ACTA IMEKO
ISSN: 2221-870X
April 2016, Volume 5, Number 1, 32-35

Spectroscopic fingerprinting techniques for food characterisation

Monica Casale, Lucia Bagnasco, Chiara Casolino, Silvia Lanteri, Riccardo Leardi

Univeristy of Genoa, Departmentof Pharmacy – Via Brigata Salerno 13, 16147 Genova, Italy

ABSTRACT
The analysis of samples by using spectroscopic fingerprinting techniques is more and more common and widespread. Such approaches are very convenient, since they are usually fast, cheap and non-destructive. In many applications no sample pretreatment is required, the acquisition of the spectrum can be performed in about one minute and no solvents are required. As a consequence, the return on investment of the related technology is very high.
The "disadvantage" of these techniques is that, the signal being non-selective, simple mathematical approaches (e.g., Lambert-Beer law) cannot be applied. Instead, a multivariate treatment must be performed by using chemometrics tools.
In what concerns food analysis, they can be applied in several steps, from the evaluation of the quality and the conformity of raw material to the assessment of the quality of the final product, to the monitoring of the shelf life of the product itself. Another interesting field of application is the verification of food-authenticity claims, this being extremely important in the case of foods labeled as protected designation of origin (PDO), protected geographical indication (PGI) and traditional speciality guaranteed (TSG).
In the present paper, it is described how non-selective signals can be used for obtaining useful information about a food.

Section: RESEARCH PAPER

Keywords: non selective signals; UV-Visible spectroscopy (UV-VIS); mid-infrared spectroscopy (MIRS); near-infrared spectroscopy (NIRS); chemometrics; food analysis

Citation: Monica Casale, Lucia Bagnasco, Chiara Casolino, Silvia Lanteri, Riccardo Leardi, Spectroscopic fingerprinting techniques for food characterisation, Acta IMEKO, vol. 5, no. 1, article 7, April 2016, identifier: IMEKO-ACTA-05 (2016)-01-07

Section Editor: Claudia Zoani, Italian National Agency for New Technologies, Energy and Sustainable Economic Development affiliation, Rome, Italy

Received June 26, 2015; **In final form** December 17, 2015; **Published** April 2016

Corresponding author: Riccardo Leardi, e-mail: riclea@difar.unige.it

1. INTRODUCTION

Traditional analytical methods for food analysis have several drawbacks, such as low speed, the necessity for sample pre-treatments, a requirement for highly-skilled personnel and destruction of the sample.

Several fast and non-destructive instrumental methods have been proposed to overcome these hurdles. Among them, UV-Visible, Mid infrared and Near infrared spectroscopy have proven to be successful analytical methods for analysis of food since they offer a number of important advantages [1], [2]. They are fast, non-destructive methods and they require a minimal or no sample preparation. Moreover, they are less expensive because no reagents are required and thus no wastes are produced. Finally, the versatility of these instruments makes them useful tools for online process monitoring.

Chemical information contained in the spectra resides in the band positions, intensities, and shapes. Whereas band positions give information about the molecular structure of chemical compounds, the intensities of the bands are related to the concentration of these compounds as described by the Lambert-Beer law. The easiest way to determine the content of a chemical compound is to measure the change in the intensity of a well-resolved band that has been unambiguously attributed to this compound [3].

However, this is possible for a pure component system, but foods contain numerous components giving rise to complex spectra with overlapping peaks.

In fact, in order to take advantage of these spectroscopic fingerprinting techniques, the analyst must overcome limitations in sensitivity and selectivity that arise from the

relatively weak and highly overlapping bands found in the spectra. The result is a unique "fingerprint" that can be used to confirm the identity of a sample.

Thus, for the implementation of a successful analysis the use of chemometric methods is fundamental to extract from the spectra as much information as possible about the analysed samples.

The information extracted from the non-selective signals can be used for two types of analysis:

(1) quantitative analysis to link features of the spectra to quantifiable properties of the samples.

Spectroscopic techniques are commonly used to obtain calibration models able to predict the concentration of a compound or a specific characteristic of a food product. Several studies have been performed, for example on the detection of adulterants in foods.

(2) qualitative analysis, i.e. classification of samples.

Recently, a rather specific type of fraud has become important involving claims regarding the geographical origin of food ingredients. It is generally true that deceptions regarding the geographical origin of foods have few health implications, but they may nonetheless represent a serious commercial fraud. In fact, consumers often pay a considerable price premium when a food is labelled with a declaration of production within a specific region, since such a label may be perceived as an implicit guarantee of a traditional and, perhaps, healthier manufacturing process. In response, several national and international institutions have issued directives to support the differentiation of agricultural products and foodstuffs on a regional basis by introducing an integrated framework for the protection of both geographical origin and traditional production techniques. In these cases, non-selective signals can be used to obtain classification models able to discriminate samples according to a quality/characteristic.

In this paper, as an example, the construction of a reliable quantitative model for the detection of addition of barley to coffee using NIR spectroscopy and chemometrics is shown.

2. HOW TO BUILD A MODEL

Figure 1 shows the scheme for building a quantitative or qualitative multivariate model.

The main steps to be performed are:

1) Sample selection

The analysed samples should fully represent the population studied; this means that all the variability sources (or at least the most important ones) should be taken into account. Moreover, when a calibration model is developed the samples should cover a wide range of response values. Unfortunately, there are many reports based on poor sampling, and this affects the result of the whole analysis. For example, in order to discriminate extra virgin olive oils on the basis of the olive cultivar, oil samples should be collected from different oil mills in order to take into account the different sources of variability, such as geographical origin, production year, harvest period and production technologies (fertilizer, olive fly control tools, extraction process, etc).

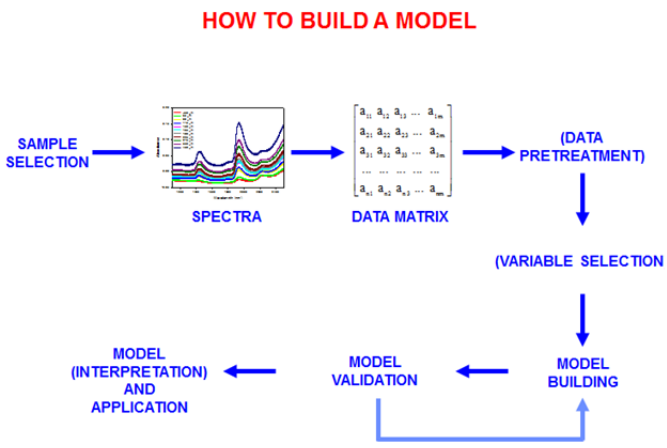

Figure 1. Scheme for building a quantitative or qualitative multivariate model.

2) Spectra acquisition and data matrix

The acquisition of the spectra is very simple and generally requires few minutes. Then, spectra have to be arranged in a data matrix in which each row corresponds to a sample and each column to a variable (wavelength). This trivial operation is not always easy to be performed since many instruments do not allow the direct exportation of a set of spectra.

3) Sample pretreatment

Spectra can be affected by turbidity in liquid samples or different granulometry in solid samples and by variations in the optical path length. To avoid or decrease these interferences, mathematical pretreatments are required.

The most current data pretreatments are normalisation methods such as standard normal variate (SNV) [4] and derivatives.

The SNV, or row autoscaling, mainly corrects both baseline shifts and global intensity variations, which are related to the granulometry of the sample. Each spectrum is row-centered, by subtracting its mean from each single value, and then scaled by dividing it by its standard deviation. As a result, each spectrum has mean equal to 0 and standard deviation equal to 1.

Since SNV removes the possibly shifting informative regions along the signal range, the interpretation of the results referring to the original signals should be performed with caution.

Other common pretreatments applied to spectroscopic signals are the derivatives. First derivative provides a correction for baseline shifts, while the second derivative represents a measure of the curvature of the original signal, i.e., the rate of change of its slope. Such transform provides a correction for both baseline shifts and drifts. A drawback of derivatives is the increase of random noise.

4) Variable selection

Since the UV-Visible, NIR and MIR spectra are characterized by a very high number of variables, the selection of the informative ones is an important task, both to obtain simpler qualitative or quantitative models and to identify the most useful wavelengths. Several algorithms can be applied. Among them, some commonly used are the stepwise selection methods (i.e: Stepwise Linear Discriminant Analysis [5], Iterative Stepwise Elimination [6]), Interval-PLS [7] and Genetic Algorithms [8]. Compared with the other techniques, Genetic Algorithms have the great advantage of selecting well defined spectral regions, often corresponding to relevant spectral features, instead of single wavelengths.

5) Model Building and validation

The choice of the classification or regression method should be performed following the simplicity criterion, since simpler models are also more robust and stable. The most used ones are Linear Discriminant Analysis (LDA) [9] and K Nearest Neighbors (KNN) [10] for classification, Soft Independent

Modeling of Class analogy (SIMCA) [11] as class-modeling technique and Partial Least Square (PLS) [12] for multivariate calibration.

A chemometrical model must be evaluated on the basis of its predictive ability. Cross-validation is the most common validation procedure: the N objects are divided into G cancellation groups, the model is computed G times and each time the objects in the corresponding cancellation group are predicted. At the end of the procedure, each object has been predicted once.

6) External test set

The real predictive ability of each model should be tested on an external set of samples.

The test set has to be totally independent of the calibration set, not only mathematically as in cross-validation, but also from the 'spatial' and 'temporal' point of view. For instance, in case of industrial data the samples of the test set must be produced and analysed after the samples of the training set; in case of samples whose origin can be very different the samples of the test set must be produced by producers that are not the same as the producers of the samples of the training set; in case of natural products the samples of the test set must come from crops subsequent to those of the samples of the training set.

3. EXAMPLE OF CONSTRUCTION OF A RELIABLE QUANTITATIVE MODEL FOR FOOD ANALYSIS

This study [13] presents an application of near infrared spectroscopy for detection and quantification of the fraudulent addition of barley in roasted and ground coffee samples.

Nine different types of coffee including pure Arabica, Robusta and mixtures of them at different roasting degrees were blended with four types of barley. The types of coffee and barley were selected in order to be as representative as possible of the Italian market.

Since it was decided to perform 100 experiments out of the 360 combinations resulting from nine coffees, four barleys and ten different concentrations (from 2 % to 20 % w/w), the calibration set was defined by applying a D-optimal design. The validation was performed on thirty experiments selected by applying a subsequent D-optimal design on the combinations not constituting the training set.

After that, a further validation was performed on a completely external set made by mixtures of a type of coffee and a type of barley, different from those used in the previous steps.

Partial least squares regression (PLS) was employed to build the models aimed at predicting the amounts of barley in coffee samples. In order to obtain simplified models, taking into account only informative regions of the spectral profiles, genetic algorithms were applied for feature selection. This allowed to reduce the number of data points from 1501 to 188.

The models showed excellent predictive ability with root mean square errors (RMSE) for the test and external sets equal to 1.4 % w/w and 1.1 % w/w, respectively. As it can be noticed in Figure 2, the model obtained is capable to predict barley concentration with a very satisfactory accuracy, not only in calibration but also on external samples.

This application clearly shows that the representativity of the training set is a key point in the success of a calibration model. The achievement of very low prediction errors on a totally external test set (i.e., on mixtures composed by qualities of coffee and barley unknown to the model) has been possible only as a consequence of the fact that the training set was made by taking into account a relatively large number of varieties of coffee and barley. Another key point is the application of D-optimal design for the selection of a subset of adequate size from the very high set of candidate experiments. Moreover, the variable selection by using genetic algorithms helped to determine the spectral regions most useful to identify the adulteration of coffee with barley and to increase calibration model performances.

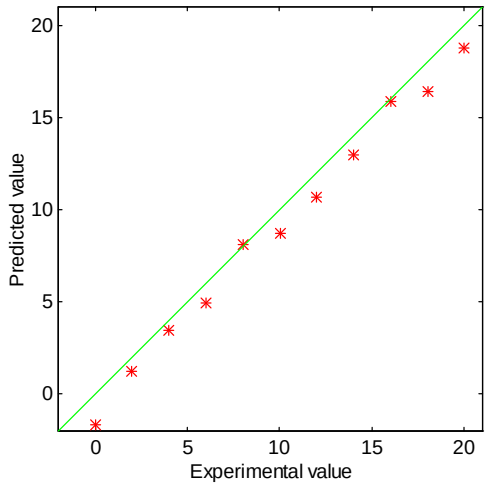

Figure 2. Experimental vs. predicted values of concentration (% w/w) of barley in the coffee samples of the external test set (PLS model on the spectral regions selected by GA).

4. CONCLUSIONS

In the present paper it has been shown, from a methodological point of view, how non-selective signals can be used for obtaining useful information about food.

The chemometrical elaboration is fundamental in order to obtain useful information from the spectral data; the main steps of this approach have been identified and their importance discussed also showing a real application.

REFERENCES

[1] T. Woodcock, G. Downey, C.P. O'Donnell, Review: Better quality food and beverages: the role of near infrared spectroscopy, Journal of Near Infrared Spectroscopy, 16(1), (2008), pp. 1-29.

[2] L. Wang, F. S.C. Lee, X. Wang, Y. He, Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR, Food Chemistry, 95, (2006) pp.529–536.

[3] R. Karoui, G. Downey, C. Blecker, Mid-Infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships - a review, Chemical Reviews, 110, (2010), pp.6144–6168.

[4] R.J. Barnes, M.S Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy, 43(5), (1989), pp.772-777.

[5] R.I. Jenrich, in K. Enselein, A. Ralston, H.S. Wilf (Eds)Statistical Methods for digital computers, J. Wiley and Sons, New York, 1960, pp. 76.

[6] R. Boggia, M. Forina, P. Fossa, L. Mosti, Chemometric study and validation strategy in the structure-activity relationships of new cardiotonic agents, Quantitative Structure-Activity Relationships, 16, (1997), pp. 201-213.

[7] L. Nørgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck and S.B. Engelsen, Interval Partial Least Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy, Applied Spectroscopy, 54, (2000) pp. 413-419.

[8] R. Leardi, A.L. Gonzalez, Genetic algorithms applied to feature selection in PLS regression: how and when to use them, Chemom. Intell. Lab. Syst. 41 (1998) pp.195–207.

[9] D. L Massart, B. G. M. Vandeginste, L. M. C Buydens, S. De Jong, P. L Lewi, J. Smeyers-Verbeke, "Supervised pattern recognition" in, Handbook of Chemometrics and Qualimetrics: Part B, vol. 20B, B.G.M. Vandeginste, & S.C. Rutan, Elsevier, Amsterdam, 1998, pp. 213-220.

[10] T. M. Cover and P. Hart, "The nearest neighbor decision rule," IEEE Trans. Inform. Theory, vol. IT-13, pp. 21-27 1967.

[11] S. Wold, M. Sjostrom, "SIMCA: A method for analysing chemical data in terms of similarity and analogy" in Chemometrics, Theory and Application, B. R. Kowalski, ACS Symposium Series 52, American Chemical Society, Washington, DC, 1977, pp. 243.

[12] H. Wold, "Partial Least Squares" in: S. Kotz, N.L. Johnson, (Eds.), Encyclopedia of Statistical Sciences, Wiley, New York, 1985, pp. 581–591.

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