

Confirming the geographic origin of foods by near infrared spectroscopy

Federico Marini*, Marta Bevilacqua, Remo Bucci, Andrea D. Magrì and Antonio L. Magrì

Department of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, I-00185 Rome, Italy

*Corresponding author: fmmonet@hotmail.com

Introduction

The increased information available to consumers together with competition following the institution of the single market (European Union countries) has led safety, nutritional value, eating characteristics, ethical, environmental, economic and social aspects to be more and more important and essential topics for the food industry. In a single word, there is an increasing demand for quality as a primary criterion to access the market. Accordingly, a particular aspect of quality which has been gathering more and more attention during the recent years is that of product *typicity* or uniqueness. In fact, it has been widely reported in the literature that the geographic, production and species origin of a foodstuff are important variables regulating the overall quality of the product. This issue has been recognised in the European Union by the introduction of the Designations of Origin (PDO and PGI). The products which are labeled by this denomination must be produced in a well-defined geographic area and manufactured using only one or few specified botanical varieties or animal species.

As a consequence, the availability of accurate and precise analytical methods for confirmation of the origin of a foodstuff is an issue that is becoming more and more relevant. Against this background, while the possibility of identifying chemical markers of geographic, production or species origin is currently being explored using different fingerprinting techniques, it is widely recognised that, whatever the technique(s) chosen, the use of mathematical and statistical methods (chemometrics) to process the experimental data is essential to obtain a reliable authentication of the product.¹⁻⁴ In this communication, the successful use of different chemometric pattern recognition methods for confirming the geographic origin of two high value-added food products (extra virgin olive oil and pistachios) based on NIR measurements will be presented.

Materials and Methods

Samples

Two different food matrices were considered in this study - extra virgin olive oil and pistachio nuts.

Olive oil

The study on olive oil samples focused on the possibility of authenticating products from the PDO Sabina, a PDO from central Italy. For this purpose, oils from the PDO Sabina and other geographical origins were sampled during the two harvests 2009 and 2010. In particular, 20 oils from Sabina (13 from 2009 and 7 from 2010) and 37 oils from different origins (22 from 2009 and 15 from 2010) were collected and analyzed by NIR spectroscopy.

Pistachio

The study on pistachio nuts focused on the possibility of discriminating the origin of pistachio samples coming from 4 different countries: Italy, Iran, Turkey and Syria. In this study, fruits from different vendors were sampled and at least 20 nuts per country were analyzed by NIR spectroscopy, both as whole samples and after longitudinal slicing.

NIR measurements

Olive oil

NIR spectra were recorded in the range 1100-2500 nm on a Thermo Nicolet instrument equipped with an integrating sphere. Each oil sample (1 mL) was placed in a glass vial (1.8 cm i.d.) that was then covered with a mirror, so that measurements were conducted in transreflectance. Each sample was analysed in triplicate.

Pistachio

NIR spectra were recorded in the range 1100-2500 nm on a Thermo Nicolet instrument equipped with an integrating sphere. Spectra of whole pistachio nuts after removal of the shell were recorded at different orientations and averaged. Additionally, nuts were longitudinally split in half and spectra of the flat surface were also recorded.

Reference paper as:

F. Marini, M. Bevilacqua, R. Bucci, A.D. Magrì and A.L. Magrì (2012). Confirming the geographic origin of foods by near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 136-139.

Chemometric data processing

NIR spectra collected as reported in the previous subsection were arranged in a matrix and analysed by chemometric pattern recognition techniques. In particular, both a discriminant (based on a PLS-DA classifier) and a modeling approach (using SIMCA) were tested in the case of olive oil samples, while PLS-DA only was used to process data on pistachios.

Pretreatment

As several instrumental effects can hinder the performances of the calibration models, different kinds of spectral pre-treatments (MSC, detrending, 1st and 2nd derivatives, and the combination of MSC with any of the latter three)⁵ were tested.

PLS-DA

Discriminant classification techniques can be viewed as aiming to find a relationship between a multivariate independent vector **X** and a qualitative vector of responses. Accordingly, if a suitably-designed dummy response vector is introduced, traditional regression methods can be used also to tackle classification problems. To do so, PLS-DA⁶, as the name itself suggest, uses Partial Least Squares Regression⁷ i.e. computes a bilinear decomposition of both the X- and Y- spaces on the assumption that a relationship between the two internal spaces exists. The result is a linear classifier that has proved to be statistically equivalent to Linear Discriminant Analysis (LDA⁸) but that is also applicable to cases which LDA cannot handle i.e. low number of samples: number of variables ratio and high co-linearity among the variables.

SIMCA

In SIMCA⁹⁻¹⁰, each class model is defined on the basis of a principal component model of appropriate dimensionality. Once the principal component model is computed, the class space is defined according to some statistically defined outlier detection criterion. In particular, the score matrix **T_A** and the residual matrix **E** are used to compute the probability distributions for the distances within the model space (**T²** statistics) and for the orthogonal distance to the model space (**Q** statistics) respectively. When a sample has a **T²** or a **Q** value that exceed a selected percentile of the corresponding distributions (usually 95%), the sample is considered as an outlier for the category and rejected by the class model. Otherwise, if both values are lower than the thresholds, the sample is recognised as belonging to the class.

Results and Discussion

Olive oil

In a first stage of the study, data from both years were gathered to build a general classification model. In particular, a selection procedure based on the Duplex algorithm was used to divide samples into training and test sets. Accordingly, 35 samples (13 from Sabina and 22 from other origins) were selected as a training set and the remaining 22 were left out for the external validation stage.

PLS-DA

PLS-DA models were built on the training set and the optimal complexity was chosen as that leading to the minimum classification error in cross-validation. The best classification accuracy was reached when either MSC or 1st derivative were used to pre-process the data. In particular, the PLS-DA model built on detrended data (4 LVs) was able to correctly classify 100% of Sabina and 93.9% of samples from other origins in the calibration stage and 97.4% of Sabina and 93.9% of other samples in cross-validation. On the other hand, the model built on 1st derivative pre-treated data (5 LVs) led to 100% and 95.5% classification accuracy in calibration and to 97.4% and 92.4% correct classification in cross-validation for Sabina and other origins respectively. When tested on the external validation set, both these models performed well, resulting in 100% correct predictions for Sabina samples and more than 90% for the oils coming from different origins.

Interpretation of the models based on VIP scores showed that, in both cases, the spectral regions mostly responsible of the observed discrimination were 4000-4725, 5028-5317, 5662-6048 cm⁻¹; additionally, in the case of 1st derivative, a fourth interval (7085-7384 cm⁻¹) appeared to significantly contribute to the projection.

SIMCA

SIMCA models of the class Sabina were built using training set samples. The optimal complexity of each model was chosen as that leading to the highest value of the geometric average of cross-validated sensitivity and specificity. The best results were obtained using 1st derivative only (4PCs; 94.9% sensitivity and 90.9%

Reference paper as:

F. Marini, M. Bevilacqua, R. Bucci, A.D. Magrì and A.L. Magrì (2012). Confirming the geographic origin of foods by near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 136-139.

specificity in calibration; 82.1% sensitivity and 92.4% specificity in cross-validation) or MSC followed by 1st derivative (5PCs; 94.2% sensitivity and 92.4% specificity in calibration; 82.1% sensitivity and 93.9% specificity in cross-validation). When these models were applied to the external validation set, comparable sensitivity and specificity values were observed: 85.7% sensitivity and 77.8% specificity for 1st derivative and 85.7% sensitivity and 80.0% specificity for MSC+1st derivative.

Effect of the production year

To check whether the different production year could cause variability in the spectral fingerprint, chemometric analysis was repeated using the samples from 2009 as a training set and those from 2010 as the external validation set. In this case, the most effective pre-treatments appeared to be MSC (5PCs; 94.9% sensitivity and 97.0% specificity on the calibration set; 84.6% sensitivity and 98.5% specificity in cross-validation) and 1st derivative (4PCs; 97.4% sensitivity and 100.0% specificity on the calibration set; 82.1% sensitivity and 100.0% specificity in cross-validation). Therefore, models built on the data collected from samples harvested in a single year are more sensitive and specific. However, when these models were applied to the external validation set, specificity remained high (82.2%) but sensitivity was 0%, suggesting a significant effect of production year on the experimental fingerprint.

These outcomes can be visualised in Figure 1, where the results of SIMCA modeling on 1st derivative-pre-treated data are shown in the form of model space and Coomans plots.

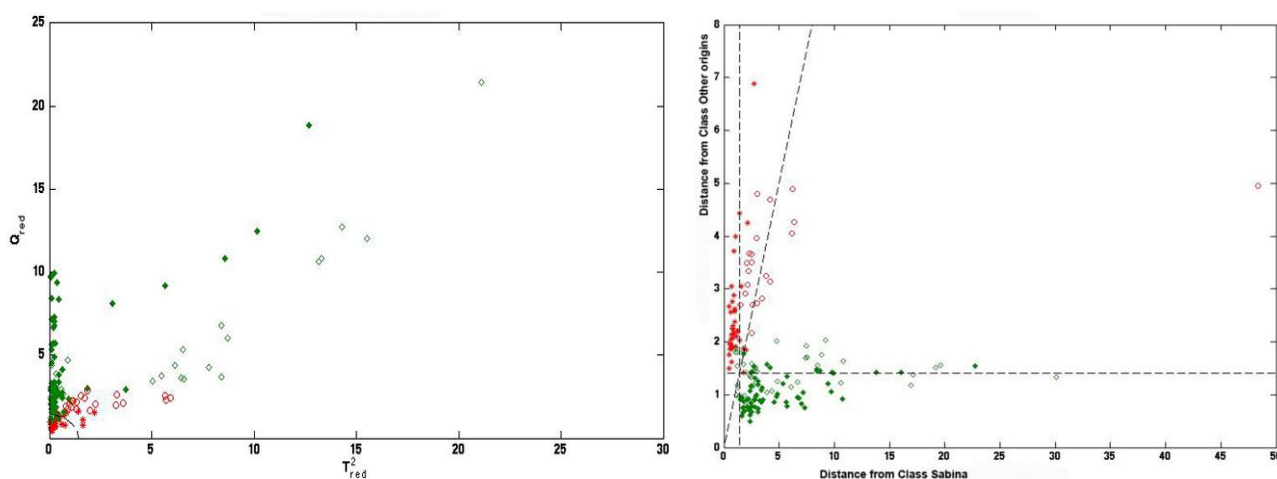


Figure 1. SIMCA modeling of olive oil data – effect of production year: (*left*) 2009 (filled symbols) and 2010 (empty symbols) samples from Sabina (red) and other origins (green) projected on the model space of class Sabina; (*right*) Coomans plot showing the distance of each sample from the two class models.

Pistachio

NIR data of whole pistachio nuts were divided into training and test sample sets using the duplex algorithm and including 14 samples for each class in the matrix used for calibration and 6 from each class in the external validation set. Only PLS-DA was used for data analysis and the optimal complexity of the models was fixed based on the non-error rate in cross-validation. Accordingly, best results were obtained using MSC as a spectral pre-treatment (4 LVs), resulting in classification accuracies higher than 95% in training, cross-validation and testing for all classes except Turkey. Indeed, only about 70% of Turkish samples were correctly assigned both in the internal and external validation.

This result can be graphically explained by observing the position of the Turkish samples in the space spanned by the first three latent vectors (Figure 2).

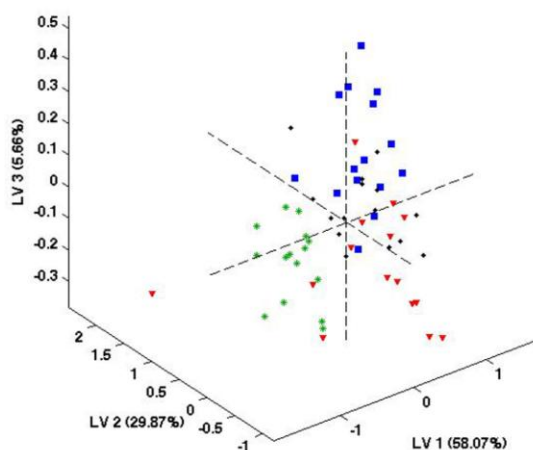


Figure 2. PLS-DA modeling of pistachios: cross-validated scores projected onto the first three LVs (blue: Italy, green: Iran, red: Syria; black: Turkey)

Conclusion

In the two examples reported, the coupling of chemometric pattern recognition methods to NIR spectroscopy proved to be a valuable and versatile tool for authentication of the origin of two foodstuffs. In the case of olive oil, the modeling approach allowed to not only confirmation of the geographic origin of the samples but also evaluation of the impact of production year on the experimental fingerprint.

References

1. D.M.A.M. Luykx and S.M. van Ruth, *Food Chem.* **107**, 897-911 (2008).
2. R. Consonni and L.R. Cagliani, *Adv. Food Nutrit. Res.* **59**, 87-165 (2010).
3. S. Kelly, K. Heaton and J. Hoogewerff, *Trends Food Sci. Technol.* **16**, 555-567 (2005).
4. <http://www.trace.eu.org>
5. A. Rinnan, F. van den Berg and S.B. Engelsen, *Trends Anal. Chem.* **28**, 1201-1222 (2009).
6. M. Barker and W. Rayers, *J. Chemom.* **17**, 166-173 (2003).
7. S. Wold, H. Martens and H. Wold, "The multivariate calibration method in chemistry solved by the PLS method", in *Matrix Pencils: Proceedings Of A Conference Held At Pite Havsbad (Lecture Notes in Mathematics)*, Ed by A. Ruhe and B. Kagstrom, Springer-Verlag, Heidelberg, pp. 286-293 (1983).
8. R.A. Fisher, *Ann. Eugen.* **7**, 179-188 (1936).
9. S. Wold, *Pattern Recognit.* **8**, 127-139 (1976).
10. S. Wold and M. Sjöström, "SIMCA: a method for analysing chemical data in terms of similarity and analogy", in *Chemometrics, Theory and Application. ACS Symposium Series No. 52*, Ed by B.R. Kowalski, American Chemical Society, Washington DC, pp 243-282 (1977).