Near infrared spectroscopy and class-modelling techniques to discriminate Taggiasca table olives

Lucia Bagnasco[1]*, Chiara Casolino[2] and Luca Medini[3]

[1]Department of Chemistry and Industrial Chemistry, University of Genoa, Genoa, 16146, Italy
[2]Department of Pharmaceutical and Food Chemistry and Technology, University of Genoa, Genoa, 16147, Italy
[3]Laboratorio Chimico Merceologico, Camera di Commercio di Savona, Albenga (SV), 17031, Italy
*Corresponding author: lucia.bagnasco@unige.it

Introduction

The Mediterranean coastal areas have a mild, warm climate that fully meets the climatic requirements of *Olea europaea* L. trees and they are thus considered an ideal habitat for their growth and development.
The Taggiasca cultivar is certainly the best known of the olive cultivars in Liguria, an Italian region, and the extra-virgin olive oils obtained from these olives are highly valued all over the world. A minor part of Taggiasca crop is used to prepare table olives of high commercial value that can be easily adulterated by olives having similar morphological characteristics but less appreciated sensory properties.

The identification of the cultivar of table olives is a topic of great economic relevance since the demand for table olives is increasing and there is a growing commercial interest in high quality products. Thus, reliable analytical techniques are necessary to certify olive quality and origin. Identification of olive cultivar has been traditionally carried out by morphological, agronomic and sensory traits, genetic approaches and gas chromatography analysis. Although these methodologies provide useful tools for cultivar identification, they present some limitations, i.e. low selectivity and time requirements.

In a previous study,[1] results obtained by NIR spectroscopic analysis of Taggiasca table olive samples were compared to those obtained using the fatty acid composition of the oils extracted from the same olives. The results obtained using the NIR spectra are even better than the results obtained using the chemical information from the GC analysis (Figure 1).

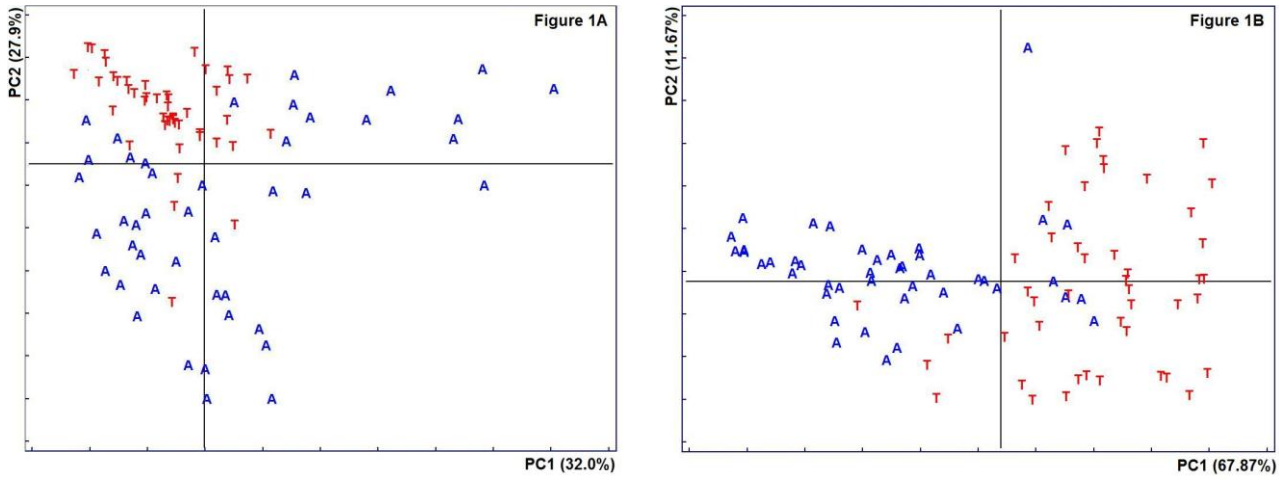

Figure 1. Score plot on PC1 and PC2 obtained using GC **(A)** and NIR **(B)** data. Samples are represented by their class symbol: T (red) = Taggiasca; A (blue) = other cultivar. GC analysis was performed following the European Commission Regulation (EC) No 1989/2003; NIR measurements were performed in accordance with the procedure detailed below**.**

In this work, the potential application of NIR spectroscopy, coupled with multivariate analysis, as an innovative tool to discriminate Taggiasca table olives from other varieties is confirmed and the repeatability and reproducibility of the method is evaluated.

Materials and Methods

Samples

The sampling was organised by Laboratorio Chimico Merceologico, so the traceability and representativity of the tested matrix was assured. Table olive samples (n=161) were analysed: 90 (46 from Taggiasca cultivar and 43 from other cultivars) were obtained from the 2007-2008 olive crop (SET1) and 71 (35 of Taggiasca cultivar and 36 of other cultivars) were obtained from the 2008-2009 olive crop (SET2).

Reference paper as:
L. Bagnasco, C. Casolino and L. Medini (2012).Near infrared spectroscopy and class-modelling techniques to discriminate Taggiasca table olives, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 119-122.

Near infrared spectroscopy

For the analysis, olives were washed with water, dried, stoned and ground. Then, the olive paste was analysed in reflectance mode by NIR spectroscopy in the 4000 – 10000 cm^{-1} range at 4 cm^{-1} resolution. NIR measurements of SET1 were taken with an FT-NIR Buchi (*NIRFlex N-500*) using standard glass Petri dishes. Nine replicates of each sample were recorded, using three different glass Petri dishes. NIR measurement of SET2 were taken with a FT NIR Thermo Scientific (*AntarisII FT-NIR Analyzer*) using a quartz sample cup spinner. In this case, three replicates of each sample were recorded. For both sets, the average spectrum was calculated for each sample and used for the multivariate analysis.

Chemometrics analysis

Data analysis was performed using the chemometric package V-PARVUS[2]. Data were transformed into absorbance units, then a column centering scaling was applied to the NIR spectra of SET1 and SET2 separately. In this way, the differences related to the cultivar are retained while the systematic differences due to instrument or year changes are deleted.

Two segments of the signal, exactly from 10000 to 9000 and from 4200 to 4000 cm^{-1}, were removed since they were not informative. Thus, the NIR data matrix had 161 rows (samples) and 1199 columns (variables, absorbance at different wavenumbers). In order to get rid of the effect of particle size, scatter and multicollinearity, several mathematical transformations of NIR spectra were tested. The use of SNV and first derivative was effective in reducing the major factor preventing NIR spectral interpretation. For visualisation and evaluation of the information contained in the NIR signals, principal component analysis (PCA) was applied to the spectra of Taggiasca table olives of SET1, projecting the other samples (SET2 and the remaining olives of SET1) into the same space.

In order to build a class-model for the Taggiasca cultivar, UNEQ (Unequal Variances Class-Model)[3] and PFM (Potential Function Methods)[4] class-modelling techniques were applied to the scores on the first two PCs computed. SET1 spectra were used to develop the Taggiasca olive class-model, evaluating its sensitivity with Taggiasca samples and its specificity with the samples from different cultivars. SET2 spectra were used as an external test set in order to obtain a reliable estimate of model sensitivity (percentage of SET2 Taggiasca samples accepted) and model specificity (percentage of SET2 different samples rejected).

Results and Discussion

PCA was applied intitally to the raw spectra to visualise differences among SET1 and SET2 samples. The scores on PC1-PC2 of the autoscaled data (93.89% of the total variance) show a clear separation between SET1 and SET2 (Figure 2A). In this case, olive discrimination on the basis of the cultivar is only partially detectable. This dataset separation disappears after a column centering scaling was applied to the NIR spectra of SET1 and SET2 separately. The score plot on PC1-PC2 (91.10% of the total variance) after scaling shows an elimination of the systematic differences between the two datasets retaining only the differences due to the olive cultivar (Figure 2B). Olives from cultivar Taggiasca (labeled as T) are well separated in the upper left of the score plot, without significant differences between samples of SET1 (T blue) and SET2 (T green).

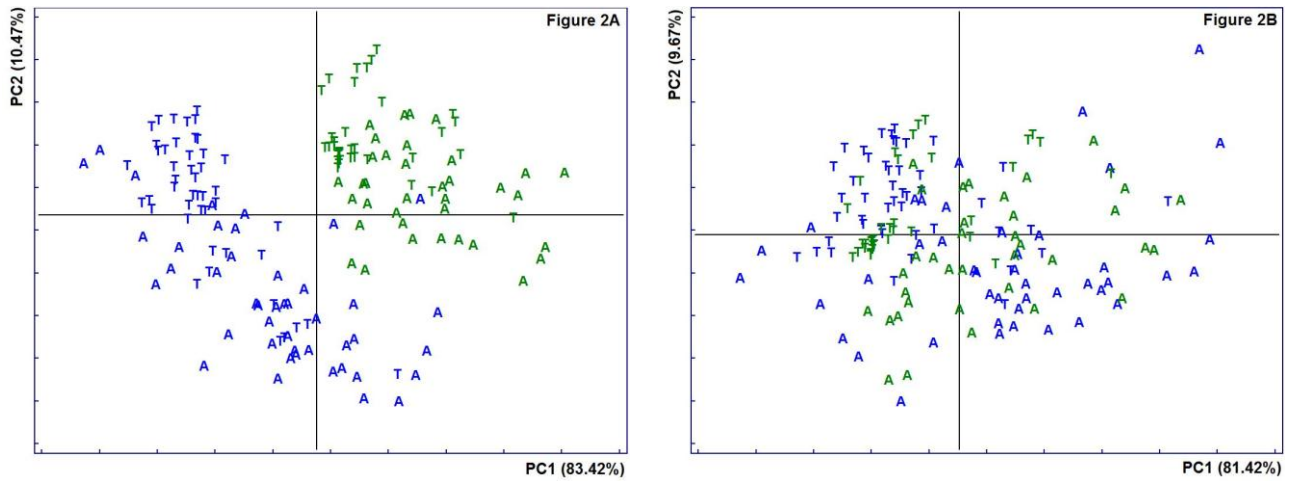

Figure 2. Score plots on PC1 and PC2 of the autoscaled data obtained by raw NIR spectra without category scaling **(A)** and with category scaling **(B)**. Samples are represented by their class symbol: SET1 = blue; SET2 = green; T = Taggiasca; A = other cultivar.

Reference paper as:
L. Bagnasco, C. Casolino and L. Medini (2012).Near infrared spectroscopy and class-modelling techniques to discriminate Taggiasca table olives, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 119-122.

After elimination of the non-informative part of the spectrum, the application of SNV and computation of the first derivative of the signals, PCA was applied as a display method. The PCs have been computed only on the NIR spectra of the samples belonging to the SET1 Taggiasca table olives, projecting the other samples (SET2 and the remaining olives of SET1) in the same space. The score plot on the PC1-PC2 of the autoscaled data (73.64% of total variance) shows an increase in the separation between the samples on the basis of the olive cultivar (Figure 3). Most of the olives from Taggiasca cultivar (T blue) appear at high values of PC1 and Taggiasca samples of SET2 (T green) are projected in the same region of the score plot.

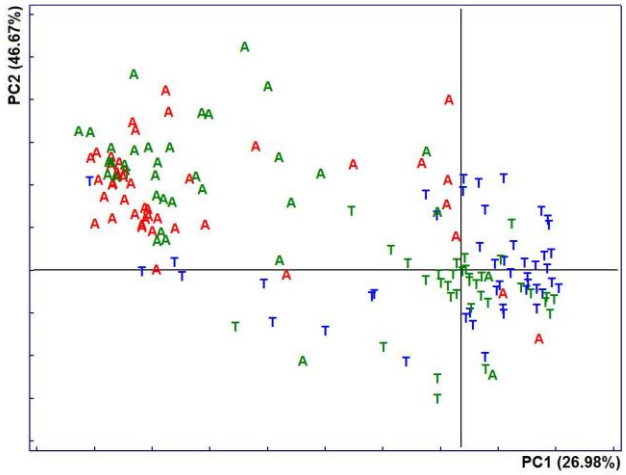

Figure 3. Score plot on PC1 and PC2 obtained by pre-treated NIR spectra corresponding to Taggiasca table olives of SET1. In the same space are projected also the other samples (SET2 and the remaining olives of SET1).
Samples are represented by their class symbol: SET1: T blue = Taggiasca; A red = other cultivar; SET2: T green = Taggiasca; A green = other cultivar.

A class-modelling study was performed using the scores on the first two PCs. In Table 1, the UNEQ and PFM results are listed as sensitivity and specificity of the respective models related to the Taggiasca olive samples at a confidence level of 90%. Predictions on the external test set are also reported in the same table. The same results are visualised by the Coomans plot[5] in Figure 4.

Table 1. UNEQ and PFM results.

			EXTERNAL TEST SET	
Class-modelling technique	Sensitivity %	Specificity %	Sensitivity %	Specificity %
UNEQ	91.30	84.09	91.43	88.89
PFM	97.83	72.73	94.29	69.44

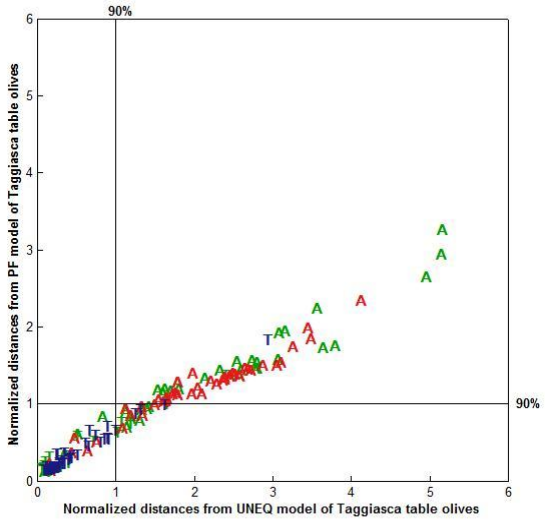

Figure 4. Coomans plot of the distances from the models of class 'Taggiasca' obtained by UNEQ and PFM.
Samples are represented by their class symbol: SET1: T blue = Taggiasca; A red = other cultivar; SET2: T green = Taggiasca; A green = other cultivar.

Reference paper as:
L. Bagnasco, C. Casolino and L. Medini (2012).Near infrared spectroscopy and class-modelling techniques to discriminate Taggiasca table olives, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 119-122.

Most samples of Taggiasca olives (labeled as T blue for SET1 and T green for SET2 external test set) are projected in the left bottom small square that represents the overlapping space between the two models. Objects in this space fit both models. Similarly, most samples belonging to other cultivar (labeled as A red for SET1 and A green for SET2 external test set) are in the right top large square, the space of the outliers, objects rejected by both the models. It is possible to notice the high sensitivity of PFM model in the Coomans plot: a lot of Taggiasca samples present a PFM model distance less than confidence level, corresponding to a UNEQ model distance exceeding the confidence level. This behavior of the PFM model is found also in the minor specificity with a higher numbers of olives belonging to the other cultivar accepted by the Taggiasca PFM model. This major modelling quality of PF technique is due to its characteristics: potential functions evaluate the probability function by using the local density of the objects instead of location and dispersion parameters of the class as UNEQ. So, the probability function is calculated as the sum of the individual contribution of the objects in the training set.

Conclusion

In this study the application of NIR spectroscopy as innovative tool to the identification of Taggiasca olive cultivar was presented. The class models obtained with UNEQ and PFM achieve excellent sensitivity and specificity, in agreement with preliminary work[1]. These results confirm the potential of NIR spectroscopy, coupled with multivariate analysis, to discriminate Taggiasca table olives against other cultivar. Moreover, these good results directed our efforts towards the study of repeatability and reproducibility of the method. The class models obtained seem to be useful to discriminate Taggiasca olives belonging to the next crop analyzed by different NIR spectrometer. This information is very interesting to develop a validated NIRS method to exploit and to certify the ligurian table olives.

Acknowledgements

Authors are grateful to the “Italian Society for NIR Infrared Spectroscopy (SISNIR)” for financially supporting the participation to the 15th International Conference on NIR Infrared Spectroscopy.

References

1. M. Casale, P. Zunin, M.E. Cosulich, E. Pistarino, P. Perego and S. Lanteri, *Food Chem.* **122**, 1261-1265 (2010).
2. M. Forina, S. Lanteri, C. Armanino, M.C. Casolino, M. Casale and P. Oliveri, PARVUS, available at http://www.parvus.unige.it (2011).
3. M.P. Derdre and D.L. Massart, *Anal.Chim. Acta* **184,** 33-51 (1986).
4. M. Forina, C. Armanino, R. Leardi, G. Drava, *J. Chemom.* **5**, 435-453 (1991).
5. D. Coomans, PhD Thesis, Vrije Univesiteit Brussels (1982).

Reference paper as:
L. Bagnasco, C. Casolino and L. Medini (2012).Near infrared spectroscopy and class-modelling techniques to discriminate Taggiasca table olives, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 119-122.