transfer of NIR instruments, *Anal. Chem.* 91 (2019) 3516–3524. <https://doi.org/10.1021/acs.analchem.8b05188>.
- [78] F. Ahmed, I.D. Fuller, D.P. Killeen, S.J. Fraser-Miller, K.C. Gordon, Raman and infrared spectroscopic data fusion strategies for rapid, multicomponent quantitation of krill oil compositions, *ACS Food Sci. Technol.* (2021) 570–578. <https://doi.org/10.1021/acfoodsctech.0c00139>.
- [79] A.A. Gomes, L. Khvalbota, L. Onça, A. Machynáková, I. Špánik, Handling multiblock data in wine authenticity by sequentially orthogonalized one class partial least squares, *Food Chem.* 382 (2022), 132271. <https://doi.org/10.1016/j.foodchem.2022.132271>.
- [80] E. Aguado-Sarrió, J.M. Prats-Montalbán, R. Sanz-Requena, C. Duchesne, A. Ferrer, Sequential multiblock partial least squares discriminant analysis for assessing prostate cancer aggressiveness with multiparametric magnetic resonance imaging, *Chemometr. Intell. Lab. Syst.* 226 (2022), 104588. <https://doi.org/10.1016/j.chemolab.2022.104588>.
- [81] G. Robert, R. Gosselin, Evaluating the impact of NIR pre-processing methods via multiblock partial least-squares, *Anal. Chim. Acta* 1189 (2022), 339255. <https://doi.org/10.1016/j.aca.2021.339255>.
- [82] P. Mishra, K.H. Liland, Swiss knife partial least squares (SKPLS): one tool for modelling single block, multiblock, multiway, multiway multiblock including multi-responses and meta information under the ROSA framework, *Anal. Chim. Acta* 1206 (2022), 339786. <https://doi.org/10.1016/j.aca.2022.339786>.
- [83] E.T. Mangamana, R.G. Kakaí, E.M. Qannari, A general strategy for setting up supervised methods of multiblock data analysis, *Chemometr. Intell. Lab. Syst.* 217 (2021), 104388. <https://doi.org/10.1016/j.chemolab.2021.104388>.
- [84] A. de Gomes, S.M. Azcarate, P.H.G. Diniz, D.D. de S. Fernandes, G. Veras, Variable selection in the chemometric treatment of food data: a tutorial review, *Food Chem.* 370 (2022), 131072. <https://doi.org/10.1016/j.foodchem.2021.131072>.
- [85] Y.H. Yun, H.D. Li, B.C. Deng, D.S. Cao, An overview of variable selection methods in multivariate analysis of near-infrared spectra, *Trends Anal. Chem.* 113 (2019) 102–115. <https://doi.org/10.1016/j.trac.2019.01.018>.
- [86] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in Partial Least Squares Regression, *Chemometr. Intell. Lab. Syst.* 118 (2012) 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>.
- [87] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, *Anal. Chim. Acta* 667 (2010) 14–32. <https://doi.org/10.1016/j.aca.2010.03.048>.
- [88] 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. <https://doi.org/10.1016/j.aca.2014.04.019>.
- [89] 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, Quality assurance of qualitative analysis in the framework of the European project 'MEQUALAN', *Accred. Qual. Assur.* 8 (2003) 68–77. <https://doi.org/10.1007/s00769-002-0556-x>.
- [90] S.L.R. Ellison, T. Fearn, Characterising the performance of qualitative analytical methods: statistics and terminology, *Trends Anal. Chem.* 24 (2005) 468–476. <https://doi.org/10.1016/j.trac.2005.03.007>.
- [91] G.B. Costa, D.D.S. Fernandes, A.A. Gomes, V.E. Almeida, G. Veras, Using near infrared spectroscopy to classify soybean oil according to expiration date, *Food Chem.* 196 (2016) 539–543. <https://doi.org/10.1016/j.foodchem.2015.09.076>.
- [92] F. Allegrini, A.C. Olivieri, A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis, *Anal. Chim. Acta* 699 (2011) 18–25. <https://doi.org/10.1016/j.aca.2011.04.061>.
- [93] A.S. Pontes, A. Araújo, W. Marinho, P.H. Gonçalves Dias Diniz, A. de Araújo Gomes, H.C. Goicoechea, E.C. Silva, M.C.U. Araújo, Ant colony optimization for variable selection in discriminant linear analysis, *J. Chemom.* 34 (2020), e3292. <https://doi.org/10.1002/cem.3292>.
- [94] A. Biancolillo, I. Måge, T. Næs, Combining SO-PLS and linear discriminant analysis for multi-block classification, *Chemometr. Intell. Lab. Syst.* 141 (2015) 58–67. <https://doi.org/10.1016/j.chemolab.2014.12.001>.
- [95] A. Machynáková, M.P. Schneider, L. Khvalbota, O. Vyviurska, I. Špánik, A.A. Gomes, A fast and inexpensive approach to characterize Slovak Tokaj selection wines using infrared spectroscopy and chemometrics, *Food Chem.* 357 (2021), 129715. <https://doi.org/10.1016/j.foodchem.2021.129715>.
- [96] A.A. Gomes, L. Khvalbota, A. Machynáková, K. Furdíková, C.A. Zini, I. Špánik, Slovak Tokaj wines classification with respect to geographical origin by means of one class approaches, *Spectrochim. Acta Mol. Biomol. Spectrosc.* 257 (2021), 119770. <https://doi.org/10.1016/j.saa.2021.119770>.
- [97] O.Y. Rodionova, P. Oliveri, A.L. Pomerantsev, Rigorous and compliant approaches to one-class classification, *Chemometr. Intell. Lab. Syst.* 159 (2016) 89–96. <https://doi.org/10.1016/j.chemolab.2016.10.002>.
- [98] R. Ríos-Reina, S.M. Azcarate, J.M. Camiña, H.C. Goicoechea, Multi-level data fusion strategies for modeling three-way electrophoresis capillary and fluorescence arrays enhancing geographical and grape variety classification of wines, *Anal. Chim. Acta* 1126 (2020) 52–62. <https://doi.org/10.1016/j.aca.2020.06.014>.
- [99] D.A. Magdas, A. Pirnau, I. Feher, F. Guyon, B.I. Cozar, Alternative approach of applying 1H NMR in conjunction with chemometrics for wine classification, *LWT—Food Sci. Technol.* 109 (2019) 422–428. <https://doi.org/10.1016/j.lwt.2019.04.054>.
- [100] L. Gougeon, G. Da Costa, I. Le Mao, W. Ma, P.L. Teissedre, F. Guyon, T. Richard, Wine analysis and authenticity using 1H-NMR metabolomics data: application to Chinese wines, *Food Anal. Methods* 11 (2018) 3425–3434. <https://doi.org/10.1007/s12161-018-1310-2>.
- [101] J.M. Jurado, A. Alcázar, A. Palacios-Morillo, F. De Pablos, Classification of Spanish DO white wines according to their elemental profile by means of support vector machines, *Food Chem.* 135 (2012) 898–903. <https://doi.org/10.1016/j.foodchem.2012.06.017>.
- [102] C.M. de Lima, D.D.S. Fernandes, G.E. Pereira, A. de A. Gomes, M.C.U. de Araújo, P.H.G.D. Diniz, Digital image-based tracing of geographic origin, winemaker, and grape type for red wine authentication, *Food Chem.* 312 (2020). <https://doi.org/10.1016/j.foodchem.2019.126060>.
- [103] R.K.R. Ranaweera, A.M. Gilmore, D.L. Capone, S.E.P. Bastian, D.W. Jeffery, Spectrofluorometric analysis combined with machine learning for geographical and varietal authentication, and prediction of phenolic compound concentrations in red wine, *Food Chem.* 361 (2021). <https://doi.org/10.1016/j.foodchem.2021.130149>.
- [104] D.A. Magdas, F. Guyon, I. Feher, S. Pinzaru, CWine discrimination based on chemometric analysis of untargeted markers using FT-Raman spectroscopy, *Food Control* 85 (2018) 385–391. <https://doi.org/10.1016/j.foodcont.2017.10.024>.
- [105] S. Fan, Q. Zhong, C. Fahl-Hassek, M.K.H. Pfister, B. Horn, Z. Huang, Classification of Chinese wine varieties using 1H NMR spectroscopy combined with multivariate statistical analysis, *Food Control* 88 (2018) 113–122. <https://doi.org/10.1016/j.foodcont.2017.11.002>.
- [106] J. Sádecká, M. Jakubíková, P. Májek, Fluorescence spectroscopy for discrimination of botrytized wines, *Food Control* 88 (2018) 75–84. <https://doi.org/10.1016/j.foodcont.2017.12.033>.
- [107] C.A. Teixeira dos Santos, R.N.M.J. Páscoa, M.C. Sarraguça, P.A.L.S. Porto, A.L. Cerdeira, J.M. González-Sáiz, C. Pizarro, J.A. Lopes, Merging vibrational spectroscopic data for wine classification according to the geographic origin, *Food Res. Int.* 102 (2017) 504–510. <https://doi.org/10.1016/j.foodres.2017.09.018>.
- [108] C. Herrero-Latorre, J. Barciela-García, S. García-Martín, R.M. Peña-Crecente, Detection and quantification of adulterations in aged wine using RGB digital images combined with multivariate chemometric techniques, *Food Chem. X* (2019) 3. <https://doi.org/10.1016/j.focx.2019.100046>.
- [109] S.Y. Li, B.Q. Zhu, M.J. Reeves, C.Q. Duan, Phenolic analysis and theoretic design for Chinese commercial wines' authentication, *J. Food Sci.* 83 (2018) 30–38. <https://doi.org/10.1111/1750-3841.13961>.
- [110] J.L.Z. Zaukuu, J. Soós, Z. Bodor, J. Felföldi, I. Magyar, Z. Kovacs, Authentication of Tokaj wine (hungaricum) with the electronic tongue and near infrared spectroscopy, *J. Food Sci.* 84 (2019) 3437–3444. <https://doi.org/10.1111/1750-3841.14956>.
- [111] T. Nardi, M. Petrozziello, R. Girotto, M. Fugaro, R.A. Mazzei, S. Scuppa, Wine aging authentication through near infrared spectroscopy: a feasibility study on chips and barrel aged wines, *Oeno One* 54 (2020) 165–173. <https://doi.org/10.20870/oeno-one.2020.54.1.2921>.
- [112] I.K. Karabagias, V.K. Karabagias, A.V. Badeka, Volatilome of white wines as an indicator of authenticity and adulteration control using statistical analysis, *Aust. J. Grape Wine Res.* 27 (2021) 269–279. <https://doi.org/10.1111/ajgw.12486>.
- [113] R.D.B. Galán, H. Del Valle-Herrero, M. Bueno-Herrera, P. L. de la Cuesta, S. Pérez-Magariño, Volatile and non-volatile characterization of white and rosé wines from different Spanish protected designations of origin, *Beverages* 7 (2021). <https://doi.org/10.3390/beverages7030049>.
- [114] Z. Liu, H.L. Wu, L.X. Xie, Y. Hu, H. Fang, X.D. Sun, T. Wang, R. Xiao, R.Q. Yu, Direct and interference-free determination of thirteen phenolic compounds in red wines using a chemometrics-assisted HPLC-DAD strategy for authentication of vintage year, *Anal. Methods* 9 (2017) 3361–3374. <https://doi.org/10.1039/c7ay00415j>.
- [115] M. Basalekou, C. Pappas, P. Tarantilis, Y. Kotseridis, S. Kallithraka, Wine authentication with Fourier Transform Infrared Spectroscopy: a feasibility study on variety, type of barrel wood and ageing time classification, *Int. J. Food Sci. Technol.* 52 (2017) 1307–1313. <https://doi.org/10.1111/ijfs.13424>.

- [116] E.I. Geana, C.T. Ciucure, C. Apetrei, V. Artem, Application of spectroscopic UV-Vis and FT-IR screening techniques coupled with multivariate statistical analysis for red wine authentication: varietal and vintage year discrimination, *Molecules* (2019) 24. <https://doi.org/10.3390/molecules24224166>.
- [117] D.A. Magdas, B.I. Cozar, I. Feher, F. Guyon, A. Dehelean, S. Cinta Pinzaru, Testing the limits of FT-Raman spectroscopy for wine authentication: cultivar, geographical origin, vintage and terroir effect influence, *Sci. Rep.* 9 (2019). <https://doi.org/10.1038/s41598-019-56467-y>.
- [118] A.E. Springer, Wine authentication: a fingerprinting multiclass strategy to classify red varieties through profound chemometric analysis of volatiles, *Eur. Food Res. Technol.* 245 (2019) 179–190. <https://doi.org/10.1007/s00217-018-3151-z>.
- [119] I. Sen, F. Tokatli, Differentiation of wines with the use of combined data of UV-visible spectra and color characteristics, *J. Food Compos. Anal.* 45 (2016) 101–107. <https://doi.org/10.1016/j.jfca.2015.09.018>.
- [120] S.M. Azcarate, A. De Araújo Gomes, L. Vera-Candioti, M.C.U. De Araújo, J.M. Camiña, H.C. Goicoechea, Second-order capillary electrophoresis diode array detector data modeled with the tucker3 algorithm: a novel strategy for argentinean white wine discrimination respect to grape variety, *Electrophoresis* 37 (2016) 1902–1908. <https://doi.org/10.1002/elps.201600052>.
- [121] L. Lvova, I. Yaroshenko, D. Kirsanov, C.D. Natale, R. Paolesse, A. Legin, Electronic tongue for brand uniformity control: a case study of apulian red wines recognition and defects evaluation, *Sensors* 18 (2018) 2584. <https://doi.org/10.3390/s18082584>.
- [122] A. Mascellani, G. Hoca, M. Babisz, P. Kraska, P. Kloucek, J. Havlik, 1H NMR chemometric models for classification of Czech wine type and variety, *Food Chem.* 339 (2021). <https://doi.org/10.1016/j.foodchem.2020.127852>.
- [123] E.G. Alves Filho, L.M.A. Silva, P.R.V. Ribeiro, E.S. de Brito, G.J. Zocolo, P.C. Souza-Leão, A.T.B. Marques, A.L. Quintela, F.H. Larsen, K.M. Canuto, 1H NMR and LC-MS-based metabolomic approach for evaluation of the seasonality and viticultural practices in wines from São Francisco River Valley, a Brazilian semi-arid region, *Food Chem.* 289 (2019) 558–567. <https://doi.org/10.1016/j.foodchem.2019.03.103>.
- [124] A. Philippidis, E. Poulakis, R. Kontzadaki, E. Orfanakis, A. Symianaki, A. Zoumi, M. Velegarakis, Application of ultraviolet-visible absorption spectroscopy with machine learning techniques for the classification of cretan wines, *Foods* 10 (2021). <https://doi.org/10.3390/foods10010009>.
- [125] J. Tan, R. Li, Z.T. Jiang, Y. Zhang, Y.M. Hou, Y.R. Wang, X. Wu, L. Gong, Geographical classification of Chinese Cabernet Sauvignon wines by data fusion of ultraviolet-visible and synchronous fluorescence spectroscopies: the combined use of multiple wavelength differences, *Aust. J. Grape Wine Res.* 22 (2016) 358–365. <https://doi.org/10.1111/ajgw.12214>.
- [126] R.K.R. Ranaweera, A.M. Gilmore, D.L. Capone, S.E.P. Bastian, D.W. Jeffery, Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling, *Food Chem.* 335 (2021). <https://doi.org/10.1016/j.foodchem.2020.127592>.
- [127] K. Furdíková, L. Khvalbota, A. Machyňáková, I. Španík, Volatile composition and enantioselective analysis of chiral terpenoids in Tokaj varietal wines, *J. Chromatogr. B* 1167 (2021). <https://doi.org/10.1016/j.jchromb.2021.122565>.
- [128] A. Machyňáková, M.P. Schneider, L. Khvalbota, O. Vyviurska, I. Španík, A.A. Gomes, A fast and inexpensive approach to characterize Slovak Tokaj selection wines using infrared spectroscopy and chemometrics, *Food Chem.* 357 (2021). <https://doi.org/10.1016/j.foodchem.2021.129715>.
- [129] A.A. Gomes, L. Khvalbota, A. Machyňáková, K. Furdíková, C.A. Zini, I. Španík, Slovak Tokaj wines classification with respect to geographical origin by means of one class approaches, *Spectrochim. Acta, Part A* 257 (2021). <https://doi.org/10.1016/j.saa.2021.119770>.
- [130] H. Zhang, W. Shao, S. Qiu, J. Wang, Z. Wei, Collaborative analysis on the marked ages of rice wines by electronic tongue and nose based on different feature data sets, *Sensors* 20 (2020). <https://doi.org/10.3390/s20041065>.
- [131] R. Ríos-Reina, S. Elcoroaristizabal, J.A. Ocaña-González, D.L. García-González, J.M. Amigo, R.M. Callejón, Characterization and authentication of Spanish PDO wine vinegars using multidimensional fluorescence and chemometrics, *Food Chem.* 230 (2017) 108–116. <https://doi.org/10.1016/j.foodchem.2017.02.118>.
- [132] R. Ríos-Reina, R.M. Callejón, F. Savorani, J.M. Amigo, M. Cocchi, Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars, *Talanta* 198 (2019) 560–572. <https://doi.org/10.1016/j.talanta.2019.01.100>.
- [136] R. Ríos-Reina, S.M. Azcarate, J. Camiña, R.M. Callejón, J.M. Amigo, Application of hierarchical classification models and reliability estimation by bootstrapping, for authentication and discrimination of wine vinegars by UV-vis spectroscopy, *Chemometr. Intell. Lab. Syst.* 191 (2019) 42–53. <https://doi.org/10.1016/j.chemolab.2019.06.001>.
- [137] R. Ríos-Reina, D.L. García-González, R.M. Callejón, J.M. Amigo, NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin, *Food Control* 89 (2018) 108–116. <https://doi.org/10.1016/j.foodcont.2018.01.031>.
- [138] K.H. Čiča, M. Pezer, J. Mrvcic, D. Stanzer, J. Čacic, V. Jurak, M. Krajnovic, J.G. Kljusurić, Identification of phenolic and alcoholic compounds in wine spirits and their classification by use of multivariate analysis, *J. Serb. Chem. Soc.* 84 (2019) 663–677. <https://doi.org/10.2298/JSC190115020H>.
- [139] O. Anjos, M.M. Comesaña, I. Caldeira, S.I. Pedro, P.E. Oller, S. Canas, Application of functional data analysis and FTIR-ATR spectroscopy to discriminate native spirits ageing technologies, *Mathematics* 8 (2020). <https://doi.org/10.3390/MATH8060896>.
- [140] C. Medina-Plaza, J.A. de Saja, J.A. Fernández-Escudero, E. Barajas, G. Medrano, M.L. Rodríguez-Mendez, Array of biosensors for discrimination of grapes according to grape variety, vintage and ripeness, *Anal. Chim. Acta* 947 (2016) 16–22. <https://doi.org/10.1016/j.aca.2016.10.032>.
- [141] K. Włodarska, I. Khmelinskii, E. Sikorska, Authentication of apple juice categories based on multivariate analysis of the synchronous fluorescence spectra, *Food Control* 86 (2018) 42–49. <https://doi.org/10.1016/j.foodcont.2017.11.004>.
- [142] E. Coelho, J. Magalhães, F.B. Pereira, F. Macieira, L. Domingues, J.M. Oliveira, Volatile fingerprinting differentiates diverse-aged craft beers, *LWT-Food Sci. Technol.* 108 (2019) 129–136. <https://doi.org/10.1016/j.lwt.2019.03.044>.
- [143] R. Gordon, J. Chapman, A. Power, S. Chandra, J. Roberts, D. Cozzolino, Unfrazzled by fizziness: identification of beers using attenuated total reflectance mid-infrared spectroscopy and multivariate analysis, *Food Anal. Methods* 11 (2018) 2360–2367. <https://doi.org/10.1007/s12161-018-1225-y>.
- [144] L.A. da Silva, D.L. Flumignan, H.R. Pezza, L. Pezza, 1H NMR spectroscopy combined with multivariate data analysis for differentiation of Brazilian lager beer according to brewery, *Eur. Food Res. Technol.* 245 (2019) 2365–2372. <https://doi.org/10.1007/s00217-019-03354-5>.
- [145] C.G. Viejo, S. Fuentes, Low-cost methods to assess beer quality using artificial intelligence involving robotics, an electronic nose, and machine learning, *Fermentation* 6 (2020). <https://doi.org/10.3390/fermentation6040104>.
- [146] L. Mannina, F. Marini, R. Antiochia, S. Cesa, A. Magri, D. Capitani, A.P. Sobolev, Tracing the origin of beer samples by NMR and chemometrics: trappist beers as a case study, *Electrophoresis* 37 (2016) 2710–2719. <https://doi.org/10.1002/elps.201600082>.
- [147] L.A. da Silva, D.L. Flumignan, A.G. Tininis, H.R. Pezza, L. Pezza, Discrimination of Brazilian lager beer by 1H NMR spectroscopy combined with chemometrics, *Food Chem.* 272 (2019) 488–493. <https://doi.org/10.1016/j.foodchem.2018.08.077>.
- [148] V. Giannetti, M.B. Mariani, F. Marini, P. Torrelli, A. Biancolillo, Flavour fingerprint for the differentiation of Grappa from other Italian distillates by GC-MS and chemometrics, *Food Control* 105 (2019) 123–130. <https://doi.org/10.1016/j.foodcont.2019.05.028>.
- [149] M. Ocvirk, M. Nečemer, I.J. Košir, The determination of the geographic origins of hops (*Humulus lupulus* L.) by multi-elemental fingerprinting, *Food Chem.* 277 (2019) 32–37. <https://doi.org/10.1016/j.foodchem.2018.10.070>.
- [150] D. Markechová, P. Májek, J. Sádecká, Fluorescence spectroscopy and multivariate methods for the determination of brandy adulteration with mixed wine spirit, *Food Chem.* 159 (2014) 193–199. <https://doi.org/10.1016/j.foodchem.2014.02.085>.
- [151] F. Tosato, R.M. Correia, B.G. Oliveira, A.M. Fontes, H.S. França, W.K.T. Coltro, P.R. Filgueiras, W. Romão, Paper spray ionization mass spectrometry allied to chemometric tools for quantification of whisky adulteration with additions of sugarcane spirit, *Anal. Methods* 10 (2018) 1952–1960. <https://doi.org/10.1039/c8ay00071a>.
- [152] J.A.R. Teodoro, H.V. Pereira, M.M. Sena, E. Piccin, J.J. Zacca, R. Augusti, Paper spray mass spectrometry and chemometric tools for a fast and reliable identification of counterfeit blended Scottish whiskies, *Food Chem.* 237 (2017) 1058–1064. <https://doi.org/10.1016/j.foodchem.2017.06.062>.
- [153] W. Kew, I. Goodall, D. Uhrin, Analysis of Scotch Whisky by 1H NMR and chemometrics yields insight into its complex chemistry, *Food Chem.* 298 (2019). <https://doi.org/10.1016/j.foodchem.2019.125052>.
- [154] C. Roullier-Gall, J. Signoret, C. Coelho, D. Hemmler, M. Kajdan, M. Lucio, B. Schäfer, R.D. Gougeon, P. Schmitt-Kopplin, Influence of regionality and maturation time on the chemical fingerprint of whisky, *Food Chem.* 323 (2020). <https://doi.org/10.1016/j.foodchem.2020.126748>.
- [155] P. Wiśniewska, R. Boqué, E. Borrás, O. Busto, W. Wardencki, J. Namieśnik, T. Dymerski, Authentication of whisky due to its botanical origin and way of production by instrumental analysis and multivariate classification methods, *Spectrochim. Acta, Part A* 173 (2017) 849–853. <https://doi.org/10.1016/j.saa.2016.10.042>.
- [156] A.R. Martins, M. Talhavini, M.L. Vieira, J.J. Zacca, J.W.B. Braga, Discrimination of whisky brands and counterfeit identification by UV-Vis spectroscopy and multivariate data analysis, *Food Chem.* 229 (2017) 142–151. <https://doi.org/10.1016/j.foodchem.2017.02.024>.
- [157] H.H. Jelen, M. Majcher, A. Szwengiel, Key odorants in peated malt whisky and its differentiation from other whisky types using profiling of flavor and volatile compounds, *LWT-Food Sci. Technol.* 107 (2019) 56–63. <https://doi.org/10.1016/j.lwt.2019.02.070>.
- [158] M. Stupak, I. Goodall, M. Tomaniova, J. Pulkrabova, J. Hajslova, A novel approach to assess the quality and authenticity of Scotch Whisky based on gas chromatography coupled to high resolution mass spectrometry, *Anal. Chim. Acta* 1042 (2018) 60–70. <https://doi.org/10.1016/j.aca.2018.09.017>.
- [159] B.L. Smith, D.M. Hughes, A.K. Badu-Tawiah, R. Eccles, I. Goodall, S. Maher, Rapid scotch whisky analysis and authentication using desorption atmospheric pressure chemical ionisation mass spectrometry, *Sci. Rep.* 9 (2019). <https://doi.org/10.1038/s41598-019-44456-0>.
- [160] K. Sujka, P. Koczón, The application of FT-IR spectroscopy in discrimination of differently originated and aged whisky, *Eur. Food Res. Technol.* 244 (2018) 2019–2025. <https://doi.org/10.1007/s00217-018-3113-5>.

- [161] H. Hopfer, G. Gilleland, S.E. Ebeler, J. Nelson, Elemental profiles of whisk(E)y allow differentiation by type and region, *Beverages* 3 (2017). <https://doi.org/10.3390/beverages3010008>.
- [162] C. Roullier-Gall, J. Signoret, D. Hemmler, M.A. Witting, B. Kanawati, B. Schäfer, R.D. Gougeon, P. Schmitt-Kopplin, Usage of FT-ICR-MS metabolomics for characterizing the chemical signatures of barrel-aged whisky, *Front. Chem.* 6 (2018). <https://doi.org/10.3389/fchem.2018.00029>.
- [163] J. Kiefer, A.L. Cromwell, Analysis of single malt Scotch whisky using Raman spectroscopy, *Anal. Methods* 9 (2017) 511–518. <https://doi.org/10.1039/c6ay02907h>.
- [164] G.C. Silvello, A.R. Alcarde, Experimental design and chemometric techniques applied in electronic nose analysis of wood-aged sugar cane spirit (cachaça), *J. Agric. Food Res.* (2020) 2. <https://doi.org/10.1016/j.jafr.2020.100037>.
- [165] G. Pérez-Caballero, J.M. Andrade, P. Olmos, Y. Molina, I. Jiménez, J.J. Durán, C. Fernandez-Lozano, F. Miguel-Cruz, Authentication of tequilas using pattern recognition and supervised classification, *TrAC, Trends Anal. Chem.* 94 (2017) 117–129. <https://doi.org/10.1016/j.trac.2017.07.008>.
- [166] J.M. Andrade, D. Ballabio, M.P. Gómez-Carracedo, G. Pérez-Caballero, Nonlinear classification of commercial Mexican tequilas, *J. Chemom.* 31 (2017). <https://doi.org/10.1002/cem.2939>.
- [167] A. Carreon-Alvarez, A. Suárez-Gómez, F. Zurita, S. Gómez-Salazar, J.F.A. Soltero, M. Barceña-Soto, N. Casillas, Porfirio-Gutiérrez, E.D. Moreno-Medrano, Assessment of physicochemical properties of tequila brands: authentication and quality, *J. Chem.* (2016). <https://doi.org/10.1155/2016/6254942>.
- [168] J.R. Belmonte-Sánchez, R. Romero-González, J.L. Martínez Vidal, F.J. Arrebola, A. Garrido Frenich, 1H NMR and multi-technique data fusion as metabolomic tool for the classification of golden rums by multivariate statistical analysis, *Food Chem.* 317 (2020). <https://doi.org/10.1016/j.foodchem.2020.126363>.
- [169] J.R. Belmonte-Sánchez, R. Romero-González, F.J. Arrebola, J.L.M. Vidal, A. Garrido Frenich, An innovative metabolomic approach for golden rum classification combining ultrahigh-performance liquid chromatography-orbitrap mass spectrometry and chemometric strategies, *J. Agric. Food Chem.* 67 (2019) 1302–1311. <https://doi.org/10.1021/acs.jafc.8b05622>.
- [170] N. Govindaraj, S. Gangadoo, V.K. Truong, J. Chapman, H. Gill, D. Cozzolino, The use of derivatives and chemometrics to interrogate the UV–Visible spectra of gin samples to monitor changes related to storage, *Spectrochim. Acta, Part A* 227 (2020). <https://doi.org/10.1016/j.saa.2019.117548>.
- [171] Z. Li, K.S. Suslick, A hand-held optoelectronic nose for the identification of liquors, *ACS Sens.* 3 (2018) 121–127. <https://doi.org/10.1021/acssens.7b00709>.
- [172] M. Śliwińska, P. Wiśniewska, T. Dymerski, W. Wardencki, J. Namieśnik, Application of electronic nose based on fast GC for authenticity assessment of polish homemade liqueurs called nalewka, *Food Anal. Methods* 9 (2016) 2670–2681. <https://doi.org/10.1007/s12161-016-0448-z>.
- [173] M.L. Xu, Y. Yu, H.S. Ramaswamy, S.M. Zhu, Characterization of Chinese liquor aroma components during aging process and liquor age discrimination using gas chromatography combined with multivariable statistics, *Sci. Rep.* 7 (2017). <https://doi.org/10.1038/srep39671>.
- [174] R.L. Burns, R. Alexander, L. Snaychuk, J.C. Edwards, N. Fitzgerald, P. Gao, D. Quan, C. Douvris, T. Vaughan, D.D. Bussan, A fast, straightforward and inexpensive method for the authentication of baijiu spirit samples by fluorescence spectroscopy, *Beverages* 7 (2021). <https://doi.org/10.3390/beverages7030065>.
- [175] V. Giannetti, M.B. Mariani, F. Marini, P. Torrelli, A. Biancolillo, Grappa and Italian spirits: multi-platform investigation based on GC–MS, MIR and NIR spectroscopies for the authentication of the Geographical Indication, *Microchem. J.* (2020) 157. <https://doi.org/10.1016/j.microc.2020.104896>.
- [176] V. Giannetti, M. Boccacci Mariani, P. Torrelli, F. Marini, Flavour component analysis by HS-SPME/GC–MS and chemometric modeling to characterize Pilsner-style Lager craft beers, *Microchem. J.* 149 (2019). <https://doi.org/10.1016/j.microc.2019.103991>.
- [177] X. Li, M. Tsuta, F. Tanaka, M. Tsukahara, K. Tsukahara, Assessment of Japanese awamori spirits using UV–VIS spectroscopy, *Food Anal. Methods* 13 (2019) 726–734. <https://doi.org/10.1007/s12161-019-01692-4> / [Published](https://doi.org/10.1007/s12161-019-01692-4).
- [178] D.I. Ellis, H. Muhamadali, Y. Xu, R. Eccles, I. Goodall, R. Goodacre, Rapid through-container detection of fake spirits and methanol quantification with handheld Raman spectroscopy, *Analyst* 144 (2019) 324–330. <https://doi.org/10.1039/c8an01702f>.
- [179] M. Śliwińska, P. Wiśniewska, T. Dymerski, W. Wardencki, J. Namieśnik, Evaluation of the suitability of electronic nose based on fast GC for distinguishing between the plum spirits of different geographical origins, *Eur. Food Res. Technol.* 242 (2016) 1813–1819. <https://doi.org/10.1007/s00217-016-2680-6>.
- [180] G. Binetti, L. Del Coco, R. Ragone, S. Zelasco, E. Perri, C. Montemurro, R. Valentini, D. Naso, F.P. Fanizzi, F.P. Schena, Cultivar classification of Apulian olive oils: use of artificial neural networks for comparing NMR, NIR and merceological data, *Food Chem.* 219 (2017) 131–138. <https://doi.org/10.1016/j.foodchem.2016.09.041>.
- [181] N.P. Kalogiouri, N.A. Alygizakis, R. Aalizadeh, N.S. Thomaidis, Olive oil authenticity studies by target and nontarget LC–QTOF-MS combined with advanced chemometric techniques, *Anal. Bioanal. Chem.* 408 (2016) 7955–7970. <https://doi.org/10.1007/s00216-016-9891-3>.
- [182] K. Georgouli, J. Martinez Del Rincon, A. Koidis, Continuous statistical modelling for rapid detection of adulteration of extra virgin olive oil using mid infrared and Raman spectroscopic data, *Food Chem.* 217 (2017) 735–742. <https://doi.org/10.1016/j.foodchem.2016.09.011>.
- [183] T. Mu, S. Chen, Y. Zhang, H. Chen, P. Guo, F. Meng, Portable detection and quantification of olive oil adulteration by 473-nm laser-induced fluorescence, *Food Anal. Methods* 9 (2016) 275–279. <https://doi.org/10.1007/s12161-015-0199-2>.
- [184] A. Bajoub, S. Medina-Rodríguez, M. Gómez-Romero, E.A. Ajal, M.G. Bagur-González, A. Fernández-Gutiérrez, A. Carrasco-Pancorbo, Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics, *Food Chem.* 215 (2017) 245–255. <https://doi.org/10.1016/j.foodchem.2016.07.140>.
- [185] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Acuña, A. Calvo, O. Busto, Olive oil sensory defects classification with data fusion of instrumental techniques and multivariate analysis (PLS-DA), *Food Chem.* 203 (2016) 314–322. <https://doi.org/10.1016/j.foodchem.2016.02.038>.
- [186] S. Schwolow, N. Gerhardt, S. Rohn, P. Weller, Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Anal. Bioanal. Chem.* 411 (2019) 6005–6019. <https://doi.org/10.1007/s00216-019-01978-w>.
- [187] G. Ozcan-Sinir, Detection of adulteration in extra virgin olive oil by selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics, *Food Control* 118 (2020). <https://doi.org/10.1016/j.foodcont.2020.107433>.
- [188] A. Rohman, Y. bin Che Man, A. Ismail, P. Hashim, FTIR spectroscopy coupled with chemometrics of multivariate calibration and discriminant analysis for authentication of extra virgin olive oil, *Int. J. Food Prop.* 20 (2017) S1173–S1181. <https://doi.org/10.1080/10942912.2017.1336718>.
- [189] N. Jurado-Campos, N. Arroyo-Manzaneres, P. Viñas, L. Arce, Quality authentication of virgin olive oils using orthogonal techniques and chemometrics based on individual and high-level data fusion information, *Talanta* 219 (2020). <https://doi.org/10.1016/j.talanta.2020.121260>.
- [190] C. Pizarro, S. Rodríguez-Tecedor, N. Pérez-Del-Notario, I. Esteban-Díez, J.M. González-Sáiz, Classification of Spanish extra virgin olive oils by data fusion of visible spectroscopic fingerprints and chemical descriptors, *Food Chem.* 138 (2013) 915–922. <https://doi.org/10.1016/j.foodchem.2012.11.087>.
- [191] T.K. de Lima, M. Musso, D. Bertoldo Menezes, Using Raman spectroscopy and an exponential equation approach to detect adulteration of olive oil with rapeseed and corn oil, *Food Chem.* 333 (2020). <https://doi.org/10.1016/j.foodchem.2020.127454>.
- [192] Y. Li, S. Chen, H. Chen, P. Guo, T. Li, Q. Xu, Effect of thermal oxidation on detection of adulteration at low concentrations in extra virgin olive oil: study based on laser-induced fluorescence spectroscopy combined with KPCA–LDA, *Food Chem.* 309 (2020). <https://doi.org/10.1016/j.foodchem.2019.125669>.
- [193] F.T. Borghi, P.C. Santos, F.D. Santos, M.H.C. Nascimento, T. Corrêa, M. Cesconetto, A.A. Pires, A.V.F.N. Ribeiro, V. Lacerda, W. Romão, P.R. Figueiras, Quantification and classification of vegetable oils in extra virgin olive oil samples using a portable near-infrared spectrometer associated with chemometrics, *Microchem. J.* 159 (2020). <https://doi.org/10.1016/j.microc.2020.105544>.
- [194] A.M. Jiménez-Carvelo, E. Pérez-Castaño, A. González-Casado, L. Cuadros-Rodríguez, One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction, *Food Chem.* 221 (2017) 1784–1791. <https://doi.org/10.1016/j.foodchem.2016.10.103>.
- [195] K.A. Obisesan, A.M. Jiménez-Carvelo, L. Cuadros-Rodríguez, I. Ruisánchez, M.P. Callao, HPLC-UV and HPLC-CAD chromatographic data fusion for the authentication of the geographical origin of palm oil, *Talanta* 170 (2017) 413–418. <https://doi.org/10.1016/j.talanta.2017.04.035>.
- [196] C. Berghian-Grosan, D.A. Magdas, Raman spectroscopy and machine-learning for edible oils evaluation, *Talanta* 218 (2020). <https://doi.org/10.1016/j.talanta.2020.121176>.
- [197] Y.Y. Yuan, S.T. Wang, J.Z. Wang, Q. Cheng, X.J. Wu, D.M. Kong, Rapid detection of the authenticity and adulteration of sesame oil using excitation-emission matrix fluorescence and chemometric methods, *Food Control* 112 (2020). <https://doi.org/10.1016/j.foodcont.2020.107145>.
- [198] C. Berghian-Grosan, D.A. Magdas, Raman spectroscopy and machine-learning for edible oils evaluation, *Talanta* 218 (2020). <https://doi.org/10.1016/j.talanta.2020.121176>.
- [199] M. Esteki, B. Farajmand, S. Amanifar, R. Barkhordari, Z. Ahadiyan, E. Dashtaki, M. Mohammadlou, Y. Vander Heyden, Classification and authentication of Iranian walnuts according to their geographical origin based on gas chromatographic fatty acid fingerprint analysis using pattern recognition methods, *Chemometr. Intell. Lab. Syst.* 171 (2017) 251–258. <https://doi.org/10.1016/j.chemolab.2017.10.014>.
- [200] E. Bona, I. Marquetti, J.V. Link, G.Y.F. Makimori, V. da Costa Arca, A.L. Guimarães Lemes, J.M.G. Ferreira, M.B. dos Santos Scholz, P. Valderrama, R.J. Poppi, Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee, *LWT—Food Sci. Technol.* 76 (2017) 330–336. <https://doi.org/10.1016/j.lwt.2016.04.048>.
- [201] M.I. Milani, E.L. Rossini, T.A. Catelani, L. Pezza, A.T. Toci, H.R. Pezza, Authentication of roasted and ground coffee samples containing multiple adulterants using NMR and a chemometric approach, *Food Control* 112 (2020). <https://doi.org/10.1016/j.foodcont.2020.107104>.

- [202] L. Mandrile, L. Barbosa-Pereira, K.M. Sorensen, A.M. Giovannozzi, G. Zeppa, S.B. Engelsens, A.M. Rossi, Authentication of cocoa bean shells by near- and mid-infrared spectroscopy and inductively coupled plasma-optical emission spectroscopy, *Food Chem.* 292 (2019) 47–57. <https://doi.org/10.1016/j.foodchem.2019.04.008>.
- [203] L. Barbosa-Pereira, O. Rojo-Poveda, I. Ferrocino, M. Giordano, G. Zeppa, Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the classification of cocoa bean shells using chemometrics, *Food Res. Int.* 123 (2019) 684–696. <https://doi.org/10.1016/j.foodres.2019.05.041>.
- [204] N. Cain, O. Alka, T. Segelke, K. von Wuthenau, O. Kohlbacher, M. Fischer, Food fingerprinting: mass spectrometric determination of the cocoa shell content (Theobroma cacao L.) in cocoa products by HPLC-QTOF-MS, *Food Chem.* 298 (2019). <https://doi.org/10.1016/j.foodchem.2019.125013>.
- [205] R. Vanderschueren, D. Montalvo, B. De Ketelaere, J.A. Delcour, E. Smolders, The elemental composition of chocolates is related to cacao content and origin: a multi-element fingerprinting analysis of single origin chocolates, *J. Food Compos. Anal.* 83 (2019). <https://doi.org/10.1016/j.jfca.2019.103277>.
- [206] S. Fernandes, A. Gois, F. Mendes, R. Perestrelo, S. Medina, J.S. Câmara, Typicality assessment of onions (*Allium cepa*) from different geographical regions based on the volatile signature and chemometric tools, *Foods* 9 (2020). <https://doi.org/10.3390/foods9030375>.
- [207] A. Mir-Marqués, C. Elvira-Sáez, M.L. Cervera, S. Garrigues, M. de la Guardia, Authentication of protected designation of origin artichokes by spectroscopy methods, *Food Control* 59 (2016) 74–81. <https://doi.org/10.1016/j.foodcont.2015.05.004>.
- [208] B. Richter, S. Gurk, D. Wagner, M. Bockmayr, M. Fischer, Food authentication: multi-elemental analysis of white asparagus for provenance discrimination, *Food Chem.* 286 (2019) 475–482. <https://doi.org/10.1016/j.foodchem.2019.01.105>.
- [209] S. Amiry, M. Esmailli, M. Alizadeh, Classification of adulterated honeys by multivariate analysis, *Food Chem.* 224 (2017) 390–397. <https://doi.org/10.1016/j.foodchem.2016.12.025>.
- [210] M. Spiteri, E. Dubin, J. Cotton, M. Poirer, B. Corman, E. Jamin, M. Lees, D. Rutledge, Data fusion between high resolution 1H-NMR and mass spectrometry: a synergetic approach to honey botanical origin characterization, *Anal. Bioanal. Chem.* 408 (2016) 4389–4401. <https://doi.org/10.1007/s00216-016-9538-4>.
- [211] S. Schwolow, N. Gerhardt, S. Rohn, P. Weller, Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Anal. Bioanal. Chem.* 411 (2019) 6005–6019. <https://doi.org/10.1007/s00216-019-01978-w>.
- [212] M.S. Azevedo, S.K.T. Seraglio, G. Rocha, C.B. Balderas, M. Piovezan, L.V. Gonzaga, D. de B. Falkenberg, R. Fett, M.A.L. de Oliveira, A.C.O. Costa, Free amino acid determination by GC-MS combined with a chemometric approach for geographical classification of bracing honeydew honey (*Mimosa scabrella* Benth), *Food Control* 78 (2017) 383–392. <https://doi.org/10.1016/j.foodcont.2017.03.008>.
- [213] D. Ballabio, E. Robotti, F. Grisoni, F. Quasso, M. Bobba, S. Vercelli, F. Gosetti, G. Calabrese, E. Sangiorgi, M. Orlandi, E. Marengo, Chemical profiling and multivariate data fusion methods for the identification of the botanical origin of honey, *Food Chem.* 266 (2018) 79–89. <https://doi.org/10.1016/j.foodchem.2018.05.084>.
- [214] Y. Riswahyuli, A. Rohman, F.M.C.S. Setyabudi, S. Raharjo, Indonesian wild honey authenticity analysis using attenuated total reflectance-fourier transform infrared (ATR-FTIR) spectroscopy combined with multivariate statistical techniques, *Heliyon* 6 (2020). <https://doi.org/10.1016/j.heliyon.2020.e03662>.
- [215] Q. Li, J. Zeng, L. Lin, J. Zhang, J. Zhu, L. Yao, S. Wang, Z. Yao, Z. Wu, Low risk of category misdiagnosis of rice syrup adulteration in three botanical origin honey by ATR-FTIR and general model, *Food Chem.* 332 (2020). <https://doi.org/10.1016/j.foodchem.2020.127356>.
- [216] C. Guellis, D.C. Valério, G.G. Bessegato, M. Boroski, J.C. Dragunski, C.A. Lindino, Non-targeted method to detect honey adulteration: combination of electrochemical and spectrophotometric responses with principal component analysis, *J. Food Compos. Anal.* 89 (2020). <https://doi.org/10.1016/j.jfca.2020.103466>.
- [217] D.P. Aykas, M.-L. Shotts, L.E. Rodriguez-Saona, Authentication of commercial honeys based on Raman fingerprinting and pattern recognition analysis, *Food Control* 117 (2020), 107346. <https://doi.org/10.1016/j.foodcont.2020.107346>.
- [218] C. He, Y. Liu, H. Liu, X. Zheng, G. Shen, J. Feng, Compositional identification and authentication of Chinese honeys by 1H NMR combined with multivariate analysis, *Food Res. Int.* 130 (2020). <https://doi.org/10.1016/j.foodres.2019.108936>.
- [219] H. Chen, C. Tan, Z. Lin, T. Wu, Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares, *Spectrochim. Acta, Part A* 173 (2017) 832–836. <https://doi.org/10.1016/j.saa.2016.10.051>.
- [220] C. de S. Gondim, R.G. Junqueira, S. V. C. de Souza, I. Ruisánchez, M.P. Callao, Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies, *Food Chem.* 230 (2017) 68–75. <https://doi.org/10.1016/j.foodchem.2017.03.022>.
- [221] C. de S. Gondim, R.G. Junqueira, S.V.C. de Souza, M.P. Callao, I. Ruisánchez, Determining performance parameters in qualitative multivariate methods using probability of detection (POD) curves. Case study: two common milk adulterants, *Talanta* 168 (2017) 23–30. <https://doi.org/10.1016/j.talanta.2016.12.065>.
- [222] P.M. Santos, E.R. Pereira-Filho, L.A. Colnago, Detection and quantification of milk adulteration using time domain nuclear magnetic resonance (TD-NMR), *Microchem. J.* 124 (2016) 15–19. <https://doi.org/10.1016/j.microc.2015.07.013>.
- [223] J. Griboff, M.V. Baroni, M. Horacek, D.A. Wunderlin, M.V. Monferran, Multielemental + isotopic fingerprint enables linking soil, water, forage and milk composition, assessing the geographical origin of Argentinean milk, *Food Chem.* 283 (2019) 549–558. <https://doi.org/10.1016/j.foodchem.2019.01.067>.
- [224] J.L. da P. Teixeira, E.T. dos S. Caramês, D.P. Baptista, M.L. Gigante, J.A.L. Pallone, Vibrational spectroscopy and chemometrics tools for authenticity and improvement the safety control in goat milk, *Food Control* 112 (2020). <https://doi.org/10.1016/j.foodcont.2020.107105>.
- [225] H. Liu, Q. Zhao, X. Guo, C. Tang, X. Yu, T. Zhan, Y. Qin, J. Zhang, Application of isotopic and elemental fingerprints in identifying the geographical origin of goat milk in China, *Food Chem.* 277 (2019) 448–454. <https://doi.org/10.1016/j.foodchem.2018.10.144>.
- [226] G. Riuzzi, A. Tata, A. Massaro, V. Bisutti, I. Lanza, B. Contiero, M. Bragolusi, B. Miano, A. Negro, F. Gottardo, R. Piro, S. Segato, Authentication of forage-based milk by mid-level data fusion of (+/-) DART-HRMS signatures, *Int. Dairy J.* 112 (2021). <https://doi.org/10.1016/j.idairyj.2020.104859>.
- [227] S.H. Gregorić, D. Potočnik, F. Camin, N. Ogrinc, Milk authentication: stable isotope composition of hydrogen and oxygen in milks and their constituents, *Molecules* (2020) 25. <https://doi.org/10.3390/molecules25174000>.
- [228] N. Liu, H.A. Parra, A. Pustjens, K. Hettinga, P. Mongondry, S.M. van Ruth, Evaluation of portable near-infrared spectroscopy for organic milk authentication, *Talanta* 184 (2018) 128–135. <https://doi.org/10.1016/j.talanta.2018.02.097>.
- [229] Y.T. Wang, H. Bin Ren, W.Y. Liang, X. Jin, Q. Yuan, Z.R. Liu, D.M. Chen, Y.H. Zhang, A novel approach to temperature-dependent thermal processing authentication for milk by infrared spectroscopy coupled with machine learning, *J. Food Eng.* 311 (2021). <https://doi.org/10.1016/j.jfoodeng.2021.110740>.
- [230] I.M. Chung, J.K. Kim, C.T. Yarnes, Y.J. An, C. Kwon, S.Y. Kim, Y.J. Yang, H.Y. Chi, S.H. Kim, Fatty acid- and amino acid-specific isotope analysis for accurate authentication and traceability in organic milk, *J. Agric. Food Chem.* 67 (2019) 711–722. <https://doi.org/10.1021/acs.jafc.8b05063>.
- [231] E.V. dos Santos Pereira, D.D. de Sousa Fernandes, M.C.U. de Araújo, P.H.G.D. Diniz, M.I.S. Maciel, In-situ authentication of goat milk in terms of its adulteration with cow milk using a low-cost portable NIR spectrophotometer, *Microchem. J.* 163 (2021). <https://doi.org/10.1016/j.microc.2020.105885>.
- [232] M. Bergamaschi, C. Cipolat-Gotet, A. Cecchinato, S. Schiavon, G. Bittante, Chemometric authentication of farming systems of origin of food (milk and ripened cheese) using infrared spectra, fatty acid profiles, flavor fingerprints, and sensory descriptions, *Food Chem.* 305 (2020). <https://doi.org/10.1016/j.foodchem.2019.125480>.
- [233] M. Bergamaschi, C. Cipolat-Gotet, A. Cecchinato, S. Schiavon, G. Bittante, Chemometric authentication of farming systems of origin of food (milk and ripened cheese) using infrared spectra, fatty acid profiles, flavor fingerprints, and sensory descriptions, *Food Chem.* 305 (2020). <https://doi.org/10.1016/j.foodchem.2019.125480>.
- [234] S.M. Zain, S. Behkani, S. Bakirdere, I.B. Koki, Milk authentication and discrimination via metal content clustering - a case of comparing milk from Malaysia and selected countries of the world, *Food Control* 66 (2016) 306–314. <https://doi.org/10.1016/j.foodcont.2016.02.015>.
- [235] S.R. Karunathilaka, B.J. Yakes, K. He, J.K. Chung, M. Mossoba, Non-targeted NIR spectroscopy and SIMCA classification for commercial milk powder authentication: a study using eleven potential adulterants, *Heliyon* 4 (2018) 806. <https://doi.org/10.1016/j.heliyon.2018>.
- [236] O. Taylan, N. Cebi, M. Tahsin Yilmaz, O. Sagdic, A.A. Bakhsh, Detection of lard in butter using Raman spectroscopy combined with chemometrics, *Food Chem.* 332 (2020). <https://doi.org/10.1016/j.foodchem.2020.127344>.
- [237] M. Bergamaschi, C. Cipolat-Gotet, A. Cecchinato, S. Schiavon, G. Bittante, Chemometric authentication of farming systems of origin of food (milk and ripened cheese) using infrared spectra, fatty acid profiles, flavor fingerprints, and sensory descriptions, *Food Chem.* 305 (2020). <https://doi.org/10.1016/j.foodchem.2019.125480>.
- [238] D.A. Magdas, I. Feher, G. Cristea, C. Voica, A. Tabaran, M. Mihaiu, D.V. Cordea, V.A. Bălăneanu, S.D. Dan, Geographical origin and species differentiation of Transylvanian cheese. Comparative study of isotopic and elemental profiling vs. DNA results, *Food Chem.* 277 (2019) 307–313. <https://doi.org/10.1016/j.foodchem.2018.10.103>.
- [239] L.G. Visconti, M.S. Rodríguez, C.V. Di Anibal, Determination of grated hard cheeses adulteration by near infrared spectroscopy (NIR) and multivariate analysis, *Int. Dairy J.* 104 (2020). <https://doi.org/10.1016/j.idairyj.2020.104647>.
- [240] R. Calvini, S. Michelini, V. Pizzamiglio, G. Foca, A. Ulrici, Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano cheese, *Food Control* 112 (2020). <https://doi.org/10.1016/j.foodcont.2020.107111>.
- [241] L. Bontempo, A. Barbero, D. Bertoldi, F. Camin, R. Larcher, M. Perini, A. Sepulcri, L. Zicarelli, E. Piasentier, Isotopic and elemental profiles of Mediterranean buffalo milk and cheese and authentication of Mozzarella di

- Bufala Campana PDO: an initial exploratory study, *Food Chem.* 285 (2019) 316–323. <https://doi.org/10.1016/j.foodchem.2019.01.160>.
- [242] C. Márquez, M.I. López, I. Ruisánchez, M.P. Callao, FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud, *Talanta* 161 (2016) 80–86. <https://doi.org/10.1016/j.talanta.2016.08.003>.
- [243] K.M. Nunes, M.V.O. Andrade, A.M.P. Santos Filho, M.C. Lasmar, M.M. Sena, Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy, *Food Chem.* 205 (2016) 14–22. <https://doi.org/10.1016/j.foodchem.2016.02.158>.
- [244] W. Sun, X. Zhang, Z. Zhang, R. Zhu, Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb, *Spectrochim. Acta, Part A* 171 (2017) 72–79. <https://doi.org/10.1016/j.saa.2016.07.039>.
- [245] Y. Li, Y. Wang, Synergistic strategy for the geographical traceability of wild Boletus tomentipes by means of data fusion analysis, *Microchem. J.* 140 (2018) 38–46. <https://doi.org/10.1016/j.microc.2018.04.001>.
- [246] S. Yao, J. Li, T. Li, Z. Duan, Y. Wang, Geographical traceability of Boletaceae mushrooms using data fusion of FT-IR, UV, and ICP-AES combined with SVM, *Int. J. Food Prop.* 22 (2019) 414–426. <https://doi.org/10.1080/10942912.2019.1588299>.
- [247] A. Dankowska, W. Kowalewski, Tea types classification with data fusion of UV–Vis, synchronous fluorescence and NIR spectroscopies and chemometric analysis, *Spectrochim. Acta, Part A* 211 (2019) 195–202. <https://doi.org/10.1016/j.saa.2018.11.063>.
- [248] Z. Małjurek, B. Zawisza, D. de Beer, E. Joubert, B. Walczak, Authentication of honeybush and rooibos herbal teas based on their elemental composition, *Food Control* 123 (2021). <https://doi.org/10.1016/j.foodcont.2020.107757>.
- [249] J. Zhu, F. Zhu, L. Li, L. Cheng, L. Zhang, Y. Sun, X. Wan, Z. Zhang, Highly discriminant rate of Dianhong black tea grades based on fluorescent probes combined with chemometric methods, *Food Chem.* 298 (2019). <https://doi.org/10.1016/j.foodchem.2019.125046>.
- [250] M. Xu, J. Wang, L. Zhu, The qualitative and quantitative assessment of tea quality based on E-nose, E-tongue and E-eye combined with chemometrics, *Food Chem.* 289 (2019) 482–489. <https://doi.org/10.1016/j.foodchem.2019.03.080>.
- [251] Q.Q. Wang, H.Y. Huang, Y.Z. Wang, Geographical authentication of macrohyporia cocos by a data fusion method combining ultra-fast liquid chromatography and fourier transform infrared spectroscopy, *Molecules* (2019) 24. <https://doi.org/10.3390/molecules24071320>.
- [252] P.Q. Thaiyz, Q.T.L. Hoangzy, T.T. Tuan, Bottled water classification using spectroscopy and deep belief network, *ISEE* (2021), 20593592. <https://doi.org/10.1109/ISEE51682.2021.9418702>.
- [253] H. Liu, Y. Wei, H. Lu, S. Wei, T. Jiang, Y. Zhang, B. Guo, Combination of the  $87\text{Sr}/86\text{Sr}$  ratio and light stable isotopic values ( $\delta^{13}\text{C}$ ,  $\delta^{15}\text{N}$  and  $\delta\text{d}$ ) for identifying the geographical origin of winter wheat in China, *Food Chem.* 212 (2016) 367–373. <https://doi.org/10.1016/j.foodchem.2016.06.002>.
- [254] K. Porker, M. Zerner, D. Cozzolino, Classification and authentication of barley (hordeum vulgare) malt varieties: combining attenuated total reflectance mid-infrared spectroscopy with chemometrics, *Food Anal. Methods* (2016). <https://doi.org/10.1007/s12161-016-0627-y>.
- [255] C.H.V. Bazoni, E.I. Ida, D.F. Barbin, L.E. Kurozawa, Near-infrared spectroscopy as a rapid method for evaluation physicochemical changes of stored soybeans, *J. Stored Prod. Res.* 73 (2017) 1–6. <https://doi.org/10.1016/j.jspr.2017.05.003>.
- [256] E. Achten, D. Schütz, M. Fischer, C. Faulh-Hassek, J. Riedl, B. Horn, Classification of grain maize (Zea mays L.) from different geographical origins with FTIR spectroscopy—a suitable analytical tool for feed authentication? *Food Anal. Methods* 12 (2019) 2172–2184. <https://doi.org/10.1007/s12161-019-01558-9>.
- [257] Z. Liu, Y. Yuan, T. Xie, Y. Zhang, S. Shao, J. Nie, W. Xia, K.M. Rogers, W. Zhang, Long-term agricultural effects on the authentication accuracy of organic, green, and conventional rice using isotopic and elemental chemometric analyses, *J. Agric. Food Chem.* 68 (2020) 1213–1225. <https://doi.org/10.1021/acs.jafc.9b06847>.
- [258] (a) I.M. Chung, J.K. Kim, K.J. Lee, S.K. Park, J.H. Lee, N.Y. Son, Y.I. Jin, S.H. Kim, Geographic authentication of Asian rice (*Oryza sativa* L.) using multi-elemental and stable isotopic data combined with multivariate analysis, *Food Chem.* 240 (2018) 840–849. <https://doi.org/10.1016/j.foodchem.2017.08.023>;  
(b) J. Wang, T. Chen, W. Zhang, Y. Zhao, S. Yang, A. Chen, Tracing the geographical origin of rice by stable isotopic analyses combined with chemometrics, *Food Chem.* 313 (2020). <https://doi.org/10.1016/j.foodchem.2019.126093>.
- [259] S. Hurtado-Bermúdez, J.A. Jurado-González, J.L. Santos, C.F. Díaz-Amigo, I. Aparicio, J.L. Más, E. Alonso, Geographical origin of bivalve molluscs in coastal areas using natural radioactivity fingerprinting and multivariate statistical analyses: andalusian coast as case of study, *J. Hazard Mater.* 367 (2019) 706–714. <https://doi.org/10.1016/j.jhazmat.2019.01.027>.
- [260] X. Zhang, D. Han, X. Chen, X. Zhao, J. Cheng, Y. Liu, Combined use of fatty acid profile and fatty acid  $\delta^{13}\text{C}$  fingerprinting for origin traceability of scallops (*Patinopecten yessoensis*, *Chlamys farreri*, and *Argopecten irradians*), *Food Chem.* 298 (2019). <https://doi.org/10.1016/j.foodchem.2019.124966>.
- [261] X. Zhang, J. Cheng, D. Han, X. Zhao, X. Chen, Y. Liu, Geographical origin traceability and species identification of three scallops (*Patinopecten yessoensis*, *Chlamys farreri*, and *Argopecten irradians*) using stable isotope analysis, *Food Chem.* 299 (2019). <https://doi.org/10.1016/j.foodchem.2019.125107>.
- [262] X. Zhang, J. Cheng, D. Han, X. Chen, X. Zhao, Y. Liu, Regional differences in fatty acid composition of sea cucumber (*Apostichopus japonicus*) and scallop (*Patinopecten yessoensis*) in the coastal areas of China, *Reg. Stud. Mar. Sci.* 31 (2019), 100782. <https://doi.org/10.1016/j.rsma.2019.100782>.
- [267] L. Morrison, M. Bennion, S. Gill, C.T. Graham, Spatio-temporal trace element fingerprinting of king scallops (*Pecten maximus*) reveals harvesting period and location, *Sci. Total Environ.* (2019) 697. <https://doi.org/10.1016/j.scitotenv.2019.134121>.
- [268] X. Zhang, Y. Liu, Y. Li, X. Zhao, Identification of the geographical origins of sea cucumber (*Apostichopus japonicus*) in northern China by using stable isotope ratios and fatty acid profiles, *Food Chem.* 218 (2017) 269–276. <https://doi.org/10.1016/j.foodchem.2016.08.083>.
- [269] X. Kang, Y. Zhao, D. Shang, Y. Zhai, J. Ning, H. Ding, X. Sheng, Identification of the geographical origins of sea cucumbers in China: the application of stable isotope ratios and compositions of C, N, O and H, *Food Control* 11 (2020), 107036. <https://doi.org/10.1016/j.foodcont.2019.107036>.
- [270] X. Zhang, J. Cheng, D. Han, X. Chen, X. Zhao, Y. Liu, Regional differences in fatty acid composition of sea cucumber (*Apostichopus japonicus*) and scallop (*Patinopecten yessoensis*) in the coastal areas of China, *Reg. Stud. Mar. Sci.* 31 (2019), 100782. <https://doi.org/10.1016/j.rsma.2019.100782>.
- [271] K. Gopi, D. Mazumder, J. Sammut, N. Saintilan, J. Crawford, P. Gadd, Isotopic and elemental profiling to trace the geographic origins of farmed and wild-caught Asian seabass (*Lates calcarifer*), *Aquaculture* 502 (2019) 56–62. <https://doi.org/10.1016/j.aquaculture.2018.12.012>.
- [272] C. Wang, W.Z. He, D.Y. Zhao, Z. Liu, Y. ying Fan, W.N. Tian, W.L. Wu, K.M. Rogers, Modeling of stable isotope and multi-element compositions of jujube (*Ziziphus jujuba* Mill.) for origin traceability of protected geographical indication (PGI) products in Xinjiang, China, *J. Food Compos. Anal.* 92 (2020). <https://doi.org/10.1016/j.jfca.2020.103577>.
- [273] S. Munera, J.M. Amigo, N. Aleixos, P. Talens, S. Cubero, J. Blasco, Potential of VIS-NIR hyperspectral imaging and chemometric methods to identify similar cultivars of nectarine, *Food Control* 86 (2018) 1–10. <https://doi.org/10.1016/j.foodcont.2017.10.037>.
- [274] M.J. Hidalgo, D.C. Fechner, E.J. Marchevsky, R.G. Pellerano, Determining the geographical origin of Sechium edule fruits by multielement analysis and advanced chemometric techniques, *Food Chem.* 210 (2016) 228–234. <https://doi.org/10.1016/j.foodchem.2016.04.120>.
- [275] I. Coelho, A.S. Matos, R. Teixeira, A. Nascimento, J. Bordado, O. Donard, I. Castanheira, Combining multielement analysis and chemometrics to trace the geographical origin of Rocha pear, *J. Food Compos. Anal.* 77 (2019) 1–8. <https://doi.org/10.1016/j.jfca.2018.12.005>.
- [276] C.A. Esquerre, E.M. Achata, M. García-Vaquero, Z. Zhang, B.K. Tiwari, C.P. O'Donnell, Use of an NIR MEMS spectrophotometer and visible/NIR hyperspectral imaging systems to predict quality parameters of treated ground peppercorns, *LWT—Food Sci. Technol.* 131 (2020). <https://doi.org/10.1016/j.lwt.2020.109761>.
- [277] M.M. Oliveira, J.P. Cruz-Tirado, J.V. Roque, R.F. Teófilo, D.F. Barbin, Portable near-infrared spectroscopy for rapid authentication of adulterated paprika powder, *J. Food Compos. Anal.* 87 (2020). <https://doi.org/10.1016/j.jfca.2019.103403>.
- [278] F. Gao, D. Ben-Amotz, S. Zhou, Z. Yang, L. Han, X. Liu, Comparison and chemical structure-related basis of species discrimination of animal fats by Raman spectroscopy using near-infrared and visible excitation lasers, *LWT—Food Sci. Technol.* 134 (2020). <https://doi.org/10.1016/j.lwt.2020.110105>.
- [279] E.M. Achata, M. Oliveira, C.A. Esquerre, B.K. Tiwari, C.P. O'Donnell, Visible and NIR hyperspectral imaging and chemometrics for prediction of microbial quality of beef *Longissimus dorsi* muscle under simulated normal and abuse storage conditions, *LWT—Food Sci. Technol.* 128 (2020). <https://doi.org/10.1016/j.lwt.2020.109463>.
- [280] H. Jiang, X. Jiang, Y. Ru, J. Wang, L. Xu, H. Zhou, Application of hyperspectral imaging for detecting and visualizing leaf lard adulteration in minced pork, *Infrared Phys. Technol.* 110 (2020). <https://doi.org/10.1016/j.infrared.2020.103467>.
- [281] Z. Zhao, H. Yu, S. Zhang, Y. Du, Z. Sheng, Y. Chu, D. Zhang, L. Guo, L. Deng, Visualization accuracy improvement of spectral quantitative analysis for meat adulteration using Gaussian distribution of regression coefficients in hyperspectral imaging, *Optik* 212 (2020). <https://doi.org/10.1016/j.ijleo.2020.164737>.
- [282] Y. Dixit, H.Q. Pham, C.E. Realini, M.P. Agnew, C.R. Craigie, M.M. Reis, Evaluating the performance of a miniaturized NIR spectrophotometer for predicting intramuscular fat in lamb: a comparison with benchtop and hand-held Vis-NIR spectrophotometers, *Meat Sci.* 162 (2020). <https://doi.org/10.1016/j.meatsci.2019.108026>.
- [283] F. Mabeod, R. Boqué, A.Y. Alkindi, A. Al-Harrasi, I.S. Al Amri, S. Boukra, F. Jabon, J. Hussain, G. Abbas, Z. Naureen, Q.M.I. Haq, H.H. Shah, A. Khan, S.K. Khalaf, I. Kadim, Fast detection and quantification of pork meat in other meats by reflectance FT-NIR spectroscopy and multivariate analysis, *Meat Sci.* 163 (2020). <https://doi.org/10.1016/j.meatsci.2020.108084>.
- [284] P. Dumalisile, M. Manley, L. Hoffman, P.J. Williams, Discriminating muscle type of selected game species using near infrared (NIR) spectroscopy, *Food Control* 110 (2020). <https://doi.org/10.1016/j.foodcont.2019.106981>.
- [285] J.S. Kim, I.M. Hwang, G.H. Lee, Y.M. Park, J.Y. Choi, N. Jamila, N. Khan,

- K.S. Kim, Geographical origin authentication of pork using multi-element and multivariate data analyses, *Meat Sci.* 123 (2017) 13–20. <https://doi.org/10.1016/j.meatsci.2016.08.011>.
- [286] A. Ait-Kaddour, M. Loudiyi, A. Ferlay, D. Gruffat, Performance of fluorescence spectroscopy for beef meat authentication: effect of excitation mode and discriminant algorithms, *Meat Sci.* 137 (2018) 58–66. <https://doi.org/10.1016/j.meatsci.2017.11.002>.
- [287] S.W. Erasmus, M. Muller, M. Van Der Rijst, L.C. Hoffman, Stable isotope ratio analysis: a potential analytical tool for the authentication of South African lamb meat, *Food Chem.* 192 (2016) 997–1005. <https://doi.org/10.1016/j.foodchem.2015.07.121>.
- [288] A.N. Gatmaitan, J.Q. Lin, J. Zhang, L.S. Eberlin, Rapid analysis and authentication of meat using the MasSpec pen technology, *J. Agric. Food Chem.* 69 (2021) 3527–3536. <https://doi.org/10.1021/acs.jafc.0c07830>.
- [289] L. Chiesa, S. Panseri, S. Bonacci, A. Procopio, A. Zecconi, F. Arioli, F.J. Cuevas, J.M. Moreno-Rojas, Authentication of Italian PDO lard using NIR spectroscopy, volatile profile and fatty acid composition combined with chemometrics, *Food Chem.* 212 (2016) 296–304. <https://doi.org/10.1016/j.foodchem.2016.05.180>.