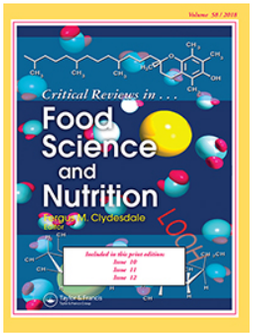

Taylor & Francis
Taylor & Francis Group

Critical Reviews in Food Science and Nutrition

ISSN: 1040-8398 (Print) 1549-7852 (Online) Journal homepage: https://www.tandfonline.com/loi/bfsn20

# Implementation of chemometrics in quality evaluation of food and beverages

Magdalena Efenberger-Szmechtyk, Agnieszka Nowak & Dorota Kregiel

**To cite this article:** Magdalena Efenberger-Szmechtyk, Agnieszka Nowak & Dorota Kregiel (2018) Implementation of chemometrics in quality evaluation of food and beverages, Critical Reviews in Food Science and Nutrition, 58:10, 1747-1766, DOI: 10.1080/10408398.2016.1276883

**To link to this article:** https://doi.org/10.1080/10408398.2016.1276883

Accepted author version posted online: 27 Jan 2017.
Published online: 21 Jul 2017.

CRITICAL REVIEWS IN FOOD SCIENCE AND NUTRITION
2018, VOL. 58, NO. 10, 1747–1766
https://doi.org/10.1080/10408398.2016.1276883

Taylor & Francis
Taylor & Francis Group



# Implementation of chemometrics in quality evaluation of food and beverages

Magdalena Efenberger-Szmechtyk, Agnieszka Nowak, and Dorota Kregiel

Institute of Fermentation Technology and Microbiology, Lodz University of Technology, Lodz, Poland

**ABSTRACT**

Conventional methods for food quality evaluation based on chemical or microbiological analysis followed by traditional univariate statistics such as ANOVA are considered insufficient for some purposes. More sophisticated instrumental methods including spectroscopy and chromatography, in combination with multivariate analysis—chemometrics, can be used to determine food authenticity, identify adulterations or mislabeling and determine food safety. The purpose of this review is to present the current state of knowledge on the use of chemometric tools for evaluating quality of food products of animal and plant origin and beverages. The article describes applications of several multivariate techniques in food and beverages research, showing their role in adulteration detection, authentication, quality control, differentiation of samples and comparing their classification and prediction ability.

**KEYWORDS**

Chemometrics; food quality control; authenticity; adulteration; multivariate tools

**GLOSSARY OF ACRONYMS:** AHC: agglomerative hierarchical clustering; ANN: artificial neural networks; BAI: biogenic amine index; BP-ANN: back-propagation artificial neural network; BP-MLP: back-propagation multilayer linear perceptron; CA: cluster analysis; CART: classification and regression tree; DRIFTS: diffuse reflectance infrared Fourier transform spectroscopy; EEMs: excitation-emission matrixes; ELISA: enzyme linked immunosorbent assays; ETAAS: electrothermal atomic absorption spectrometry; EVOO: extra-virgin olive oils; FAAS: flaming atomic absorption spectrometry; FDA: Food and Drug Administration; FDA: Fisher discriminant analysis; FGC-E nose: flash gas chromatography electronic nose; FTIR: Fourier transform infrared spectroscopy; FTIR-ATR/HT: FTIR attenuated total reflectance spectroscopy/high throughput; FT-MIR: Fourier transform mid-infrared spectroscopy; GC-MS: gas chromatography mass spectrometry; HCA: hierarchical cluster analysis; $^{1}$H NMR: proton nuclear magnetic resonance; HPLC: high performance liquid chromatography; HSI: hyperspectral imaging; ICP-AES: inductively coupled plasma atomic emission spectrometry; ICP-MS: inductively coupled plasma mass spectrometry; IRMS: isotope-ratio mass spectrometry; KNN: k nearest neighbor; LC-MS: liquid chromatography-mass spectrometry; LDA: linear discriminant analysis; LS-SVM: least squares support vector machines; MIR: mid-infrared; MLP: multilayer perceptron; MLR: multiple linear regression; NIRS: near-infrared spectroscopy; PCA: principal component analysis; PCR: principal component regression; PDO: protected designations of origin; PLS-DA: partial least squares discriminant analysis; PLSR: partial least squares regression; PNN: probabilistic neural networks; PTR-ToF-MS: proton transfer reaction-time of flight-mass spectrometry; QDA: quadratic discriminant analysis; RC-MLR/PLSR: regression coefficient MLR/PLSR; RDA: regularized discriminant analysis; RMS: root mean square; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; SEP: standard error of prediction; SERS: surface-enhanced Raman spectroscopy; SI-PLS: synergy interval PLS; SIMCA: soft independent modeling of class analogy; SOMs: Kohonen self-organizing map; SPA-MLR/PLSR: successive projections algorithm MLR/PLSR; SPME-GC: solid phase microextraction gas chromatography; SVM: support vector machine; SW-NIRS: short-wavelength near infrared spectroscopy; RC-MLR/PLSR: rating curve MLR/PLSR; TLC: thin-layer chromatography; TVC: total viable counts; UHT: ultra-high temperature; UPLC-QToF MS: ultra-performance liquid chromatography quadrupole time of flight mass spectrometry; UVE-PLS: uninformative variable elimination-PLS; VOO: virgin olive oils; WPTER: wavelet packet transform for efficient pattern recognition

## Introduction

Chemometrics is a chemical approach which enables the analysis of multidimensional data using mathematical statistics, probability theory and information technology. Unlike traditional statistical methods, which are often inadequate for accurate and deep interpretation of results, chemometrics allows valuable information to be obtained from a wide range of complex data sets, and facilitates the detection of hidden relationships between variables. Analyzing a number of processes occurring simultaneously in different environments is often challenging, and requires a multidimensional methodology. Statistical methods analyze only single variables, while chemometric methods provide a multi-pronged approach. Chemometric methods also reduce the number of analyses and measurements required, saving time and minimizing costs (Arvanitoyannis and van Houwelingen-Koukaliaroglou, 2003; Tyszkiewicz and Tyszkiewicz, 2004; Kamal and Karoui, 2015). Chemometric tools are therefore of great interest to industry.

Chemometric methods of pattern recognition can be divided into two types: supervised and unsupervised. Unsupervised pattern recognition determines the structure of a dataset on the basis of measurements. It reveals clusters without assumptions

**CONTACT** Magdalena Efenberger-Szmechtyk magdalena.efenberger-szmechtyk@dokt.p.lodz.pl Institute of Fermentation Technology and Microbiology, Lodz University of Technology, Wolczanska 171/173, 90-924 Lodz, Poland.

regarding the number or type of classes. The results are presented as dendrograms or two/three dimensional graphs. The most commonly used unsupervised pattern techniques are PCA and CA, including HCA. In supervised pattern recognition, the number of classes is already known, as is the membership of objects to those classes. New unknown samples can be classified into known classes according to experimental results. Supervised pattern recognition techniques include: LDA, QDA, RDA, K-NN, SIMCA, SVM, PLSR, and ANN methods (Berrueta et al., 2007; Zielinski et al., 2014b) (Table 1).

In recent years, consumer awareness regarding issues related to food authenticity and labeling has increased. Detection of adulteration is of great importance to ensure the quality and safety of food. Unfortunately, regulations regarding the labelling of food products have in some cases proven to be ineffective at preventing trade fraud and the adulteration of food. Therefore, new techniques are being developed to enable fast and inexpensive evaluation of food authenticity. Attention has been focused on chemometric tools, which classify products based on their chemical composition. Food products can be differentiated according to their geographical and botanical origins or production techniques (Zielinski et al., 2014b; Kamal and Karoui, 2015). Chemometrics is also useful for detecting the replacement of original ingredients with other additives. Numerous research studies have been conducted into the use of chemometrics in combination with analytical methods, especially chromatography and spectroscopy (Aculey et al., 2010; Kamal and Karoui, 2015; Elzey et al., 2016; Melucci et al., 2016). Other studies have investigated the possibility of using chemometric methods in food microbiology, to evaluate the level of microbial spoilage in food products, identify microorganisms or detect mycotoxins (He and Sun, 2015; Sendin et al., 2016).

This article describes the chemometric methods most commonly used in food chemistry and food microbiology. It provides a comprehensive overview of their application for the analysis of food of animal origin (meat and meat products, fish and seafood, milk, and dairy products) as well as beverages and food of plant origin (fruits and fruit products, honey, crops, olive oils, coffee, tea, and alcohols.

**Table 1.** Most commonly used chemometric tools.

| Chemometric tool | Overall features |
|---|---|
| | **Unsupervised Pattern Recognition** |
| Cluster analysis (CA) | Objects are grouped into clusters according to their similarity (proximity). The correctness of clusters is monitored and deviating points are detected. The results of CA are usually presented in the form of a dendrogram, a diagram (tree) showing the grouping of objects. |
| Hierarchical cluster analysis (HCA) | The simplest and most commonly used form of cluster analysis. HCA creates a classification hierarchy starting from a division in which each object is a single cluster and ending with a division in which all the objects constitute a single cluster. |
| Principal component analysis (PCA) | This method reduces the dimensionality of the original data set and leads to the creation of new dimensions of data. PCA converts strongly correlated input variables into uncorrelated values, called principal components (PCs). The first components account for the majority of the variability in the original data set and the rest can be omitted without significant loss of information. This technique enables a graphical presentation of the multi-dimensional structure of a data set. |
| | **Supervised Pattern Recognition** |
| **Discriminant analysis** | Techniques that assign samples to specific classes |
| Linear discriminant analysis (LDA) and Fisher discriminant analysis (FDA) | Based on a linear function, LDA is used for linear classification or leads to dimensionality reduction. Although similar to Fisher's discriminant analysis (FDA), LDA includes assumptions as to the normal distribution of classes and equal covariance of classes. |
| Quadratic discriminant analysis (QDA) | A technique closely related to LDA. It omits the assumption of equal covariance classes, but the assumption of normality is maintained. It should not be used if the sample size is very small. |
| Regularized discriminant analysis (RDA) | A compromise between LDA and QDA, which uses regularization of the covariance matrix. Consequently, it can be used with small sample sizes. |
| K-nearest neighbor (KNN) | Nonparametric method used for object classification and prediction. Classification is based on similarities between objects. Samples are classified based on class membership of its k-nearest neighbors. The model is based on parameter K, selected by cross-validation procedures to give the lowest classification error. Due to limits on memory and processing power, the KNN method should not be used as a single classification algorithm, but as a reference in comparative analysis for other non-linear classifiers, such as ANN. |
| Soft independent modeling of class analogy (SIMCA) | SIMCA is based on PCA model and defines classes according principal components (PCs) describing the greatest variability. Classes are created according to distances (similarities) between objects. Classes may overlap with a sample belonging to any number of classes, or may not fit into any class. |
| Support vector machine (SVM) | A linear classification method, usually applied, however, to solve nonlinear problems. This model is based on the nuclear function (kernel), and is used when a linear separation of objects is impossible. Original objects are transformed into a new space in which classes can be linearly separated. |
| Partial least squares discriminant analysis (PLS-DA) | Discriminatory variant of PLS model (PLS-DA), used to classify samples. |
| **Correlation and Regression Analysis** | |
| Partial least squares regression (PLSR) | Multivariate calibration method. PLS regression constructs a single quantitative model for many analyzed variables and is used in combination with one dependent variable. It is used as a prognostic model. |
| **Artificial Neural Networks** (ANN) | A prognostic model built using an extensive network of neural nodes which exchange messages, simulating the operation of the human brain. Taught by stimulus and response, not by an algorithm. It is impossible to obtain a correlation between the input and the output of a neural network. |

## Chemometrics in quality evaluation of food of animal origin

### *Chemometrics in quality evaluation of meat and meat products*

Consumers are increasingly aware of issues regarding the authenticity and quality of meat and meat products, with high-profile cases of food adulteration and false labeling being reported in the media. Identification of adulteration is essential not only from the point of view of the law, but also for moral, religious, cultural, diet and health reasons. Sometimes, meat products are made with lower value meats or meat species other than those declared (Balog et al., 2016). Furthermore, muscle proteins can be replaced with less expensive vegetable proteins, such as lupine, pea and soy proteins. These substitutes are potential allergens, and can have a significant effect on human health, especially among people suffering from allergies (Hoffmann et al., 2016). Therefore, research is underway to develop fast, cheap and reliable methods of assessing the quality and authenticity of meat products. Of particular interest is the application of chemometrics, which enables classification of meat products on the basis of their chemical composition (Kamruzzaman et al., 2015).

Studies show that Raman spectroscopy and PCA can be successfully used to differentiate salami products based on fat composition according to meat origin (Boyaci et al., 2014). Similarly, Raman spectroscopy in combination with PLS-DA makes it possible to distinguish three types of sausage (poultry, turkey and mixed) according to their contents of fatty compounds. PLS-DA models revealed specificity and sensitivity values of 88.9–100%, accuracy and efficiency values in the range of 94.7–100% and 94.4–100%, respectively and correlation coefficients of between 0.90 and 1.00 (Campos et al., 2014). Traditional Indonesian meatballs (Bakso) are made of beef, but they can be also prepared with chicken, pork or fish meat. Due to its high price, producers substitute beef with a wild boar meat, which is considered by Muslims as haram and forbidden for consumption. In studies by Guntarti et al. (2015), FTIR spectroscopy in combination with PCA was used to distinguish meatballs made from beef and wild boar meat. Only two samples of meatballs labeled as 100% beef did not belong to any group, which suggests that they could have been made from a different kind of meat. A PLSR model was also built to quantify wild boar fat. The correlation coefficient between the actual value and the value predicted by FTIR was 0.998 (RMSEC = 2.00%).

Irradiation is a very effective food preservation technology, which is used to reduce pathogenic bacteria. The use of irradiation in the European Union is regulated by two directives: Framework Directive 1999/2/EC and Implementing Directive 1999/3/EC. Currently, the products authorized for irradiation are: fruit and vegetables including root vegetables, cereals, cereal flakes, rice flour, spices, condiments, fish, shellfish, fresh meats, poultry, frog legs, raw milk camembert, gum arabic, casein/caseinates, egg whites and blood products (http://ec.europa.eu/food/safety/biosafety/irradiation/legislation/index_en.htm last update: 26.04.2017). Previously, use of irradiation was only allowed for preserving dried aromatic herbs, spices and vegetable seasoning. Meat and meat products were not included. Zanardi et al. (2015) reported the use of PCA for the classification of irradiated and non-irradiated beef samples based on $^{1}$H NMR spectra. Analysis revealed the occurrence of reliable biomarkers for distinguishing between irradiated and non-irradiated beef: glycerol, lactic acid esters, tyramine and p-substituted phenolic compound. 184 It was possible to separate samples treated with higher irradiation doses (8 and 4.5 kGy) from the controls, but samples treated with low irradiation doses (2.5 kGy) were grouped together with the controls. These results show that PCA is not always an effective chemometric technique. Although the first three PCs explained 86.73% of the total variance, separation was not achieved. With unsupervised learning techniques, no information is given concerning the data and class labels. In order to confirm the lack of differences between control samples and samples with low doses of irradiation, supervised learning should additionally be applied.

Other studies have described the use of chemometric models for predicting meat spoilage processes. The microbiological safety of meat and meat products is crucial for public health. It is therefore of great importance to detect pathogens and to ensure that levels of microorganisms are within acceptable limits. Contamination of meat and meat products is caused by many different microorganisms, mainly bacteria, such as Acinetobacter sp., *Aeromonas* sp., *Brochothrix thermosphacta, Flavobacterium* sp., *Moraxella* sp., *Pseudomonas* sp., *Psychrobacter* sp., *Streptococcus* sp., and lactic acid bacteria. During metabolic processes, bacteria synthesize undesirable compounds, such as indole, skatole, ammonia, hydrogen sulfide, biogenic amines, lactic acid, $CO_2$, acetoin, diacetyl, acetic acid, and valeric acid, which adversely affect the organoleptic characteristics of meat and meat products. Therefore, the levels of these metabolites can be correlated with the degree of microbiological contamination in food products (Ellis and Goodacre, 2001).

Papadopoulou et al. (2011) applied FTIR spectroscopy in combination with a PLSR model to monitor microbial counts during aerobic storage of pork meat samples at different temperatures (0°C, 5°C, 10°C, and 15°C). Spectral data were correlated with microbiological load (TVC), the number of *Pseudomonas* sp., *B. thermosphacta* and lactic acid bacteria. In each case, the correlation was greater than 0.80. Moreover, the PLS-DA approach achieved 100% accuracy for the classification of spoiled meat samples, with 93.3% and 86.7% accuracy for fresh and semi-fresh samples, respectively. Classification errors could in some cases result in uncertainty during sensory analysis and variability in the samples. Analysis was performed by a sensory panel, whose evaluations may not be reliable. In studies by Ammor et al. (2009), chemometric techniques were used to monitor spoilage processes in ground beef stored under different atmospheric conditions (aerobically, under modified atmosphere and using an active packaging) and temperatures (0°C, 5°C, 10°C, and 15°C). The results indicate that FTIR spectroscopy can reveal metabolic fingerprints, reflecting the degree of food spoilage. The FDA method enables beef to be differentiated in terms of its freshness and storage atmosphere and in both cases 100% of samples were correctly classified. When cross-validated, FDA provided 76.3% correct classification according to freshness and 92.5% according to storage atmosphere. No fresh sample was included in the spoiled group or vice versa. Analysis was also performed by a sensory panel,

whose evaluations may not be reliable. PLSR was used to predict the number of microorganisms in ground beef and its pH, with fits of $R^2$ 0.80 and 0.92, respectively. Wang et al. (2011) built a model using LS-SVM to predict total viable bacteria count in pork ($R^2 = 0.9426$). Eight selected optimal wavelengths were investigated, to construct a TVC prediction model ($R^2 = 0.9236$). Huang et al. (2013) developed a TVC predictive model for pork using BP-ANN. These authors report that their model based on fusion data (image and spectra data) was far superior ($R^2 = 0.83$; RMSEP = 0.243 log CFU/g) to either of two separate models: one model based on images ($R^2 = 0.4104$; RMSEP = 1.169 log CFU/g) and another based on spectra ($R^2 = 0.7867$; RMSEP = 0.459 log CFU/g). This observation is very important. Through data fusion it is possible to assess both the internal (chemical composition, tissue structure, etc.) and external (color, texture, etc.) attributes of pork meat.

Lin et al. (2004) describe the use of SW-NIR spectroscopy coupled with PCA and PLSR to detect and quantify microbial loads in chicken meat. PCA separated samples held for 8 h or longer more clearly than the control (held for 0 h). However, samples stored for less than 8 h were not separated from the control. PLSR model showed potential to predict microbial load in chicken meat ($R^2 = 0.91$; SEP = 0.48 log CFU/g). PCA was used to cluster chicken meat stored in a modified atmosphere in terms of microbial load, sensory evaluation and the concentration of metabolites (Vainionpää et al., 2004). The use of ANN to assess meat freshness has also been reported (Argyri et al., 2010). Beef samples were classified concerning their biochemical profiles, determined by FTIR spectroscopy. MLP—artificial neural networks showed high accuracy. The network was able to classify correctly 91.7% of fresh samples 94.1% of spoiled samples and 81.2% of semi-fresh samples with none of the fresh samples being categorized as spoiled meat, or vice versa. The authors also used MLP-neural networks to predict the number of microorganisms in meat based on the results of FTIR spectroscopy (Bias factor $B_f = 0.951–1.031$).

Montel et al. (1996) reported the use of PCA for the identification of *Staphylococcus* and *Micrococcus* strains, based on their biochemical properties. PCA revealed a relation between the qualitative and quantitative compositions of volatile compounds in meat products and inoculation strains. The strongest dry-cured odor was observed in samples inoculated with *Staphylococcus carnosus* and *Staphylococcus xylosus*, which have low lipolytic and proteolytic activities and do not produce acetoin, but which reduce nitrate very effectively.

Biogenic amines can occur naturally in foods. However, they are also released during proteolytic processes. Excessive levels of biogenic amines are carcinogenic. Meat, due to its high protein content, is an excellent environment for the synthesis of biogenic amines. Meat spoilage bacteria, including *Escherichia coli, Pseudomonas* sp., *Proteus* sp., *Micrococus* sp., and *Lactobacillus* sp., reveal high activity in the presence of proteolytic enzymes. The number of these microorganisms can therefore be associated with the concentration of biogenic amines in meat and meat products (Bota and Harrington, 2006), and biogenic amines may be used as indicators of meat quality (Lázaro et al., 2015). The effect of microorganisms, including starter cultures, on the synthesis of biogenic amines in fermented meat products was considered as an indicator of hygienic conditions and good manufacturing practices by Parente et al. (2001). PCA showed no association between the content of biogenic amines, ripening time, pH level, microbial load or type of starter culture used. The first four components explained 75% of total variance. Cheng et al. (2016) studied several PLSR and MLR models (RC- PLSR, RC-MLR, SPA-PLSR, and SPA-MLR) to determine BAI in pork. The most accurate results ($R^2 = 0.957$; RMSEP = 4.866 mg/kg) were obtained using an optimized and simplified RC-MLR model, while the classical PLSR model showed the lowest prediction accuracy ($R^2 = 0.895$; RMSEP = 7.559 mg/kg). These results suggest that instead of classical regression models, such as MLR and PLSR, simplified models can be used with higher prediction ability than their classical equivalents.

### Chemometrics in quality evaluation of fish and seafood products

Seafood products can be classified in terms of biological species, geographical origin and method of production. This information must be verified not only to avoid mislabeling and trade fraud but also to ensure food safety and regulatory compliance.

Ortea and Gallardo (2015) used stable isotope ratio analysis ($C^{13}$, $N^{15}$) and multi-element analysis (As, Cd, Pb, S, and P) combined with chemometrics to study seafood product authenticity. PCA and CA well separated samples in terms of production technique but failed to differentiate in terms of origin and species. DA achieved 100% correct classification of shrimps was achieved in terms of origin and production method, and 93.5% accuracy in terms of species when all seven variables were considered. This paper confirms that, in some cases, supervised learning methods are more appropriate than unsupervised learning methods. Using unsupervised techniques may be suitable when there are significant differences between the samples. Moreover, the number of samples can influence differentiation. For wild/farmed shrimps, the classification ability of DA decreased with smaller amounts of data.

Ottavian et al. (2012) reported effective use of NIR spectroscopy for the authentication of wild European sea bass. Models based on PLS-DA and WPTER confirmed that NIR spectroscopy can be applied to distinguish farmed from wild sea bass. Classification results obtained using NIR spectroscopy were almost identical to those obtained by analyzing chemical properties and morphometric traits. The most predictive spectral regions were related to fat, fatty acids and water content.

Similar studies have been conducted for turbot (*Psetta maxima*), which were classified according to production method and fishing areas (Denmark, the Netherlands, Spain). Fatty acid composition and stable isotope ratios ($C^{13}$ and $N^{15}$) were determined in muscle tissue from turbot. PCA separated farmed from wild fish but failed to distinguish between fish with different geographical origins. Better separation was achieved using LDA. Recognition accuracy was 100% for samples classified according to production method and 93.3% in terms of catching zone. Only one wild sample was incorrectly classified as a farmed sample using SIMCA (Busetto et al., 2008).

Chemometric techniques have been used to identify mussels (*Mytilus galloprovincialis*) from Galicia in Spain. These mussels

have been given Protected Designation of Origin (PDO) status by the European Commission. However, many mussels sold in Galicia (in particular frozen and canned mussels) do not come from this region. Identification of original mussels from Galicia was therefore performed, based on the content of trace elements determined by ICP-MS. Using LDA, prediction accuracy was 95.6%, although false positive and false negative results occurred. SIMCA failed to classify samples according to different parts of Galicia (prediction ability of 39.1–85.7%), but managed to accurately distinguish mussels from the region as a whole and those from elsewhere (prediction ability of 100%). However, these results were obtained after reducing the number of variables. Before this reduction, 100% of non-Galician samples but only 75% of Galician samples were classified correctly. It should be mentioned that even in the training set, less than 100% recognition ability (over 90%) was achieved for Galician mussels. The ANN model was the most effective, however, classifying samples with recognition accuracy of 100%, including according to local area of origin. (Costas-Rodríguez et al., 2010). SIMCA has also been used to identify different shrimp species (*Litopenaeus vanname, Penaeus monodon*) and their geographical origins (Ecuador, Philippines, Thailand, United States), based on multidimensional fluorescence fingerprinting (Eaton et al., 2012).

Short-wavelength near infrared spectroscopy combined with chemometric techniques have been used to determine the levels of microorganisms and monitor the spoilage process in rainbow trout fillets (*Oncorhynchus mykiss*). PCA based on SW-NIRS results showed that it is possible to group trout fillets according to storage time and temperature. The control samples (day 1) were clearly separated from the samples stored for 4 days or more at 4 °C. The stored samples held for 10 h or longer at 21 °C were also separated from the control. Validation of SW-NIR spectroscopy using PLSR indicated that this method can be successfully used to assess microbiological counts in trout meat ($R^2 = 0.97$. SEP = 0.38 log CFU/g for fresh side at 4 °C; $R^2 = 0.94$, SEP = 0.53 log CFU/g for skin side at 4°C; $R^2 = 0.82$, SEP = 0.82 log CFU/g for minced sample at 21 °C) (Lin et al., 2006).

Certain bacteria species isolated as contaminants in salmon can be characterized on the basis of synthesized volatile compounds (e.g. *Carnobacterium piscicola* according to the concentration of diacetyl and *B. thermosphacta* based on the content of 2-heptanone and 2-propanone). Using PCA and GC-MS, Joffraud et al. (2001) identified *Aeromonas* sp., *Shewanella putrefaciens* and *Enterobacteriaceae, Lactobacillus* sp., *C. piscicola* and *B. thermosphacta* as the bacteria mainly responsible for the sensory deterioration of products. The composition of volatile compounds was found to be characteristic for certain bacterial species.

Chemometric techniques are widely used for the analysis of toxic trace elements in seafood. PCA based on FAAS and ETAAS has been used to compare different methods of seafood pre-treatment for removing As, Cd, Cr, Cu, Fe, Mg, Mn, Ni, Pb, Se, and Zn (Bermejo-Barrera et al., 2001). Aflaki et al. (2016) used PCA cluster biogenic amines based on fish species and storage time. Fish species differed considerably in terms of their contents of spermine, spermidine, and phenylethylamine. The profiles of biogenic amines can therefore be associated with fish species and storage time.

### Chemometrics in quality evaluation of dairy products

Milk and dairy products are consumed all over the world. Whole milk products contain high levels of nutritious proteins, and can reach relatively high prices on the market. This food group is particularly susceptible to adulteration, as producers seek to gain more profit. They modify the composition of dairy products by replacing some ingredients with non-dairy or other dairy additives. Chemometric methods provide a useful tool for verifying the authenticity and quality of dairy products (Kamal and Karoui, 2015).

Souza et al. (2011) analyzed the presence of adulterants such as starch, chlorine, formol, hydrogen peroxide, and urine in Brazilian UHT milk. Although starch was not detected, other adulterants were found in all the samples. Urine was the most common adulterant (detected in 55% of samples). Urine is added to disguise the addition of water, increasing the possibility of microbial contamination. PCA and HCA methods were successfully applied to verify the occurrence of adulterants in milk, and associate them with the geographical locations of industrial plants. A similar study was performed by Santos et al. (2013), who report the use of SIMCA to discriminate control milk samples from samples adulterated with whey, synthetic milk, urea, and hydrogen peroxide. 343 Classification ability was from 90% (control) to 98% (samples adulterated with urea). Five of the samples did not match any of the groups. PLSR was used to estimate levels of adulteration in dairy products based on MIR-microspectroscopy and $R^2$ ranged from 0.90 (hydrogen peroxide) to 0.98 (urea and synthetic urine)..

Melamine is a substance containing high level of nitrogen (66.7% by mass). It is illegally used in industry to increase protein content in food products. Conventional methods (Kiejfahl and Dumas tests), which are used to examine total nitrogen content, do not allow identification of the nitrogen source (Elsheikh et al., 2016). Wu et al. (2016) used NIR spectroscopy combined with chemometrics to identify and quantify melamine in milk. PLSR and UVE-PLS analysis were used to construct quantitative models with $R^2$ values of 0.93 and 0.97, respectively. However, UVE-PLS was more accurate. A PLS-DA model was constructed to classify unadulterated and adulterated milk samples, with sensitivity and specificity of 100%. The specificity of the SIMCA method was also 100%, but sensitivity was 81.8%.

Dairy products are a suitable environment for the growth of microorganisms, so maintaining their microbiological safety is of great importance. Traditional methods used to quantify spoilage bacteria are time-consuming. Nicolaou and Goodacre (2008) used FTIR (FTIR-ATR and FTIR-HT) spectroscopy in combination with PLSR to detect the relationship between metabolic fingerprints and microbial load. PLSR enabled quantification of total viable counts, based on FTIR spectra but FTIR-ATR was more accurate (for the test set: RMS error = 0.25) than FTIR-HT (for the test set: RMS error = 0.84). According to Paradkar and Irudayaraj (2002), PLSR combined with FTIR spectroscopy has also been used to estimate cholesterol content, with an $R^2 > 0.99$ and SEP < 0.98

Cruz et al. (2013) discriminated between low and full-fat yoghurts in terms of pH, color and firmness, using different chemometric methods (PCA, HCA, SIMCA, K-NN, and PLS-DA).

Total separation was not achieved using PCA or HCA. However, KNN and PLS-DA methods resulted in prediction ability of 100% for full fat yoghurts, 100% for low-fat yoghurts (KNN) and 95% for low-fat yoghurts (PLS-DA). SIMCA classified full fat yoghurts with 100% prediction ability but low-fat yoghurts with only 50% prediction ability. A large set of samples (83 full-fat and 43 low-fat yogurts) was studied. When there are a large number of samples ($n > 20$), unsupervised techniques may not be appropriate, and supervised pattern recognition techniques may provide better separation. Moreover, the main drawback of HCA is that it requires a decision to be made regarding the number of clusters in which the variables should be grouped. If an algorithm is applied, the results may not be unique.

Raw cow milk varies in terms of the concentration of minerals and trace elements. This variation is the result of a range of factors, including animal species, health condition, lactation state, diet, the season, the geographical situation of the farm and soil type. LDA provided information that complemented the results of PCA and enabled cow milk to be differentiated in terms of seasonal criteria (97.1% of samples correctly classified), but very low classification accuracy (47.8%) was observed in terms of origin. The geographical origin of milk is connected with different feeding practices, such as use of nutritional additives and supplements, as well as to contamination and stress factors (climate, disease or lactation), which could explain such low classification accuracy. Use of PCA reduced the number of elements from 16 to 5 principal components (Sola-Larrañaga and Navarro-Blasco, 2009).

Cheese is another group of dairy products prone to adulteration practices. Vegetable oils and fats can be substituted for milk fat to produce imitation cheese. Monakhova et al. (2013) used $^{1}$H NMR and $^{13}$C NMR spectroscopy together with chemometric techniques to detect cheese adulterated with vegetable fat. PCA separated imitation from unadulterated cheese samples. Imitation cheese samples were located in the range of negative PC1 values ($^{1}$H NMR) or positive PC3 values ($^{13}$C NMR). However, $^{1}$H NMR spectroscopy provided better differentiation. By removing imitation products, it was possible to group the cheeses by type (Edamer, Gouda, Feta, or Emmentaler). Using a PLSR model, NMR spectra were correlated with the content of certain compounds (saturated and unsaturated fatty acids and their esters), determined using the GC method ($r^2 = 0.75–0.95$). This enabled quantification of vegetable fat substitutes.

De Sá Oliveira et al. (2016) employed Raman spectroscopy and chemometric tools to detect starch (another commonly used adulterant in dairy products) in spreadable cheese. PLS-DA identified adulterated and unadulterated cheese samples with 100% accuracy. PLSR was used to quantify starch content. The correlation between the reference values and the values estimated by the model was 0.989 for the calibration set and 0.984 for the validation set. Raman spectroscopy can detect starch with a minimum concentration of 6%. However, implementation of chemometrics can improve its sensitivity. The limit of detection was 0.34% for the model and the limit of quantification was 1.14%.

Oscypek is a Polish cheese with PDO status. Authentic PDO Oscypek is made in the Polish Tatra Mountains (Podhale Region), using raw ewe's milk and traditional technology. The addition of cow's milk is allowed in quantities up to 40% if the milk comes from a mountain cow breed known as Polish Red. PDO Oscypek has a characteristic composition of volatile compounds, which differentiates it from Oscypek-like cheeses. Majcher et al. (2015) used the GC-MS method (electronic nose) in combination with chemometrics to identify original Oscypek and detect mislabeling and adulteration. PCA perfectly separated PDO Oscypek from adulterated cheeses. LDA and SIMCA enabled classification of original cheese with the same classification accuracy (100%). Using SVM, classification accuracy of 97.9% was achieved and only industrially-produced samples were misclassified. This may have been due to the influence of technological processes on the chemical composition of the industrially-produced samples.

## Chemometrics in quality evaluation of beverages and food of plant origin

### *Chemometrics in quality evaluation of fruits and fruit products*

Verifying the authenticity of food products of plant origin is extremely difficult, due to the combination of factors involved, including cultivar, growing region, soil conditions, degree of plant maturity, storage conditions and processing technology. Fruit juices are often adulterated by the addition of water, inexpensive ingredients such as sugar, acids or colorants and even with peel extracts and other less expensive juices (Muntean, 2010; Jandrić et al., 2014).

Jandrić and Cannavan (2015) classified citrus fruits/juices according to origin, variety and production techniques using UPLC-QToF MS coupled with chemometrics. PCA, PLS-DA, and SIMCA enabled effective separation of authentic from adulterated juice samples. SIMCA and PLS-DA enabled to detect samples adulterated with other citrus juices at 1% and with water at 5%. PLS-DA and SIMCA provided recognition ability of 100% for citrus fruits, oranges of various geographical origins and fresh squeezed commercial orange juices. Lower recognition ability (80%) was achieved for orange juices made from concentrates. In this study, PCA successfully differentiated fruit juice samples and enabled the data to be structured. However, since PCA is not a classification model, PLS-DA and SIMCA methods were used as classification tools to confirm the results. Discrimination was observed between hand-squeezed and commercial juices, probably due to differences in polyphenol content related to the use of each technique.

Fruits are characterized by the occurrence of phenolic compounds. Abad-García et al. (2012) reported the use of CA, PCA and LDA to group different citrus species based on the concentration of their polyphenols. In this paper, unsupervised techniques were used to reveal the data structure and reduce the number of variables. Classification was then performed with a limited amount of variables. The prediction accuracy of the LDA classification model was 100% for tangerine and grapefruit juices, but lower for sweet orange and lemon juices (95.7% and 89.9%, respectively). PLSR was found to be a promising predictive model for estimating the percentage of adulteration in sweet orange juices ($R^2 = 0.9541$ and after cross-validation

$R^2 = 0.9508$) For PCR model, the correlation coefficients were lower ($R^2 = 0.9254$ and after cross-validation $R^2 = 0.9240$).

Braga et al. (2013) found that aroma compounds can be used as markers to differentiate apple juices and fermented apple juices made from different varieties of apple. PCA and HCA methods separated juices based on aroma composition and physicochemical properties. The major compound identified in all samples was ethanol.

Guo et al. (2016) constructed multivariate calibration models (PLSR, LS-SVM, and BP-ANN) based on NIR spectra for quantitative analysis of total sugar, total acid, total phenolic content and antioxidant activity in jujube fruits. The best classification results were obtained using the LS-SVM model with prediction rate of 93.8% and calibration rate 98.5%, whereas BP-ANN provided prediction rate of 81.2% and calibration rate of 100%. The LS-SVM model was also the most suitable quantification model ($R^2 = 0.904–0.978$). PCA and LDA revealed differences between jujube fruits of different geographical origins. Guo et al. (2009) used HPLC to determine levels of triterpenoid acids in dried *Ziziphus jujuba* fruits. PCA and HCA methods differentiated and separated cultivars of *Z. jujuba* based on their triterpenoid acid content. Interestingly, the Dongzao cultivar was completely separated, suggesting considerable genetic distance from the other cultivars. Moreover, fruits from the same cultivation region were grouped together, suggesting similar triterpenoid acid content. This was probably due to the fruits having been grown under the same climatic conditions and in the same soil type.

Brazilian frozen pulps from a range of fruits were analyzed based on chromaticity, phenolic compounds, carotenoids and antioxidant activity. Three clusters were obtained using HCA and corroborated with PCA. Strawberry, red fruits, blackberry, acai, and grape pulps were found to contain the highest concentrations of phenolic acids and flavonoids, and to have the greatest antioxidant activity. Coconut, graviola, cocoa, pineapple, mint, pineapple, umbu, seriguela, tamarind, peach, and cashew pulps showed the lowest phenolic content and free-radical scavenging activity (Zielinski et al., 2014a).

### *Chemometrics in quality evaluation of honey*

Honey is a natural product consisting mainly of sugar and water. It also contains minor amounts of minerals, proteins, vitamins, organic acids, flavonoids, phenolic acids, enzymes, and other phytochemicals. These components determine the quality of honey. The chemical composition of honey, as well as its quality, depends primarily on the botanical origin of the plants. Therefore, much research has been conducted to identify markers for distinguishing between different types of honey (Anklam, 1998; Zhao et al., 2016).

Honeys with different botanical and geographical origins, produced under varying climatic conditions, also differ in terms of their flavonoid content. Several studies have focused on the composition of polyphenols in different types of honey (Gómez-Caravaca et al., 2006; Hadjmohammadi et al., 2009; Bertoncelj et al., 2011). Bertoncelj et al. (2011) used the LDA method to classify varieties of Slovenian honey based on their flavonoid and abscisic acid (phytohormone) contents. However, only 85% of the honey samples were correctly classified. In the case of acacia and linden honey, classification accuracy was 100%, for floral honeys—80%, while for spruce honey it was only 60%. No specific markers were found for the differentiation of honeys. However, pinobanksin galangin and abscisic acid were the most discriminating variables. Maybe, other phenolic compounds, e.g., phenolic acids and flavonoid glycosides should be also analyzed to achieve better classification results.

In studies by Zhou et al. (2014), polyphenols (kaemferol, morin, and ferulic acid) were found to be effective markers for the identification of chaste honey and rape honey. Multivariate methodologies such as PCA, PLS, PLS-DA, and SIMCA were able to differentiate honey according to floral origin. PCA successfully differentiated samples and was used to build classification models. For the calibration set using a SIMCA model, discrimination accuracy was 94.53%. For the predictive set, discrimination accuracy was 96.43%, with $R^2$ values of 0.8002 and 0.8088. The PLS model was characterized by low RMSECV = 0.1463 for the training test, and RMSEP = 0.1929 for the test set, signifying that the model was very accurate. PLS-DA classified honey samples into two groups.

Daher et al. (2008) used PCA to separate nectar honeys from honeydew honeys. The main marker was salicylic acid, an aromatic compound. PCA was also used by Flanjak et al. (2016), for the differentiation of Croatian honey types (black locust, lime, sage, chestnut, and honeydew), based on antioxidant capacity and physicochemical properties. In recent studies, 7 monofloral honey varieties made from citrus, chestnut, sunflower, honeydew, robinia, rhododendron, and linden tree were studied, according to their content of volatile organic compounds. PCA did not separate honey samples, LDA upon PCA provided better separation and correctly classified 89% samples in the test set. The most precise classification results were achieved using stepwise LDA and PNN methods, with accuracy rates of 100% and 90%, respectively. (Schuhfried et al., 2016).

Studies of honeydew, buckwheat and rape honey from Poland have shown differences in terms of mineral content (Al, B, Ba, Ca, Cd, Cr, Cu, K, Mg, Mn, Na, Ni, Pb, Sr, and Zn). The elements were determined using ICP-MS. LDA was able to differentiate between types of honey with 100% accuracy. However, the results were less satisfactory for the geographical origin of buckwheat honey (91.7%), rape honey (90.9%), and honeydew honey (97.1%) (Chudzinska and Baralkiewicz, 2011). Interestingly, studies using PCA and CA methods have revealed significant differences between the concentrations of metals in honey collected in urban, industrial and rural areas. The composition of honey may reflect the degree of contamination in honey plants, soil and air (Rodríguez García et al., 2006).

Raman spectroscopy has been applied to determine the content of selected sugars in honey. PCA was used to group honey samples with high glucose, fructose, sucrose and maltose contents. The results were validated using PLSR ($R^2 = 0.949–0.968$) for, and ANN ($R^2 = 0.956–0.978$) models, suggesting that Raman spectroscopy can be used to quantify the sugar content in honey (Özbalci et al., 2013). ANN is a relatively new predictive model, which does not require any threshold values nor use a strict mathematical equation. Contrary to traditional methods, in which the strict rules and algorithms may cause difficulties, ANN is taught by examples (Baş et al., 2007).

## Chemometrics in quality evaluation of crops

Crops such as maize, wheat, rice, and grains are consumed in large quantities and therefore deserve consideration in terms of safety and quality. Cereal products are very prone to microbial contamination, especially by fungi. Various species of *Aspergillus, Fusarium*, and *Penicilium* have been detected in cereal products. Not only fungal spores, but also their mycotoxins, are a huge problem for producers (Sendin et al., 2016). Mycotoxins are secondary fungal metabolites and can be synthesized at any stage of product processing. Most are chemically stable, heat resistant, and very difficult to inactivate. Mycotoxins can have harmful effects on the health of both humans and animals, including carcinogenic, neurotoxic, hepatotoxic nephrotoxic, teratogenic, and immunotoxic activities (Min and Cho, 2015).

The most common mycotoxins that occur in cereal products are fumonisins, trichothecenes, and zearalenones (produced by *Fusarium* sp.), as well as aflatoxins (produced by *Aspergillus* sp.), and ochratoxins (produced by *Aspergillus* sp. and *Penicilium* sp.) (Min and Cho, 2015; Köppen et al., 2010). Classical methods for the detection of mycotoxins include HPLC, ELISA, TLC, and GC. The most common method is HPLC. However, these techniques are time-consuming and demanding, and spectroscopic methods in combination with chemometrics could therefore represent a promising alternative (Min and Cho, 2015).

Giacomo and Stefania (2013) used NIRS and multivariate statistical regression (PLSR) to determine the content of fumonisins in maize. The concentration of fumonisin was measured using HPLC and correlated with NIR spectra. Two models were obtained with high correlations in terms of calibration (0.995, 0.998) and validation (0.908, 0.909). The models described the relation between fumonisin content and NIR spectra well. The authors also obtained a satisfactory limit of detection, compatible with the EU-defined threshold of 4 mg/kg. Lee and Herrman (2016) report the use of SERS for detecting fumonisins in maize. Chemometric classification methods such as KNN, PLS-DA, and LDA, and quantification methods such as PLSR, MLR, and PCR were evaluated. KNN models enabled correct classification rates in the ranges of 79.6–84.5% and 70.6–79.4% for the training and validation datasets, respectively. The LDA models revealed lower classification accuracy for the training dataset (68.0–79.6%), but had comparable classification accuracy for the validation dataset (67.6–85.3%). The disadvantage of LDA is that it is sensitive to outliers and fails when discriminatory information is contained in the variance of the data, not in the mean. The KNN algorithm is often influenced by irrelevant attributes, which contribute to meaningless results. The PLS-DA model had 100% classification accuracy for the training dataset, but the lowest classification accuracy for the validation dataset. The class separation in dataset could be nonlinear and thus analysis was problematic. Turning to the quantification models, the correlation coefficients for the MLR model were $R^2$ = 0.940–0.955 (training dataset) and 0.902–1.076 (validation dataset); for the PLSR model $R^2$ = 0.947–0.993 (training dataset) and $R^2$ = 0.946–1.096 (validation dataset); and for PCR $R^2$ = 0.889–0.930 (training dataset) and $R^2$ = 0.910-1.008 (validation dataset). However, MLR models could not predict low fumonisin content below 5 mg/kg, which is the maximum acceptable level of fumonisins in maize as defined by the FDA.

Using FTIR spectra and chemometric analysis, Kos et al. (2016) classified maize based on deoxynivalenol content. The EU regulatory limit for deoxynivalenol in unprocessed maize is 1.75 mg/kg. The PCA method showed overlapping between maize samples with lower and higher levels of deoxynivalenol contamination relative to this limit. A new chemometric approach, the bootstrap-aggregated (bagged) decision tree, was also tested. This method was able to classify 79% of the maize samples at the EU regulatory limit for deoxynivalenol. The model also revealed a relationship between varieties of maize and types of fungal infection. This may explain the low classification accuracy. Employing two maize varieties reduced classification accuracy from 90% to 79%. Removing samples infected with *Fusarium verticillioides* from those infected with *Fusarium graminearum* and *Fusarium culmorum* improved classification accuracy from 73% to 79%.

Grains can be contaminated with aflatoxins. Ge et al. (2016) performed quantitative determination of aflatoxin in grains using terahertz spectroscopy and chemometrics (PLSR, PCR, SVM, and PCA-SVM). The results indicate that linear regression models (PLSR and PCR) are more accurate for lower aflatoxin concentrations (1–50 $\mu$g/ml) ($R^2$ = 0.983 for PLSR; $R^2$ = 0.982 for PCR; $R^2$ = 0.963 for SVM; $R^2$ = 0.947 for PCA-SVM), whereas nonlinear models (SVM and PCA-SVM) are more accurate for higher aflatoxin concentrations (1–50 $\mu$g/l) (85% for SVM; 94% for PCA-SVM; 50% for PCA; 35% for PLS).

Saidi and Mirzaei (2016) compared the classical HPLC method with the spectrofluorimetric method in combination with artificial neural networks to determine ochrotaxin content in wheat and rice products. The $R^2$ coefficients were 0.995 for the training and prediction set and 0.991 for the test set. The results of their study show that spectrofluorimetric spectra matrices and ANN can successfully be applied for the determination of ochratoxin. Moreover, this method is less time consuming and simpler than conventional HPLC.

Wiliams et al. (2012) investigated the use of NIR-HSI and multivariate analysis to monitor changes in maize kernels infected with *Fusarium verticillioides*. Germ-up and Germ-down samples were divided into two major groups: kernels to be sterilized and kernels without sterilization. PCA revealed the presence of three clusters, discriminating control samples, samples inoculated 17 h, and samples inoculated for 20–90 h. However, due to significant differences, it was impossible to build a single PLS model for all treatments and the PLS models were evaluated individually. The data matrix was composed of 32 objects. However, in most cases greater accuracy for predicting the degree of fungal infection was obtained with a lower number of objects (8 of 32). In all treatments $R^2$ = 0.98. However, when 32 objects were considered, only the sterilized germ-up sample revealed an identical $R^2$ = 0.98 while for the others $R^2$ = 0.83–0.92. It is therefore possible to build PLS models to predict the degree of contamination in kernels, but the number of spectra should be limited and samples should be quite similar to each other.

## Chemometrics in quality evaluation of olives and olive oils

Olive oils are divided into two groups: virgin olive oil (VOO) and extra-virgin olive oil (EVOO). EVOO is considered higher

quality due to its health and sensory aspects, as well as because of its oxidative stability. These characteristics are related to the fatty acid composition of EVOO and to the presence of several minor compounds, such as volatile compounds, polyphenols, tocopherols, and squalene. Given that high quality entails high price, EVOO is often adulterated by the addition of cheaper oils: seed, corn, soybaean, sunflower and nut oils, refined or residue olive oils, or synthetic olive oil-glycerol products. Moreover, the properties of olive oils are influenced by the conditions of cultivation, harvesting, and technological processing (Arvanitoyannis and Vlachos, 2007; Gómez-Caravaca et al., 2016).

Gurdeniz and Ozen (2009) used Mid-IR spectroscopy and chemometrics to detect EVOO adulterated with vegetable oils (rapeseed, cottonseed, and corn-sunflower binary mixture). PCA separated authentic from adulterated EVOO. However, PCA was not able to group samples with low levels of adulterants (2% and 5%), and these samples were excluded. PLS-DA correctly classified most samples but several could not be classified as either pure or adulterated. However, a Cooman's plot based on the SIMCA model was able to classify these nonmembers, making the PLS-DA model more reliable. Quantification of adulterated oil was performed using a PLSR model. The model showed high predictive ability for all types of adulterants (97.9–99.1%) with an $R^2$ coefficient of around 0.99 for the calibration set and an $R^2$ coefficient of between 0.93 and 0.98 for the validation set. However, samples with adulterant levels of 2% were excluded. Rohman et al. (2014) employed FTIR spectroscopy to determine the authenticity of EVOO. DA was used to differentiate pure EVOO from EVOO adulterated with canola oil. Only one unadulterated sample was misclassified. PLSR and PCR were used to quantify canola oil in EVOO. PLSR ($R^2 = 0.999$) revealed slightly better quantification ability than PCR ($R^2 = 0.901$–$0.998$). For PLSR the validation set $R^2$ coefficient was 0.997.

According to the Commission Implementing Regulation, 2013 EU No. 1335/13, labels must include information on the geographical origin of olive oils. Melucci et al. (2016) studied the use of flash gas chromatography electronic nose and chemometrics to verify that the geographical origins of olive oils were the same as those given on the labels. PCA separated 100% Italian EVOO and non-100% Italian EVOO. Some samples labeled non-Italian were located in the centroid, because they may have contained a small amount of Italian EVOO. A PLS-DA model provided good discrimination between 100% Italian and non-100% Italian olive oils, with determination coefficients of 0.833 and 0.834 for the calibration and validation sets, respectively.

Gouvinhas et al. (2015) differentiated between EVOO obtained from three different olive cultivars (Olive cv, Cobrançosa, and Galega) at several stages of maturation (green, semi-ripe, and ripe) using FTIR spectroscopy and chemometrics. FDA provided 100% correct classification for the calibration set and lower accuracy (73.6%) for the cross-validation set. The highest classification accuracy was observed for ripe EVOO (87.5%) and the lowest for semi-ripe EVOO (62.5%). A PLSR model was used to predict the contents of total phenolics, o-diphenols, and flavonoids, as well as antioxidant activity based on FTIR spectra. The lowest $R^2$ coefficient was observed for antioxidant activity (0.93 for the calibration set and 0.88 for the validation set), whereas the highest $R^2$ coefficients were observed for o-diphenols ($R^2 = 0.99$) and flavonoids ($R^2 = 0.99$ - calibration set; $R^2$ 0.98 - validation set).

The cultivars of olive trees determine the quality of olive oils. Aouidi et al. (2012) analyzed the olive leaves of five Tunisian cultivars using FT-MIR and chemometrics. PCA did not discriminate all varieties. Overlapping between Chemlali and Chétoui leaves was observed. A PLS-DA classification model was constructed, with high $R^2 = 0.96$–$0.98$. Chétoui, Zarrazi and Meski cultivars were classified with 100% accuracy, whereas only 80% and 40% of Chemlali and Sayali cultivars were classified correctly.

### Chemometrics in quality evaluation of coffee

The most important species of coffee plant are *Coffea arabica* and *Coffea canephora*. These are the source of beans commonly known as Arabica and Robusta. Arabica beans are mild and aromatic, while Robusta beans are harsh and bitter and contain more caffeine. Robusta beans are easier to harvest. Arabica beans are cultivated on mountains; Robusta plantations are situated at lower altitudes. Arabica beans are 40-50% more expensive than Robusta beans. There is interest, therefore, in developing chemometric methods to distinguish between these two varieties of coffee. According to the literature, it is possible to identify coffees based on their contents of volatile compounds (Huanga et al., 2007; Ribeiro et al., 2009, 2012), polyphenols (Ribeiro et al., 2012), amino acids and caffeine (Martín et al., 1998).

Martín et al. (1998) differentiated Arabica and Robusta green coffee varieties based on their contents of chlorogenic acid, caffeine, trigelline, amino acids, aqueous extract, and polyphenols. PCA did not precisely distinguish between the varieties of coffee. However, chlorogenic acid, caffeine and polyphenols were found to be good markers. Better separation was achieved using CA. KNN analysis grouped samples with 97.6% accuracy ($K = 1$–$2$). When $K = 3$–$5$, classification accuracy decreased to 92.7%. This paper presents the use of the nonparametric classification procedure KNN. Use of KNN is justified because the variables do not have a normal distribution. A common mistake is to apply techniques designed for normal distributions without having first verified whether the distribution is normal.

Recent studies report separation of Robusta and Arabica green coffee varieties using PCA according to lipophilic and aqueous extracts. Analysis of $^1$H NMR spectra further revealed 16-O-methylcafestol (16-OMC) and kahweol as important markers for Robusta coffee and Arabica coffee, respectively (Monakhova et al., 2015). Robusta and Arabica beans may be differentiated irrespective of the degree of roasting. PLS-DA has been reported to provide around 98% classification accuracy (100% for Arabica and 95% for Robusta). SIMCA also showed high sensitivity and specificity (93% and 96% for Arabica and 77% and 96% for Robusta) (De Luca et al., 2016).

Coffee can be fraudulently mixed with less expensive materials, such as spent coffee grounds, coffee husks and other roasted grains. Coffee is generally considered impure when the adulteration level is more than 10 g/100 g coffee. Reis et al. (2013)

used FTIR spectrum data and employed PCA to differentiate pure coffee from coffee contaminated with husks, corn, barley and spent coffee grounds. Overlapping was observed between roasted corn and barley, probably due to similar starch content. LDA was also performed and 100% separation of coffees achieved. PLSR can also be used to build a model for predicting the content of barley in coffee (for the test RMSE = 1.4% for external set RMSE = 0.8%) (Ebrahimi-Najafabadi et al., 2012). On the basis of the results of FTIR-ATR spectroscopy, a PLSR model successfully predicted adulteration levels of between 0,5% and 40% with high $R^2$ coefficient ($R^2$ = 0.99 for calibration and validation set) and low error value (RMSEC = 0.69%, RMSEP = 2.00%) (Reis et al., 2016).

The flavor of coffee is determined by volatile compounds. These are usually identified using GC-MS. However, due to their number and complexity accurate separation is very difficult. Coffee producers enhance the flavor of their products using various additives, such as coconut flavoring, and this misleads consumers. Huanga et al. (2007) analyzed the chemical composition of coffee flavor and coffee enriched with coconut aroma and classified coffees using PCA. Coffee samples from the same factories were shown to have the same compositions of volatile compounds, with similarity greater than 0.9104. Other studies report the application of chemometrics for predicting the sensory characteristics of Arabica coffee. Volatile compounds were determined using SPME-GC. Based on the chromatogram, PLS models were constructed (Ribeiro et al., 2009; Ribeiro et al., 2012). SPME-GC was found to be an effective method for predicting the sensory characteristics of coffee, including acidity, bitterness, aroma, purity, body and overall quality. The RMSEP values were 0.27 (acidity), 0.33 (bitterness and flavor), 0.41 (cleanliness), 0.34 (body), and 0.35 (overall quality). The results show a correlation between evaluations made by a sensory panel and the presence of volatile compounds in Arabica coffee beans (Ribeiro et al., 2012).

### Chemometrics in quality evaluation of teas

There are five basic types of tea: green, white, black, red (Pu-erh), and blue-green (cyan, Oolong). White tea is composed of buds and young leaves, which, over a relatively short period of plant growth, are protected from light to prevent the chlorophyll synthesis. During the withering stage, there is a delicate process of fermentation and drying. Green tea is obtained from older leaves, which are dried immediately after harvesting, in order to prevent the spontaneous fermentation process. As a result, the leaves retain their green color and properties similar to fresh tea. Turquoise tea is also obtained from older leaves, but the fermentation process is controlled. Depending on the degree of fermentation, the leaves may be pale green, brown, red or black. Black tea is made in a relatively long process of enzymatic fermentation. In the case of red tea, the dried leaves are fermented again by microorganisms, changing its chemical composition.

Teas are known to differ in terms of their content of amino acids (Alcàzar et al., 2007; McKenzie et al., 2010), sugars (Seetohul et al., 2006), minerals (Fernández-Cáceres, et al., 2001; Moreda-Pineiro et al., 2003; McKenzie et al., 2010), volatile compounds (Seetohul et al., 2006) and polyphenols (catechins) (Seetohul et al., 2006). The chemical composition of tea affects the flavor of the brew (Seetohul et al., 2006). There is also a correlation between the composition of tea and its country of origin (Fernández-Cáceres, et al., 2001; Moreda-Pineiro et al., 2003). A further relationship has been found between the content of individual amino acids and sugars and the quality of teas (Ding et al., 2002).

McKenzie et al. (2010) analyzed the concentration of minerals (Al, Ba, Ca, Cu, Fe, Mg, Mn, Ni, P, K, Na, Sr, Zn, and S) in five types of tea. It was found that white tea can be identified based on its content of P, Sr, Zn, and Al, and red tea mainly on the basis of its Na and Mg content. Black tea differs from white, red and turquoise tea in terms of Mg and Zn content, and from green tea in terms of K content. LDA was performed based on the concentrations of these markers. Overall classification ability was 81% and the best separation was obtained for Oolong teas (100%) ant the lowest for black and green teas (64%). The other types of tea were not precisely separated. Better results were obtained using PNN. Overall classification ability was 97%. PNN successfully separated white, green, Oolong, and Pu-erh teas with 100% accuracy and black teas with 96% accuracy. PNN is a type of ANN. However, it requires less training time because the smoothing factor is the only control parameter which needs to be optimized. The network must be updated for each new training set, but the smoothing factor does not require adjustment. Another advantage is that the network structure is selected automatically, making the process easier.

Attempts have also been made to apply chemometric methods for the classification of teas based on concentrations of amino acids. Accurate discrimination of teas was not achieved using PCA, with only white and green teas constituting separated groups. However, white and red teas have been distinguished using LDA with recognition ability of 100% for white, green, and Pu-erh teas, 90.48% for black teas and 83.33% for Oolong teas. KNN provided recognition ability of 100% for all teas except black (66.66%). Using ANN (BP-MLP) 100% classification ability was achieved. ANN is commonly used to solve nonlinear problems, and in this case it is possible that the data structure was nonlinear (Alcàzar et al., 2007). Fernández-Caceres et al. (2001) similarly report that PCA failed to group black and green teas, while LDA enabled correct identification with recognition ability of 97.8% and prediction ability of 93.5%. ANN provided better recognition ability of 100% and prediction ability of 95.6%.

PCA has been employed to group teas from different countries in terms of their chemical composition. In a group of 18 African teas, 4 Asian teas were included, and 6 African teas were included among 36 Asian teas. Similar results were obtained using CA. Accuracy was lower using SIMCA, with only 83.3% and 88.9% African and Asian teas classified correctly. Better results were obtained with LDA 100% of African and 97.2% of Asian teas were correctly classified. Only one Asian tea was incorrectly classified (Moreda-Pineiro et al., 2003).

Identification of teas (Oolong, green, Houji, Kenya, Assam, Ceylon, and Japanese black) can be performed based on their content of polyphenols. PCA revealed that Assam and Ceylon teas were very close in the variance space, reflecting their close geographical origins. Kenyian, Japanese and black teas were

distant from each other, because they are cultivated in regions with very different climates (Seetohul et al., 2006). Recent studies report the use of HPLC coupled with chemometrics for quality evaluation of Deepure instant pu-erh tea, produced by different manufacturers. Differences were observed in terms of the concentration of catechin derivatives. Chemometric tools (HCA, PCA, and PLS-DA) gave satisfactory results for the differentiation of teas according to variety and manufacturer (Wang et al., 2016).

Adulteration practices are common in the production of tea. Some producers add talcum powder, which makes tea more attractive and hides quality defects. Studies suggest a link between the use of talc and cancer (Neil et al., 2012). Li et al. (2016) report satisfactory results using chemometric methods for the detection of tea adulterated with talcum powder. Using PCA, pure tea samples were clearly separated from those adulterated with talc. However, PCA was not effective at separating tea samples with different doses of talcum powder.

### Chemometrics in quality evaluation of alcoholic drinks

The composition of wine, and therefore also its taste and flavor, are affected by number of factors, such as grape variety, climate and soil conditions, method of rape cultivation and production methods. The region of origin is therefore extremely important. Chemometric techniques are often used to differentiate wines, particularly in terms of their origin.

González-Centeno et al. (2015) used PLS-DA to group Spanish wines (2009 and 2010 vintages) from two wine-producing regions: Binissalem and Pla i Llevant. Discriminant analysis was conducted based on differences in soil composition (content of Pb, clay, sand, and carbonate), climatic factors (rainfall, maximum, and minimum air temperature) and the location of the vineyard. For each of the wine vintages, the researchers reported a correct classification rate of 95%.

Römisch et al. (2009) used RDA and PLS-DA methods to analyze the chemical composition of Hungarian, Czech and Romanian wines, and to divide red and white wines into groups depending on their countries of origin. Correct classification rates of 88 and 100% were achieved. Capron et al. (2007) classified authentic and commercial wines, also from Hungary, the Czech Republic and Romania, using PLS-DA and SVM. Differentiation was more difficult in the case of commercial and authentic white wines than with authentic and commercial red wines. Difficulties in classifying commercial wines are probably due to the influence of technological processes on their chemical composition. Use of SVM was more effective than PLS-DA and classified wines slightly better according to a unique set of variables. (Capron et al., 2007). SVM is used to solve nonlinear problems, whereas PLS-DA is used for linear problems. SVM is based on the kernel function. However, to build a suitable model variable selection must be performed using other methods.

White and red wines originating from different wine-growing regions can also be classified and characterized according to their contents of alcohol, saccharides, amino acids, organic acids, volatile compounds, polyphenols and tracers. Ragone et al. (2015) successfully analyzed white and red wines using PCA and PLS-DA. PLS-DA as supervised technique improved separation of white and red wines. The wines differed mainly in terms of their contents of organic acids (lactic acid, succinic acid, acetic acid, malic and citric acid), amino acids (arginine, alanine, leucine and isoleucine) and isopenthanol. Other studies report the use of PCA for classifying red wines (2007 and 2008 vintages), based on the antioxidant activity, acidity, saccharides, phenolic compounds, including flavonoids and tannins. Of the parameters investigated, the best marker was tannin content Using PCA, it was possible to explain 81.36% of the total variance, but PCA did not separate wines of different vintages. This suggests that PCA is not always an appropriate technique for evaluating similarities/dissimilarities between samples. (Lima et al., 2011).

The levels of volatile compounds in wines produced using yeast depend on the variety of grape. Grape varieties differ in terms of amino acid content, sugar and water. The content of volatile metabolites may also affect the sensory characteristics of wine. Using a PLSR model, Hernández-Orte et al. (2002) found a relationship between content of different amino acids in wines and the content of volatile compounds. These results show the effect of different grape varieties on the profile of volatile compounds in wine. However, it should be mentioned that the regression coefficient was quite low for most volatile compounds (0.49–0.96), with the highest observed for methionol ($R_v = 0.96$). Based on analysis of various elements (Zn, Sr, Pb, In, Cu, Ni, As, Cd, and P), 100% classification accuracy was achieved for sparkling wines (cava and champagne) using LDA and SIMCA chemometric techniques (Jos et al., 2004).

Chemometrics may be used for the classification not only of wines but also of different alcoholic beverages. On the basis of NIR spectrum analysis in combination with chemometric methods (PCA, SIMCA), Pontes et al. (2006) classified various types of alcoholic beverage (whiskey, brandy, rum, and vodka) with 100% efficiency. They were able to distinguish classes of alcoholic beverage, adulterated by the addition of water, methanol or ethanol.

Efforts have been made to differentiate between types of beer (lager, dark, and low-alcohol) using PCA based on mineral content (Zn, P, B, Mn, Fe, Mg, Al, Sr, Ca, Ba, Na, and K). Much better separation was achieved using LDA. Beers were grouped with recognition ability of 94% (Alcázar et al., 2002). Lager beers sold under the same brand but

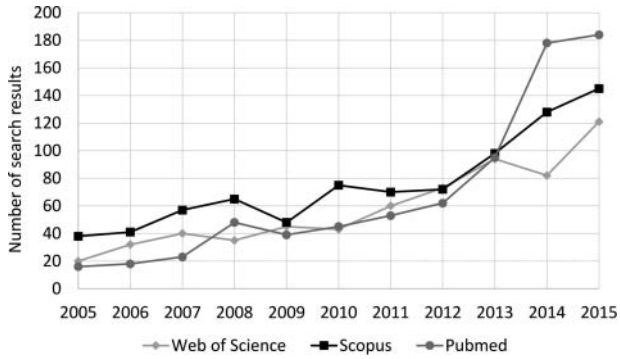


Figure 1. Number of search results for articles related to chemometrics and food on Web of Science, Scopus and PubMed published between 2005 and 2015.

**Table 2.** Applications of chemometrics for the evaluation of food and beverages quality.

| Group of food | Food product | Analytical measurement | Instrumental method | Chemometric method | Application of chemometrics | Reference |
|---|---|---|---|---|---|---|
| **Food of animal origin** | | | | | | |
| Meat and meat products | Salami products made of cattle, sheep, pig, fish, poultry, goat and buffalo meat | Fat content | Raman spectroscopy | PCA | Adulteration detection | Boyaci et al. (2014) |
| | Wild boar and beef meatballs | Fat content | FTIR | PCA<br>PLSR | Adulteration detection and authentication | Guntarti et al. (2015) |
| | Frankfurters | Proteins, carbohydrates and fat content | FR-Raman spectroscopy | PLS-DA | Adulteration detection | Campos et al. (2014) |
| | Ground beef | Metabolite profile e.g. amino acids, alcohols, amines, organic acids | $^{1}$H NMR | PCA | Irradiation control | Zanardi et al. (2015) |
| | Minced pork meat | Spectra, microbial load and sensory analysis | FTIR | PLSR<br>PLS-DA | Quality control | Papadopoulou, et al. (2011) |
| | Minced beef | Spectra, microbial load, sensory analysis, pH | FTIR | PLSR<br>PCA<br>FDA | Quality control | Ammor et al. (2009) |
| | Pork meat | Spectra and microbial load | HSI | LS-SVM | Quality control | Wang et al. (2011) |
| | Pork meat | Spectra and microbial load | HSI | PCA<br>BP-ANN<br>SI-PLS | Quality control | Huang et al. (2013) |
| | Chicken meat | Spectra and microbial load | SW-NIRS | PCA<br>PLSR | Quality control | Lin et al. (2004) |
| | Beef fillets | Spectra, microbial count and sensory analysis | FTIR | MLP neural networks (ANN) | Quality control | Argyri et al. (2010) |
| | Dry sausages | Biochemical activities | — | PCA | Identification of microorganisms | Montel et al. (1996) |
| | Dry sausages | Biogenic amines, microbial load, pH, $a_w$ | HPLC | PCA | | Parente et al. (2001) |
| | Pork meat | Spectra and BAI | HSI<br>HPLC | PLSR<br>RC-PLSR<br>SPA-PLSR<br>MLR<br>RC-MLR<br>SPA-MLR | Quality control | Cheng et al. (2016) |
| Fish and seafood products | Shrimps | Stable isotope ratio and multi-element analysis | IRMS<br>ICP-MS<br>ICP-OES | PCA<br>DA<br>CA | Authentication | Ortea and Gallardo (2015) |
| | European sea bass | Spectra and chemical properties (fatty acids, bromatological, isotopes) | NIRS | PLS-DA<br>WPTER | Authentication | Ottavian et al. (2012) |
| | Turbot | Stable isotope ratio and fatty acids | IRMS | PCA<br>SIMCA<br>LDA | Authentication | Busetto, et al., 2008 |
| | Mussels | trace elements | ICP-MS | LDA<br>SIMCA<br>ANN | Authentication | Costas-Rodríguez et al. (2010) |
| | Shrimps | spectra | EEMs | SIMCA | Authentication | Eaton et al. (2012) |
| | Rainbow trout fillets | Spectra and microbial load | SW-NIRS | PCA<br>PLSR | Quality control | Lin et al. (2006) |
| | Cold-smoked salmon | Volatile compounds | GC-MS | PCA | Identification of microorganisms | Joffraud et al. (2001) |
| | Reference material: DORM-1 (dogfish flesh) and DOLT-1 (dogfish liver) | Trace elements | FAAS<br>ETAAS | PCA | Quality control | Bermejo-Barrera et al. (2001) |
| | Kutum, common carp and Caspian carp | Biogenic amines | HPLC | PCA | Quality control and differentiation | Aflaki et al. (2016) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Milk and dairy products | Brazilian UHT milk | Physicochemial parameters (starch, chlorine, formol, hydrogen peroxide and urine) | — | PCA<br>HCA | Authentication and adulteration detection | Souza et al. (2011) |
| | Milk | Spectra | MIR-microspectroscopy | SIMCA<br>PLSR | Adulteration detection | Santos et al. (2013) |
| | Milk | Spectra | NIRS | UVE-PLS<br>PLS-DA<br>SIMCA | Adulteration detection | Wu et al. (2016) |
| | Milk | Spectra, microbial load, pH | FTIR | PLSR | Quality control | Nicolaou and Goodacre (2008) |
| | Milk, milk powder, mild Cheddar cheese, grated cheese powder, yogurt, butter | Spectra and cholesterol content | FTIR | PLSR<br>PCR | Determination of nutrition quality | Paradkar and Irudayaraj (2002) |
| | Yogurts | pH, color, firmness | — | PCA<br>HCA<br>PLS-DA<br>KNN<br>SIMCA | Determination of nutrition quality | Cruz et al. (2013) |
| | Raw cow milk | Protein, fat, minerals and trace elements content | — | PCA<br>LDA | Differentiation | Sola-Larrañaga and Navarro-Blasco (2009) |
| | Cheeses: Edamer, Gouda, Feta, and Emmentaler, ice creams | Fatty acids | $^{1}$H NMR<br>$^{13}$C NMR | PCA<br>PLSR | Adulteration detection and differentiation/ authentication/ | Monakhova et al. (2013) |
| | Spreadable cheese | Spectra and starch content | Raman spectroscopy | PLS-DA<br>PLSR | Adulteration detection | De Sá Oliveira et al. (2016) |
| | Oscypek | Volatile compounds | GC-MS | PCA<br>LDA<br>SIMCA<br>SVM | Adulteration detection and authentication | Majcher et al. (2015) |
| **Beverages and food of plant origin** | | | | | | |
| Fruits and fruit products | Fruit juices made of orange, grapefruit, mandarin, and pomelo and commercial juices | Polyphenolic componds | UPLC-QToF MS | PCA<br>SIMCA<br>PLS-DA | Adulteration detection and differentiation | Jandrić and Cannavan (2015) |
| | Sweet orange, tangerine, lemon, and grapefruit juices | Polyphenolic componds | HPLC-MS | PCA<br>CA<br>LDA<br>PCR<br>PLSR | Differentiation | Abad-García et al. (2012) |
| | Apple juices and fermented apple juices | physicochemical and aroma profile | GC | PCA<br>HCA | Differentiatiation | Braga et al. (2013) |
| | Jujube fruits | Spectra and total sugar, total acid, total phenolic content and antioxidant activity | NIRS | PCA<br>BP-ANN<br>PLSR<br>LS-SVM<br>LDA | Differentiation and characterization | Guo et al. (2016) |
| | Jujube fruits | triterpenoid acid content | HPLC | PCA<br>HCA | Differentiation | Guo et al. (2009) |
| | Brazilian frozen pulps from: acai, blackberry, caja, cashew, cocoa, coconut, grape, graviola, guava, papaya, peach, pineapple, pineapple and mint, red fruits, seriguela, strawberry, tamarind, umbu, yellow passion fruit | Chromaticity, phenolic compounds, carotenoids and antioxidant activity | — | PCA<br>HCA | Differentiation | Zielinski et al. (2014[a]) |

(*Continued on next page*)

**Table 2.** (*Continued*)

| Group of food | Food product | Analytical measurement | Instrumental method | Chemometric method | Application of chemometrics | Reference |
|---|---|---|---|---|---|---|
| Honey | Acacia linden, chestnut, fir, spruce, floral and forest honey | Flavonoids and abscisic acid | LC-MS | PCA<br>LDA | Authentication | Bertoncelj et al. (2011) |
| | Chest and rape honeys | Polyphenolic compounds | HPLC—MS | PCA<br>PLS<br>PLS-DA<br>SIMCA | Authentication | Zhou et al. (2014) |
| | Nectar honeys and honeydew honeys | Aromatic compounds | GC-MS | PCA | Authentication | Daher et al. (2008) |
| | Croatian honeys (black locust, lime, sage, chestnut and honeydew) | antioxidant capacity and physicochemical properties (moisture, electrical conductivity, HMF content, colour, phenolic content) | — | PCA | Authentication | Flanjak et al. (2016) |
| | Honeydew, buckwheat and rape honeys | Major and trace elements | ICP-MS | CA<br>PCA<br>LDA<br>CART | Authentication | Chudzinska and Baralkiewicz (2011) |
| | Honey varieties from citrus, chestnut, sunflower, honeydew, robinia, rhododendron and linden tree | Volatile compounds | PTR-ToF-MS | PCA<br>LDA<br>PLS-DA<br>PNN<br>SIMCA<br>SOMs | Authentication | Schuhfried et al. (2016) |
| | Flower and pine honeys | Heavy metals | | PCA<br>CA | Quality control | Rodríguez García et al. (2006 |
| | Flower and pine honeys | Spectra and flucose, sucrose, maltose content | Raman spectroscopy<br>HPLC | PCA<br>PLS<br>ANN | Differentiation and quality control | Özbalci et al. (2013) |
| Crops | Maize | Spectra and fumonisin content | NIRS<br>HPLC | PLSR | Quality control | Giacomo and Stefania (2013) |
| | Maize | Spectra and fumonisin content | SERS | KNN<br>PLS-DA<br>LDA<br>PLSR<br>MLR<br>PCR | Quality control | Lee and Herrman (2016) |
| | Maize | Spectra and deoxynivalenol and aflatoxin B1 content | FTIR | PCA<br>bootstrap-aggregated (bagged) decision tre | Quality control | Kos et al. (2016) |
| | Grains | Spectra and aflatoxin B1 content | terahertz spectroscopy | PLSR<br>PCR<br>SVM<br>PCA-SVM | Quality control | Ge et al. (2016) |
| | Wheat and rice | Spectra and ochratoxin content | HPLC<br>spectrofluorimetry | ANN | Quality control | Saidi and Mirzaei (2016) |
| | Maize kernels | Spectra | NIR-HSI | PLSR<br>PCA | Quality control | Williams et al. (2012) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Olives and olive oils | Unadulterated extra-virgin olive oils and oils adulterated with rapeseed, cottonseed and corn—sunflower binary mixture | Spectra | Mid-IR | PCA<br>PLS-DA<br>PLSR | Adulteration detection and authentication | Gurdeniz and Ozen (2009) |
| | Extra virgin olive oil, canola oil, grape seed oil, rice bran oil, and walnut oil | Spectra and fatty acids | FTIR | DA<br>PLSR<br>PCR | Adulteration detection and authentication | Rohman et al. (2014) |
| | Italian and non-Italian extra virgin olive oils | Volatile compounds | FGC E-nose | PCA<br>PLS-DA | Adulteration detection and authentication | Melucci et al. (2016) |
| | Extra virgin olive oils of Cobrançosa, Galega and Picual olive cultivar od different ripening stages | Spectra and total phenolics, o-diphenols, flavonoids antioxidant activity | FTIR | PCA<br>FDA<br>PLSR | Quality control | Gouvinhas et al. (2015) |
| | Olive leaves of five Tunisian cultivars: Chétoui, Zarrazi, Meski, Chemlali and Sayali | Spectra | FT-MIR | PCA<br>PLSR | Differentiation | Aouidi et al. (2012) |
| Coffee | Arabica and robusta green coffee | Chlorogenic acid, caffeine, trigonelline, aqueous extract, amino acids and polyphenols | — | PCA<br>CA<br>KNN | Authentication | Martín et al. (1998) |
| | Arabica and robusta coffee | spectra | $^{1}$H NMR | PCA | Authentication | Monakhova et al. (2015) |
| | Arabica and Robusta coffee | Spectra, and polyphenols, methylxantines | NIRS<br>HPLC | PLS-DA<br>SIMCA | Authentication | De Luca et al. (2016) |
| | Roasted and ground coffee | spectra | FTIR-ATR | PLS | Adulteration detection | Reis et al. (2016); |
| | Roasted and ground coffee | spectra | DRIFTS | PCA<br>LDA | Adulteration detection | Reis et al. (2013) |
| | Arabica, Robusta coffee and their mixtures | spectra | NIRS | PLS | Adulteration detection | Ebrahimi-Najafabadi et al. (2012) |
| | Coffee flavor | Volatile compounds | GC-MS | PCA | Quality control | Huanga et al. (2007) |
| | Brazilian arabica roasted coffees | Volatile compounds | GC<br>GC-MS | PLS-DA | Quality control | Ribeiro et al. (2009) |
| | Arabica coffee | Volatile compounds | GC<br>GC-MS | PLS | Quality control | Ribeiro et al. (2012) |
| Tea | Deepure instant pu-erh tea | catechin derivatives and caffeine | HPLC | PCA<br>PLS-DA<br>HCA | Quality control | Wang et al. (2016) |
| | White, green, black, Oolong and Pu-erh teas | Minerals content | ICP-AES | LDA<br>PNN | Differentiation/ Authentication | McKenzie et al. (2010) |
| | Green, black, Oolong, white, and Pu-erh teas | Amino acid contents | HPLC | PCA<br>KNN<br>LDA<br>ANN | Differentiation/ Authentication | Alcàzar et al. (2007) |
| | Green, black, and instant teas | Metal content | ICP-AES | PCA<br>LDA<br>ANN (BP-MLP) | Differentiation/ Authentication | Fernández-Caceres et al. (2001) |
| | Bottled liquid Japanese teas (oolong, green, houji and black teas) and leaf teas (Kenya, Assam and Ceylon teas) | spectra | TLS | PCA | Differentiation/ Authentication | Seetohul et al. (2006) |
| | Asian and African teas | Trace metals | ICP-MS<br>ICP-AES | PCA<br>CA<br>LDA<br>SIMCA | Differentiation/ Authentication | Moreda-Pineiro et al. (2003) |
| | LongJing teas | Spectra and talcum powder | FTIR | PCA<br>PLS | Adulteration detection | Li et al. (2016) |

(*Continued on next page*)

**Table 2.** (*Continued*)

| Group of food | Food product | Analytical measurement | Instrumental method | Chemometric method | Application of chemometrics | Reference |
|---|---|---|---|---|---|---|
| Alcoholic drinks | Spanish *Merlot* and *Cabernet Sauvignon* wines | soil composition climatic factors, location of the vineyard | — | PLS-DA | Authentication | González-Centeno et al. (2015) |
| | Hungarian, Czech and Romanian wines | Chemical composition (isotopic ratio, trace elements • macro elements, alcohols, organic acids, • biogenic amines) | Unknown data | CART<br>RDA<br>PLS-DA | Authentication | Römisch et al. (2009) |
| | Authentic and commercial wines | Chemical composition (isotopic ratio, trace elements • macro elements, alcohols, acids, • biogenic amines) | Unknown data | PLS-DA<br>SVM | Authentication | Capron et al. (2007) |
| | Red and white wines | amino acids, organic acid, alcohol, sugar, polyphenols contents. | $^{1}$H NMR<br>HPLC-MS | PCA<br>PLS-DA | Differentiation | Ragone et al. (2015) |
| | Red wines of different vintahes | Polyphenols content and antioxidant activity | HPLC | PCA | Authentication | Lima et al. (2011) |
| | Synthetic wines of different grape varieties Chardonnay; Cabernet Sauvignon; Grenache; Macabeo; Merlot; Moristel; Carignane, Pinot Noir; Riesling; Sauvignon Blanc; Tempranillo | amino acids, sugar, water, and yeast nutrient, volatile compounds, | GC | CA<br>PLS | Authentication | Hernández-Orte et al. (2002) |
| | Sparkling wines (cava and champagne) | Metal content | ICP-AES | LDA<br>SIMCA | Authentication | Jos et al. (2004) |
| | Whiskey, brandy, rum and vodka | Spectra | NIRS | PCA<br>SIMCA | Adulteration detection | Pontes et al. (2006) |
| | Beers: dark, lager and with low alcoholic content | Mineral content | ICP-AES | PCA<br>LDA<br>BP-MLP (ANN) | Differentiation/ Authentication | Alcázar et al. (2002) |
| | Portuguese lager beers | Volatile compounds | GC-MS | PCA<br>AHC | Quality control | Rendall et al. (2015) |

produced in different breweries were analyzed using PCA and LDA. These beers were found to have different flavors and clearly differed in terms of their chemical composition (content of esters, alcohols, aldehydes, ketones, sulfur compounds, petroleum and diacetyl). Using PCA, the first two components explained only 37% of the total variance. However, when LDA was performed on the first seven principal components, it was able to explain 66% of the total variance and provided 89.6% classification accuracy (Vera et al., 2011). Rendall et al. (2015) studied changes in the content of volatile compounds including esters, higher alcohols and selected organic acids (caprylic acid, capric acid, and acetic acid) in Portuguese beers over one year of storage. After seven months, differences in the chemical composition of the beers were observed using PCA and HCA.

## Conclusion

As an alternative to standard statistical methods, chemometrics offer effective tools for food quality analysis and control, with fast and inexpensive analysis of experimental results. According to Web of Science, Scopus, and PubMed, in recent years there has been increasing research interest into such applications of chemometrics (Fig. 1).

Chemometrics can be used to analyze a wide range of food groups (of both animal and plant origin and beverages) and beverages. The main purposes of application of chemomerics are: authentication, adulteration detection, quality control, differentiation, and determination of nutrition quality (Table 2). The results presented in this review indicate that chemometric methods can be used for differentiation and classification of food products according to their geographical origin, botanical origin, production techniques, and chemical composition, as well as for the detection of adulterants in foods. Multivariate techniques can be successfully applied in food microbiology to evaluate microbial spoilage, identify microorganisms and detect mycotoxins.

Studies have shown the potential of using multivariate techniques in combination with chromatographic and spectroscopic techniques, enabling interpretation of large and complex data sets. Raman spectroscopy, FTIR, NIR, NMR, HPLC, GC, and ICP are commonly applied with chemometrics. In recent years, HSI has also become of great interest.

The most commonly used chemometric technique is PCA. PCA is often used for classification purposes, despite not being a classification technique. PCA provides information concerning data structure, explains similarities/dissimilarities between objects and may also indicate which variables contribute most to differentiation between objects. In many cases, PCA is the first step before further supervised analysis, which helps to select the most differentiating variables. The most frequently used classification techniques are LDA, PLS-DA, KNN, and SIMCA. However, in recent years there has been growing interest in artificial neural networks, which could provide promising classification models. ANN shows much better classification ability in comparison to other classification models. The most commonly used quantification model is PLSR, which is mostly applied in combination with spectroscopic techniques, especially FTIR. PLSR is used in quality control to determine levels of microbial spoilage, mycotoxin content or of adulterants. PLSR is currently the most accurate quantification model for general recognition and prediction purposes. However, its simplified optimized models also give encouraging results and thus need further research.

In conclusion, the implementation of chemometrics for research into food and beverages requires wide knowledge of their range of applications and limitations. Building a proper calibration model is a crucial step in many forms of chemometric analysis. Several models should also be considered before choosing the most suitable. Finally, the results should always be evaluated critically.

## References


Abad-García, B., Berrueta, L. A., Garmón-Lobato, S., Urkaregi, A., Gallo, B. and Vicente, F. (2012). Chemometric characterization of fruit juices from Spanish cultivars according to their phenolic compound contents: I. citrus fruits. *J. Agric. Food Chem.* **60**:3635–3644.

Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L., Petersen, M. A. and Nielsen, D. S. (2010). Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *J. Food Sci.* **75**:S300–S307.

Aflaki, F., Ghoulipour, V., Saemian, N., Shiebani, S. and Salahinejad, M. (2017). Chemometrics approaches to monitoring of biogenic amines changes in three fish species. *J. Aquat. Food Prod. T.* **26**:43–53.

Alcázar, A., Pablos, F., Martín, M. J. and González, A. G. (2002). Multivariate characterisation of beers according to their mineral content. *Talanta* **57**:45–52.

Alcàzar, A., Ballesteros, O., Jurado, J. M., Pablos, F., Martín, M. J., Vilches, J. L. and Naválon, A. (2007). Differentiation of green, white, black, oolong, and pu-erh teas according to their free amino acids content. *J. Agric. Food Chem.* **55**:5960–5965.

Ammor, M. S., Argyri, A. and Nychas, G. J. (2009). Rapid monitoring of the spoilage of minced beef stored under conventionally and active packaging conditions using Fourier transform infrared spectroscopy in tandem with chemometrics. *Meat Sci.* **81**:507–514.

Anklam, E. (1998). A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chem.* **63**:549–562.

Aouidi, F., Dupuy, N., Artaud, J., Roussos, S., Msallem, M., Perraud-Gaime, I. and Hamdi, M. (2012). Discrimination of five Tunisian cultivars by Mid InfraRed spectroscopy combined with chemometric analyses of olive Olea europaea leaves. *Food Chem.* **131**:360–366.

Argyri, A. A., Panagou, E. Z., Tarantilis, P. A., Polysiou, M. and Nychas, G. J. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *Sens. Actuat. B Chem.* **145**:146–154.

Arvanitoyannis, I. S. and van Houwelingen-Koukaliaroglou, M. (2003). Implementation of chemometrics for quality control and authentication of meat and meat products. *Crit. Rev. Food Sci. Nutr.* **43**:173–218.

Arvanitoyannis, I. S. and Vlachos, A. (2007). Implementation of physicochemical and sensory analysis in conjunction with multivariate analysis towards assessing olive oil authentication/adulteration. *Crit. Rev. Food Sci. Nutr.* **47**:441–498.

Balog, J., Perenyi, D., Guallar-Hoyas, C., Egri, A., Pringle, S. D., Stead, S., Chevallier, O. P., Elliott, C. T. and Takats, Z. (2016). Instantaneous identification of the species of origin for meat products by rapid evaporative ionization mass spectrometry. *J. Agric. Food Chem.* **64**:4793–4800.

Baş, D., Dudak, F. C. and Boyacı, I. H. (2007). Modeling and optimization III: Reaction rate estimation using artificial neural network (ANN) without a kinetic model. *J. Food Eng.* **79**:622–628.

Bermejo-Barrera, P., Moreda-Pineiro, A. and Bermejo-Barrera, A. (2001). Sample pre-treatment methods for the trace elements determination in seafood products by atomic absorption spectrometry. *Talanta* **57**:969–984.

Berrueta, L. A., Alonso-Salces, R. M. and Héberger, K. (2007). Supervised pattern recognition in food analysis. *J. Chromatogr. A* **1158**:196–214.

Bertoncelj, J., Polak, T., Kropf, U., Korošec, M. and Golob, T. (2011). LC-DAD-ESI/MS analysis of flavonoids and abscisic acid with chemometric approach for the classification of Slovenian honey. *Food Chem.* **127**:296–302.

Bota, G. M. and Harrington, P. B. (2006). Direct detection of trimethylamine in meat food products using ion mobility spectrometry. *Talanta* **68**:629–635.

Boyaci, I. H., Uysal, R. S., Temiz, T., Shendi, E. G., Yadegari, R. J., Rishkan, M. M, Velioglu, H. M., Tamer, U., Ozay, D. S. and Vural, H. (2014). A rapid method for determination of the origin of meat and meat products based on the extracted fat spectra by using of Raman spectroscopy and chemometric method. *Eur. Food Res. Technol.* **238**:845–852.

Braga, C. M., Zielinski, A. A. F., da Silva, K. M., de Souza, F. K. F., Pietrowski, G. D. A. M., Couto, M., Granato, D., Wosiacki, G. and Nogueira, A. (2013). Classification of juices and fermented beverages made from unripe, ripe and senescent apples based on the aromatic profile using chemometrics. *Food Chem*. **141**:967–974.

Busetto, M. L., Moretti, V. M., Moreno,-Rojas, J. M., Caprino, F., Giani, I., Malandra, R., Bellagamba, F. and Guillou, C. (2008). Authentication of farmed and wild turbot (*Psetta maxima*) by fatty acid and isotopic analyses combined with chemometrics. *J. Agric. Food Chem.* **56**:2742–2750.

Campos, N. S., Oliveira, K. S., Almeida, M. R., Stephani, R. and de Oliveira, L. F. C. (2014). Classification of frankfurters by FT-Raman spectroscopy and chemometric methods. *Molecules* **19**:18980–18992.

Capron, X., Massart, D. L. and Smeyers-Verbeke, J. (2007). Multivariate authentication of the geographical origin of wines: a kernel SVM approach. *Eur. Food Res. Technol.* **225**:559–568.

Cheng, W., Sun, D. W. and Cheng, J. H. (2016). Pork biogenic amine index (BAI) determination based on chemometric analysis of hyperspectral imaging data. *LWT-Food Sci. Technol.* **73**:13–19.

Chudzinska, M. and Baralkiewicz, D. (2011). Application of ICP-MS method of determination of 15 elements in honey with chemometric approach for the verification of their authenticity. *Food Chem. Toxicol.* **49**:2741–2749.

Costas-Rodríguez, M., Lavilla, I. and Bendicho, C. (2010). Classification of cultivated mussels from Galicia (Northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Anal. Chim. Acta* **664**:121–128.

Cruz, A. G., Cadena, R. S., Alvaro, M. B. V. B., Sant'Ana, A. S., Oliveira, C. A. F., Faria, J. A. F., Bolini, H. M. A. and Ferreira, M. M. C. (2013). Assessing the use of different chemometric techniques to discriminate low-fat and full-fat yogurts. *LWT-Food Sci. Technol.* **50**:210–214.

Daher, S. and Gülaçar, F. O. (2008). Analysis of phenolic and other aromatic compounds in honeys by solid-phase microextraction followed by gas chromatography - mass spectrometry. *J. Agric. Food Chem.* **56**:5775–5780.

De Luca, S., De Filippis, M., Bucci, R., Magrì, A. D., Magrì, A. L. and Marini, F. (2016). Characterization of the effects of different roasting conditions on coffee samples of different geographical origins by HPLC-DAD, NIR and chemometrics. *Microchem. J.* **129**:348–361.

de Sá Oliveira, K., de Souza Callegaro, L., Stephani, R., Almeida, M. R. and de Oliveira, L. F. C. (2016). Analysis of spreadable cheese by Raman spectroscopy and chemometric tools. *Food Chem*. **194**:441–446.

Ding, Y., Yu, H. and Mou, S. (2002). Direct determination of free amino acids and sugars in green tea by anion-exchange chromatography with integrated pulsed amperometric detection. *J. Chromatogr. A* **982**:237–244.

Eaton, J. K., Alcivar-Warren, A. and Kenny, J. E. (2012). Multidimensional fluorescence fingerprinting for classification of shrimp by location and species. *Environ. Sci. Technol.* **46**:2276–2282.

Ebrahimi-Najafabadi, H., Leardi, R., Oliveri, P., Chiara, M., Jalali-Heravi, M. and Lanteri, S. (2012). Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. *Talanta* **99**:175–179.

Ellis, D. I. and Goodacre, R. (2001). Rapid and quantitative detection of the microbial spoilage of muscle foods: current status and future trends. *Trends Food Sci. Technol.* **12**:414–424.

El-sheikh, A. H., Al-degs, Y. S., Abu-Wardeh, A. A. H. and Al-ghouti, M. A. (2017). Quantification of melamine in milk and dairy products by liquid chromatography after a simple sample clean-up procedure. *J. Food Process, Preserv.* **26**. e12867.

Elzey, B., Pollard, D. and Fakayode, S. O. (2016). Determination of adulterated neem and flaxseed oil compositions by FTIR spectroscopy and multivariate regression analysis. *Food Control* **68**:303–309.

Fernández-Cáceres, P. L., Martín, M. J., Pablos, F. and González, A. G. (2001). Differentiation of tea (*Camellia sinensis*) varieties and their geographical origin according to their metal content. *J. Agric. Food Chem.* **49**:4775–4779.

Flanjak, I., Kenjerić, D., Bubalo, D. and Primorac, L. (2016). Characterisation of selected Croatian honey types based on the combination of antioxidant capacity, quality parameters, and chemometrics. *Eur. Food Res. Technol.* **242**:467–475.

Ge, H., Jiang, Y., Lian, F., Zhan, G Y. and Xia, S. (2016). Quantitative determination of aflatoxin B1 concentration in acetonitrile by chemometric methods using terahertz spectroscopy. *Food Chem*. **209**:286–292.

Giacomo, D. R. and Stefania, D. Z. (2014). A multivariate regression model for detection of fumonisins content in maize from near infrared spectra. *Food Chem*. **141**:4289–4294.

González-Centeno, M. R., Adrover-Obrador, S., Simal, S., Frau, M. A., Femenia, A. and Rosselló, C. (2015). A multivariate methodology to distinguish among wine appellations of origin. *Agron. Sustain. Dev.* **35**:295–304.

Gouvinhas, I., De Almeida, J. M. M. M., Carvalho, T., Machado, N. and Barros, A. I. R. N. A. (2015). Discrimination and characterisation of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics. *Food Chem*. **174**:226–232.

Gómez-Caravaca, A. M., Gómez-Romero, M., Arráez-Román, D., Segura-Carretero, A. and Fernández-Gutiérrez, A. (2006). Advances in the analysis of phenolic compounds in products derived from bees. *J. Pharm. Biomed. Anal*. **41**:1220–1234.

Gómez-Caravaca, A. M., Maggio, R. M. and Cerretani, L. (2016). Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. *Anal. Chim. Acta.* **913**:1–21.

Guntarti, A., Sudibyo, M., Agustinus, Y. and Rohman, A. (2015). FTIR Spectroscopy in combination with chemometrics for analysis of wild boar meat in meatball formulation. *Asian J. Biochem.* **10**:165–172.

Guo, S., Duan, J. A., Tang, Y., Su, S., Shang, E., Ni, S. and Qian, D. (2009). High-performance liquid chromatography—Two wavelength detection of triterpenoid acids from the fruits of *Ziziphus jujuba* containing various cultivars in different regions and classification using chemometric analysis. *J. Pharm. Biomed. Anal*. **49**:1296–1302.

Guo, Y., Ni, Y. and Kokot, S. (2016). Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **153**:79–86.

Gurdeniz, G. and Ozen, B. (2009). Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem*. **116**:519–525.

Hadjmohammadi, M. R., Nazari, S. and Kamel, K. (2009). Determination of flavonoid markers in honey with SPE and LC using experimental design. *Chromatographia* **69**:1291–1297.

He, H. J. and Sun, D. W. (2015). Microbial evaluation of raw and processed food products by Visible/Infrared, Raman and Fluorescence spectroscopy. *Trends Food Sci. Technol.* **46**:199–210.

Hernández-Orte, P., Cacho, J. F. and Ferreira, V. (2002). Relationship between varietal amino acid profile of grapes and wine aromatic composition. Experiments with model solutions and chemometric study. *J. Agric. Food Chem.* **50**:2891–2899.

Hoffmann, B., Münch, S., Schwägele, F., Neusüß, C. and Jira, W. (2016). A sensitive HPLC-MS/MS screening method for the simultaneous detection of lupine, pea, and soy proteins in meat products. *Food Control* **71**:200–209.

Huang, L., Zhao, J., Chen, Q. and Zhang, Y. (2013). Rapid detection of total viable count (TVC) in pork meat by hyperspectral imaging. *Food Res. Int.* **54**:821–828.

Huanga, L. F., Wu, M. J. and Zhong, K. J. (2007). Fingerprint developing of coffee flavor by gas chromatography–mass spectrometry and combined chemometrics methods. *Anal. Chim. Acta* **588**:216–223.

Jandrić, Z. and Cannavan, A. (2017). An investigative study on differentiation of citrus fruit/fruit juices by UPLC-QToF MS and chemometrics. *Food Control*. **72**:173–180.

Jandrić, Z., Roberts, D., Rathor, M. N., Abrahim, A., Islam, M. and Cannavan, A. (2014). Assessment of fruit juice authenticity using UPLC–QToF MS: A metabolomics approach. *Food Chem*. **148**:7–17.

Joffraud, J. J., Leroi, F., Roy, C. and Berdague, J. L. (2001). Characterisation of volatile compounds produced by bacteria isolated from the spoilage flora of cold-smoked salmon. *Int. J. Food Microbiol.* **66**:175–184.

Jos, A., Moreno, I., González, A. G., Repetto, G. and Cameán, A. M. (2004). Differentiation of sparkling wines (cava and champagne) according to their mineral content. *Talanta* **63**:377–382.

Kamal, M. and Karoui, R. (2015). Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: a review. *Trends Food Sci. Technol.* **46**:27–48.

Kamruzzaman, M., Makino, Y. and Oshita, S. (2015). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: a review. *Anal. Chim. Acta* **853**:19–29.

Kos, G, Sieger, M, McMullin, D, Zahradnik, C., Sulyok, M., Öner, T., Mizaikoff, B., and Krsk, R. (2016). A novel chemometric classification for FTIR spectra of mycotoxin-contaminated maize and peanuts at regulatory limits. *Food Addit. Contam Part A*. **33**:1596–1607.

Köppen, R., Koch, M., Siegel, D., Merkel, S., Maul, R. and Nehls, I. (2010). Determination of mycotoxins in foods: current state of analytical methods and limitations. *Appl. Microbiol. Biotechnol.* **86**:1595–1612.

Lázaro, C. A., Conte-Júnior, C. A., Canto, A. C., Monteiro, M. L. G., Costa-Lima, B., da Cruz, A. G., Mársico, E. T. and Franco, R. M. (2015). Biogenic amines as bacterial quality indicators in different poultry meat species. *LWT-Food Sci. Technol.* **60**:15–21.

Lee, K. M. and Herrman, T. J. (2016). Determination and prediction of fumonisin contamination in maize by Surface–Enhanced Raman Spectroscopy (SERS). *Food Bioprocess Technol*. **9**:588–603.

Li, X., Zhang, Y. and He, Y. (2016). Rapid detection of talcum powder in tea using FT-IR spectroscopy coupled with chemometrics. *Sci. Rep.* **6**:30313 doi: 10.1038/srep30313.

Lima, D. B. D., Agustini, B. C., Silva, E. G. D., Gaensly, F., Cordeiro, R. B., Fávero, M. L. D., Brand, D., Maraschin, M. and Bonfim, T. M. B. (2011). Evaluation of phenolic compounds content and in vitro antioxidant activity of red wines produced from *Vitis labrusca* grapes. *Food Sci Technol. (Campinas)* **31**:783–800.

Lin, M., Al-Holy, M., Mousavi-Hesary, M., Al-Qadiri, H., Cavinato, A. G. and Rasco, B. A. (2004). Rapid and quantitative detection of the microbial spoilage in chicken meat by diffuse reflectance spectroscopy (600–1100 nm). *Lett. Appl. Microbiol.* **39**:148–155.

Lin, M., Mousavi, M., Al-Holy, M., Cavinato, A. G. and Rasco, B. A. (2006). Rapid near infrared spectroscopic method for the detection of spoilage in rainbow trout (*Oncorhynchus mykiss*) fillet. *J. Food Sci.* **71**:S18–S23.

Majcher, M. A., Kaczmarek, A., Klensporf-Pawlik, D., Pikul, J. and Jeleń, H. H. (2015). SPME-MS-Based Electronic Nose as a Tool for Determination of Authenticity of PDO Cheese, Oscypek. *Food Anal. Methods* **8**:2211–2217.

Martín, M. J., Pablos, F. and Gonzàlez, A. G. (1998). Discrimination between arabica and robusta green coffee varieties according to their chemical composition. *Talanta* **46**:1259–1264.

McKenzie, J. S., Jurado, J. M. and De Pablos, F. (2010). Characterisation of tea leaves according to their total mineral content by means of probabilistic neural networks. *Food Chem*. **123**:859–864.

Melucci, D., Bendini, A., Tesini, F., Barbieri, S., Zappi, A., Vichi, S., Conte, L. and Toschi, T. G. (2016). Rapid direct analysis to discriminate geographic origin of extra virgin olive oils by flash gas chromatography electronic nose and chemometrics. *Food Chem*. **204**:263–273.

Min, H. and Cho, B. (2015). Spectroscopic techniques for nondestructive detection of fungi and mycotoxins in agricultural materials: a review. *J. of Biosystems Eng.* **40**:67–77.

Monakhova, Y. B., Godelmann, R., Andlauer, C., Kuballa, T. and Lachenmeier, D. W. (2013). Identification of imitation cheese and imitation ice cream based on vegetable fat using nmr spectroscopy and chemometrics. *Int. J. Food Sci.* **2013**, Article ID 367841, p. 9. https://doi.org/10.1155/2013/367841.

Monakhova, Y. B., Ruge, W., Kuballa, T., Ilse, M., Winkelmann, O., Diehl, B., Thomas, F. and Lachenmeier, D. W. (2015). Rapid approach to identify the presence of Arabica and Robusta species in coffee using $^{1}$H NMR spectroscopy. *Food Chem*. **182**:178–184.

Montel, M. C., Reitz, J., Talon, R., Berdagué, J. L. and Rousset-Akrim, S. (1996). Biochemical activities of *Micrococcaceae* and their effects on the aromatic profiles and odours of a dry sausage model. *Food Microbiol*. **13**:489–499.

Moreda-Pineiro, A., Fisher, A. and Hill, S. J. (2003). The classification of tea according to region of origin using pattern recognition techniques and trace metal data. *J. Food Compost. Anal.* **16**:195–211.

Muntean, E. (2010). Simultaneous carbohydrate chromatography and unsuppressed ion chromatography in detecting fruit juices adulteration. *Chromatographia* **71**:69–74.

Neill, A. S., Nagle, C. M., Spurdle, A. B. and Webb, P. M. (2012). Use of talcum powder and endometrial cancer risk. *Cancer Causes Control* **23**:513–519.

Nicolaou, N. and Goodacre, R. (2008). Rapid and quantitative detection of the microbial spoilage in milk using Fourier transform infrared spectroscopy and chemometrics. *Analyst* **133**:1424–1431.

Ortea, I. and Gallardo, J. M. (2015). Investigation of production method, geographical origin and species authentication in commercially relevant shrimps using stable isotope ratio and/or multi-element analyses combined with chemometrics: An exploratory analysis. *Food Chem*. **170**:145–153.

Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M. and Barolo, M. (2012). Use of near-infrared spectroscopy for fast fraud detection in seafood: application to the authentication of wild european sea bass (*Dicentrarchus labrax*). *J. Agric. Food Chem.* **60**:639–648.

Özbalci, B., Boyaci, İ. H., Topcu, A., Kadılar, C. and Tamer, U. (2013). Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chem*. **136**:1444–1452.

Papadopoulou, O., Panagou, E. Z., Tassou, C. C. and Nychas, G. J. E. (2011). Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. *Food Res. Int.* **44**:3264–3271.

Paradkar, M. M. and Irudayaraj, J. (2002). Determination of cholesterol in dairy products using infrared techniques: 1. FTIR spectroscopy. *Int. J. Dairy Technol.* **55**:127–132.

Parente, E., Martuscelli, M., Gardini, F., Grieco, S., Crudele, M. A. and Suzzi, G. (2001). Evolution of microbial populations and biogenic amine production in dry sausages produced in Southern Italy. *J. Appl. Microbiol.* **90**:882–891.

Pontes, M. J. C., Santos, S. R. B., Araújo, M. C. U., Almeida, L. F., Lim, R. A. C., Gaião, E. N. and Souto, U. T. C. P. (2006). Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Res. Int.* **39**:182–189.

Ragone, R., Crupi, P., Piccinonna, S., Bergamini, C., Mazzone, F., Fanizzi, F. P., Schena, F. P and Antonacci, D. (2015). Classification and chemometric study of Southern Italy monovarietal wines based on NMR and HPLC-DAD-MS. *Food Sci. Biotechnol.* **24**:817–826.

Reis, N., Franca, A. S. and Oliveira, L. S. (2013). Performance of diffuse reflectance infrared Fourier transform spectroscopy and chemometrics for detection of multiple adulterants in roasted and ground coffee. *LWT - Food Sci. Technol.* **53**:395–401.

Reis, N., Franca, A. S. and Oliveira, L. S. (2016). Concomitant use of Fourier transform infrared attenuated total reflectance spectroscopy and chemometrics for quantification of multiple adulterants in roasted and ground coffee. *J. Spectrosco.* **2016**, Article ID 4974173, p. 7. https://doi.org/10.1155/2016/4974173.

Rendall, R., Reis, M. S., Pereira, A. C., Pestana, C., Pereira, V. and Marques, J. C. (2015). Chemometric analysis of the volatile fraction evolution of Portuguese beer under shelf storage conditions. *Chemometr. Intell. Lab. Syst.* **142**:131–142.

Ribeiro, J. S., Augusto, F., Salva, T. J. G. and Ferreira, M. M. C. (2012). Prediction models for Arabica coffee beverage quality based on aroma analyses and chemometrics. *Talanta* **101**:253–260.

Ribeiro, J. S., Augusto, F., Salva, T. J. G., Thomaziello, R. A. and Ferreira, M. M. C. (2009). Prediction of sensory properties of Brazilian Arabica

roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares. *Anal. Chim. Acta* **634**:172–179.

Rodríguez García, J. C., Iglesias Rodríguez, R., Peña Crecente, R. M., Barciela Garcí, J., García Martín, S. and Herrero Latorre, C. (2006). Preliminary chemometric study on the use of honey as an environmental marker in Galicia (northwestern Spain). *J. Agric. Food Chem.* **54**:7206–7212.

Rohman, A., Che Man, Y. B. and Yusof, F. M. (2014). The use of FTIR spectroscopy and chemometrics for rapid authentication of extra virgin olive oil. *J. Am. Oil Chem. Soc.* **91**:207–213.

Römisch, U., Jäger, H., Capron, X., Lanteri, S., Forina, M. and Smeyers-Verbeke, J. (2009). Characterization and determination of the geographical origin of wines. Part III: multivariate discrimination and classification methods. *Eur. Food Res. Technol.* **230**:31–45.

Saidi, A. and Mirzaei, M. (2016). Spectrofluorimetric determination of ochratoxin A in wheat and rice products using an artificial neural network. *J. Anal. Chem.* **71**:158–162.

Santos, P. M., Pereira-Filho, E. R. and Rodriguez-Saona, L. E. (2013). Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food Chem*. **138**:19–24.

Schuhfried, E., del Pulgar, J. S., Bobba, M., Piro, R., Cappellin, L., Märk, T. D. and Biasioli, F. (2016). Classification of 7 monofloral honey varieties by PTR-ToF-MS direct headspace analysis and chemometrics. *Talanta* **147**:213–219.

Seetohul, N. L., Islam, M., O'Hare, W. T. and Ali, Z. (2006). Discrimination of teas based on total luminescence spectroscopy and pattern recognition. *J. Sci. Food Agric.* **86**:2092–2098.

Sendin, K., Williams, P. J. and Manley, M. (2016). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Crit. Rev. Food Sci. Nutr.* https://doi.org/10.1080/10408398.2016.1205548.

Sola-Larrañaga, C. and Navarro-Blasco, I. (2009). Chemometric analysis of minerals and trace elements in raw cow milk from the community of Navarra, Spain. *Food Chem*. **112**:189–196.

Souza, S. S., Cruz, A. G., Walter, E. H., Faria, J. A., Celeghini, R. M., Ferreira, M. M., Granato, D. and Sant'Ana, A. D. S. (2011). Monitoring the authenticity of Brazilian UHT milk: A chemometric approach. *Food Chem*. **124**:692–695.

Tyszkiewicz, J. and Tyszkiewicz, S. (2004). An introduction for chemometrics in food science. *Pol. J. Food Nutr. Sci.* **13**:75–86.

Vainionpää, J., Smolander, M., Alakomi, H. L., Ritvanen, T., Rajamäki, T., Rokka, M. and Ahvenainen, R. (2004). Comparison of different analytical methods in the monitoring of the quality of modified atmosphere packaged broiler chicken cuts using principal component analysis. *J. Food Eng.* **65**:273–280.

Vera, L., Aceña, L., Guasch, J., Boqué, R., Mestres, M. and Busto, O. (2011). Characterization and classification of the aroma of beer samples by means of an MS e-nose and chemometric tools. *Anal. Bioanal. Chem.* **399**:2073–2081.

Wang, W., Peng, Y., Huang, H. and Wu, J. (2011). Application of hyperspectral imaging technique for the detection of total viable bacteria count in pork. *Sens. Lett.* **9**:1024–1030.

Wang, C., Zhang, C. X., Shao, C. F., Li, C. W., Liu, S. H., Peng, X. P. and Xu, Y. Q. (2016). Chemical fingerprint analysis for the quality evaluation of deepure instant pu-erh tea by HPLC combined with chemometrics. *Food Anal. Methods*. **9**:3298–3309.

Williams, P. J., Geladi, P., Britz, T. J. and Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *J. Cereal Sci.* **55**:272–278.

Wu, T., Chen, H., Lin, Z. and Tan, C. (2016). Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. *J. Spectrosc.* 2016, Article ID 6184987, p. 8. https://doi.org/10.1155/2016/6184987.

Zanardi, E., Caligiani, A., Palla, L., Mariani, M., Ghidini, S., Di Ciccio, P. A., Palla, G. and Ianieri, A. (2015). Metabolic profiling by $^{1}$H NMR of ground beef irradiated at different irradiation doses. *Meat Sci*. **103**:83–89.

Zhao, J., Du, X., Cheng, N., Chen, L., Xue, X., Wu, L. and Cao, W. (2016). Identification of monofloral honeys using HPLC–ECD and chemometrics. *Food Chem*. **194**:167–174.

Zhou, J., Yao, L., Li, Y., Chen, L., Wu, L. and Zhao, J. (2014). Floral classification of honey using liquid chromatography–diode array detection–tandem mass spectrometry and chemometric analysis. *Food Chem*. **145**:941–949.

Zielinski, A. A. F., Ávila, S., Ito, V., Nogueira, A., Wosiacki, G. and Haminiuk, C. W. I. (2014a). The association between chromaticity, phenolics, carotenoids, and in vitro antioxidant activity of frozen fruit pulp in Brazil: an application of chemometrics. *J. Food Sci.* **79**:C510–C516. a

Zielinski, A. A. F., Haminiuk, C. W., Nunes, C. A., Schnitzler, E., Ruth, S. M. and Granato, D. (2014b). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: A critical review and guideline. *Compr. Rev. Food Sci. Food Saf.* **13**:300–316.