Controlling protected designation of origin of wine by Raman spectroscopy

Elsevier Editorial System(tm) for Food Chemistry

Manuscript Draft

Manuscript Number: FOODCHEM-D-15-02512R2

Title: Controlling Protected Designation of Origin of wine by Raman Spectroscopy

Article Type: Research Article (max 7,500 words)

Keywords: wine; raman spectroscopy; food traceability; chemometrics

Corresponding Author: Mrs. Luisa Mandrile,

Corresponding Author's Institution: INRIM

First Author: Luisa Mandrile

Order of Authors: Luisa Mandrile; Andrea M Giovannozzi, Ph.D; Giuseppe Zeppa, Professor; Andrea M Rossi, Ph.D

Abstract: In this paper, a Fourier-Transform Raman spectroscopy method, to authenticate the provenience of wine, for food traceability applications was developed. In particular, due to the specific chemical fingerprint of Raman spectra, it was possible to discriminate different wines produced in the Piedmont area (North West Italy) in accordance with i) grape varieties, ii) production area and iii) ageing time. In order to create a consistent training set, more than 300 samples from tens of different producers were analyzed, and a chemometric treatment of raw spectra was applied. A discriminant analysis method was employed in the classification procedures, providing a classification capability (percentage of correct answers) of 90 % for validation of grape analysis and geographical area provenance, and a classification capability of 84 % for ageing time classification. The present methodology was applied successfully to samples without any preliminary treatment of the sample, providing a response in a short time.

***Highlights (for review)**

- Designation of origin of wines can be controlled through Raman spectroscopy;
- Grape cultivar, provenience and ageing time of wines is determined through Raman;
- A rapid, sensitive and non-destructive method for wine analysis is proposed;
- Discriminant Analysis is applied to spectral data for wine classification.

Controlling Protected Designation of Origin of wine by Raman Spectroscopy

Luisa Mandrile[a,c] , Giuseppe Zeppa[b], Andrea Mario Giovannozzi[c] and Andrea Mario Rossi[c]

[a] *Department of Drug Science and Technology, Università degli Studi di Torino, Via Giuria 9, 10125, Torino, Italy l.mandrile@inrim.it*

[b] *Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA) - Microbiologia agraria e Tecnologie alimentari, Largo Paolo Braccini 2, 10095 Grugliasco (TO), Italy giuseppe.zeppa@unito.it*

[c] *Quality of life Division, Food Metrology program, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce, 91 10135, Torino, Italy a.rossi@inrim.it , a.giovannozzi@inrim.it*

[*] *Corresponding author: Andrea Mario Rossi, tel +39 011 3919342; fax +39 011 346384 e-mail a.rossi@inrim.it*

Keywords: wine, raman spectroscopy, food traceability, chemometrics, fingerprint

Abstract

In this paper, a Fourier Transform Raman spectroscopy method, to authenticate the provenience of wine, for food traceability applications was developed. In particular, due to the specific chemical fingerprint of the Raman spectrum, it was possible to discriminate different wines produced in the Piedmont area (North West Italy) in accordance with i) grape varieties, ii) production area and iii) ageing time. In order to create a consistent training set, more than 300 samples from tens of different producers were analyzed, and a chemometric treatment of raw spectra was applied. A discriminant analysis method was employed in the classification procedures, providing a classification capability (percentage of correct answers) of 90 % for validation of grape analysis and geographical area provenance, and a classification capability of 84 % for ageing time classification. The present

methodology was applied successfully to raw materials without any preliminary treatment of the sample, providing a response in a very short time.

Main text

1. Introduction

In order to preserve the quality of food products from particular geographical areas, and to protect consumers against imitations and false information, the European Commission has defined, via Regulations 1151/2012, the designations: Traditional Specialty Guaranteed (TSG), Protected Designation of Origin (PDO) and Protected Geographical Indication (PGI) (Regulation (Eu) No 1151/2012 Of The European Parliament And Of The Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs). Quality labels have an important role in consumer behavior and give confidence about the origins and the quality of food. Label assignment is an important market claim and represents a valuable weapon to attest and justify the economic value of products. Traceability has become a very relevant concept in association with edible products and represents an essential tool to enhance traders and consumers' confidence in the safety, quality, and authenticity of food.

Unfortunately, most of food traceability procedures involve tedious administrative documents, while scientific methodologies that identify the authenticity of food objectively are preferable. Accordingly, scientific research is focusing on the development of analytical methods for traceability to authenticate the geographical origin of foods (Peres, Barlett, Loiseau and Montet, 2007), with the aim of linking food products with distinctive features, such as ingredients, physical properties and production methods. Food traceability analysis are usually performed by means of several analytical techniques, such as mass spectrometry for isotope ratio determination (Durante, Baschieri, Bertacchini, 2015), DNA based techniques, such as polymerase chain reaction (PCR) (Pardo, 2014) and nuclear magnetic resonance spectrometry (NMR) (Mazzei, Francesca, Moschetti, Piccolo, 2010).

In the last two decades, stable isotope methodologies, based on gas chromatography-isotope ratio mass spectrometry (GC-IRMS) and GC-pyrolysis-IRMS (Fronza, Fuganti, Graselli, Reniero et al 1998; Adam, Bartels, Christoph, Stempfl 1995; Misselhorn, Grafahrend, 1990), have been applied successfully in quality control of wine following the establishment of an official wine database for stable isotope parameters (EU

regulations 2670/90, 2347/91 and 2348/91) (Rossmann, 2001). As reported by Breas et al. (Bréas, Reniero, Serrini, Martin and Rossmann, 1994), a classification of wines from different European countries can be achieved with $^{13}C/^{12}C$ analysis of ethanol and $^{18}O/^{16}O$ determination of water, underlining the importance of the photosynthetic pathway as well as the environmental and climatological conditions of the vineyard. Even if stable isotope methods provided consistent results, which could be used for routine analysis of wines, it is not always simple to find a physical, chemical or biochemical explanation for variations of isotope ratios in natural substances or to establish a relevant database for statistical evaluation.

DNA based technologies have also been exploited in this field due to their specificity in analysis, which is strictly associated with genotype (the inherited instructions that an organism carries within its genetic code), but these technologies inevitably miss the stochastic significant epigenetic differences accumulating over time across cells (Petronis, 2010). Dordevic et al. (2013) highlighted the need for new methods and better geographical discrimination between samples, demonstrating that multivariate methods are superior to univariate approaches. The NMR and vibrational spectroscopy techniques represent interesting alternatives or even complementary methods. Godelmann et al. (2013) analyzed about 600 German wines and demonstrated that ^{1}HNMR coupled with statistical data treatment could provide individual "fingerprints" for wine samples, which include information about variety, origin, vintage, physiological state, technological treatment, and other factors. The fusion of NMR profiling and stable isotope data for wine analysis has been reported in literature with good results (Monakhova et al. 2014). However, the main drawbacks of the cited techniques (i.e. MS, NMR and DNA based techniques) are related to the cost of instruments, extensive sample pre-treatments, and the duration of analysis, which often reduce the accuracy and precision of measurements. Since simple and rapid analytical methods are needed to meet the demands of European labeling legislation, vibrational spectroscopy is emerging as a new and powerful tool in authenticating food provenance.

Vibrational spectroscopy techniques usually provide non-destructive analysis of samples, rapid collection times with no or minimal sample pre-treatment, which reduce the total time of analysis and could support the development of reliable control procedures and screening methods for food traceability. Moreover, new modern, portable instruments with smart accessories have been developed, making these techniques more suitable for *in line* process monitoring and *in situ* analysis (Gallego, Guesalaga, Bordeu and Gonzàlez, 2011). These methods encompass absorption spectroscopy in the mid-infrared (MIR) and the near-infrared (NIR) for studying fundamental molecular vibrations and their harmonics (Bauer et al., 2008; Cozzolino, Dambergs, Janik, Cynkar,

& Gishen, 2006; Cozzolino, McCarthy, & Bartowsky, 2012,Cozzolino D., 2014), and absorption spectroscopy in the ultra-violet and visible (UV–vis) regions for probing electronic transitions (Acevedo, Jiménez, Maldonado, Domínguez, & Narváez, 2007; García-Jares & Médina, 1995; Harbertson & Spayd, 2006; Roig & Thomas, 2003; Urbano, Luque de Castro, Pérez, García-Olmo, & Gómez-Nieto, 2006). Raman spectroscopy, which is based on the inelastic scattering of a monochromatic light, also provides a characteristic spectroscopic pattern (i.e. "molecular fingerprint") of organic compounds based on the vibrational modes of chemical bonds (Li-Chan, Griffiths and Chalmers, 2010; Thygesen,. Løkke, Micklander and Engelsen, 2003). Moreover, Raman analysis can be easily done in aqueous media and through glass containers, because signals from both water and glass are very weak in the Raman spectrum (Schulz and Baranska, 2007; Yang, Irudayaraj 2001) and do not overlap with those from food components, such as proteins (Li-Cha, Nakai, Hirotsuka, 1994), lipids (Yang, Irudayaraj and Paradkar, 2005) and carbohydrates (Mathlouthi, Koenig, 1986), which are sensitive and specific.

Raman spectroscopy has demonstrated its value in food traceability for olive oil provenance and composition (Bernuy, Meurens, Mignolet and Larondelle, 2008), honey provenance (Özbalcia, Hakkı Boyacia, Topcua, Kadılarb, Tamerc**,** 2013; Paradkar and Irudayaray, 2001) and the authenticity of beers (Downey, 2009). As regards alcoholic beverages, Raman spectroscopy has been used for the quantification of the alcohol content in whisky, vodka and other spirituous beverages (Nordon, Mills, Burn, Cusick and Littlejohn 2005). The feasibility of exploiting Raman scattering to analyze white wines has also been investigated (Meneghini et. al., 2008). In particular, a recent work by Coralie et al. (2014) demonstrated that resonance condition of some chemical species present in wine, such as phenolic compounds, hydroxycinnamic acids and sugars, can be analyzed selectively using lasers at different wavelengths.

In this work, we evaluated the potential to use Raman spectroscopy, coupled with a chemometric data treatment, to discriminate different wines from the Piedmont area (North West Italy) in accordance with grape varieties, production area and ageing time. In particular, tests were performed on Nebbiolo, Dolcetto and Barbera wines, which were chosen for their wide distribution and their productive and economic relevance to the Italian wine market. The purpose of the work was to provide a statistically substantial classification method, based on a set of known responses (training set) through the chemometric treatment of data. The work scheme was structured on three levels: classification of wines in accordance with the (1) grapes used, (2) production area, and (3) age.

2. Material and Methods

2.1 Samples

315 commercial wines were obtained from different winemakers using Nebbiolo, Barbera and Dolcetto grapes. For each grape variety, wines from the different area and age were selected (Table 1). More than 10 Protected Designation of Origin (PDO) wines were examined. The number of samples for each PDO wine was different based on the winemaker and commercial dissemination and, inevitably, limited by the availability of samples. All the samples were furnished directly by the producers, and stored at +4°C until analysis.

Tab.1 – Distribution of wines examined in accordance with grape, PDO and production area

2.2 Raman measurements

Raman spectroscopy was performed with a Thermo Scientific NXR FT-Raman Module Nicolet Series™ equipped with an InGaAs detector (ThermoFisher Scientific, Waltham, USA), a CaF_2 beamsplitter and a 1064 nm laser line. Raman spectra were collected using a laser (power 0.9 W) in the spectral range 200 to 4000 cm^{-1} with a resolution of 4 cm^{-1}. 256 scans were collected to obtain S/N ratio higher than 15. Samples were analyzed in 4 ml glass vials positioned vertically on a motorized stage.

2.3 Multivariate Analysis.

The raw Raman spectra were subjected to discriminant analysis using TQ Analyst™ 8.0 software (ThermoFisher Scientific, Waltham, USA). Spectra were pre-processed using the Savitzky-Golay smoothing filter (Savitzky, Golay, 1964) to remove of as much noise as possible without unduly degrading the spectral information. The spectral range to be analyzed was selected in such a way that interference from random variability of spectra was minimized and did not generate spurious information in the classification model. Seven restricted spectral regions around Raman peaks were selected to optimize the classification result. The frequency regions of spectra that did not contain Raman peaks (e.g. 800-600 cm-1 and 2800-1800 cm-1) were excluded. In this way, worthless information was ignored and the best class separation was obtained. The number of PCs selected was a compromise between explained variance for each PC and the predictive capability of the model: when the cumulative variance reaches the *plateau*, further components do not provide any useful information and should be excluded so variables that represent only noise are not considered. Variables that explain only a small portion of the variability are not excluded if they improve significantly the classification capability of the model (% of

samples correctly classified). The chemometric models presented for wine classification were first validated through a leave-one-out cross-validation procedure during model optimization (mathematical pretreatment choice, selection of significant PCs, etc.). Finally, the optimized models were validated through a cross validation procedure using exclusions sets made up of five samples chosen randomly; the number of exclusion sets was proportional to the total number of calibration samples. This classification technique permits unknown distance to a class center in terms of Mahalanobis distance (Mahalanobis, 1936) to be calculated and each unknown samples to be assigned correctly. Md is based on the idea that it contains an auto-scaling process in and overcomes assumption about the spherical distribution of sample points around the center of mass; thus, non-spherical distributions can be described as well as spherical ones. In the generalized formula for Md, the observation are represented by x=(x1, x2, …xn) while μ=(μ 1, μ 2, … μ n) represents the observations' mean. The apex ^{T} indicates the transposed matrix (x- μ). S^{-1} is the inverse of the covariance matrix of the observations.

$$Md(x)=\sqrt{(x-\mu)^{T}S^{(-1)}(x-\mu)}$$

If an ellipsoidal distribution is considered, we would expect that the probability a test point belongs to the set depends not only on the distance from the center of mass but also the direction. (De Maesschalck,Jouan-Rimbaud, Massart, 2000).

The statistical reliability of results will be discussed case by case to assess the effective classification capability of the proposed Raman method, even if an external set dedicated to test set validation was not available. The work scheme of this study was divided into three consecutive steps: discrimination according to (i) grape, (ii) production area, and (iii) age.

3. Results and Discussion

Food systems are dynamic, chemically complex and, generally, heterogeneous matrices containing large numbers of biological molecules. The chemical specificity, ease of sampling, speed, and non-destructive nature of FT-Raman spectroscopy makes it an attractive tool for food analysis. Chemical specificity of the Raman technique relies on the fact that different molecular bonds or groups of chemical bonds are identified by characteristic frequency-shifts in incident light (Figure1). For this reason, the first step in compositional analysis of wine using FT-Raman is attribution of characteristic frequency shifts to vibrational modes of molecular bonds observed in spectra (Table 1S in supplementary information).

As Figure 1 shows, a large band ascribed to OH stretching at 3350 cm^{-1} was clearly visible in all the spectra analyzed. In addition, a minor band related to OH bending at 1700-1500 cm^{-1} was observed. The group of peaks between 3000-2800 cm^{-1} is due to symmetric and asymmetric stretching of CH_x bonds. Several other characteristic peaks of ethanol are present at frequencies less than 1500 cm^{-1}. These are associated with several deformation modes of CH_x as reported in Tab 1S (Mammone, Sharma, Nicol, 1980). All peaks in the wine were shifted slightly in comparison with the pure ethanol peaks; this is due to the simultaneous presence of different organic species, such as glycerol, acetaldehyde, organic acids, and polyphenols including flavonoids and non-flavonoids. At 1630 cm^{-1} a low intensity band was present in the wine spectra. This band is characteristic of C=O stretching, a relatively inactive Raman vibration. The C=O peak observed could be attributed to several species present in the matrix (e.g. organic acids and flavonoids) the carbonyl groups of which are characterized by slightly different vibration frequencies. This, a quite broad signal was registered in this spectral region.

The samples analyzed were chosen with the aim of representing a wide selection of the wines, which were purchased from different producers. Numerous samples were requested to capture the variability in the system and obtain a representative dataset for multivariate calibration. Raman spectra of the different wines were very similar to each other, as it can be seen in Figure 1 where the spectra of Dolcetto, Barbera and Nebbiolo are compared. This explains why a univariate analysis would not be effective. It was decided a multivariate approach would be employed to ensure a more complete interpretation of characteristic patterns in the spectra.

Fig. 1

From an oenological point of view, the specific features of wine are the result of synergic effects involving several factors. The wine composition is very complex and the final organoleptic features are produced by the interaction of many chemicals, such as sugar, alcohol, acids and tannins; e.g. *total acidity* refers to the sour attributes of the wine, which are evaluated in relation to how well the acidity balances out sweetness. During the course of winemaking and in the finished wines, tartaric, malic, citric, acetic and lactic acids can have significant roles and together define the characteristic acidity of the wine (Bellman et al., 1979). In the same way, from a spectroscopic point of view, the final wine spectrum is the result of a synergic interaction of many factors and none can be regarded in isolation. The literature is poor in respect of interpretative analysis of Raman spectra from wine because of the complexity, and only chemometric analysis permits extraction of the more interesting information and selective parameters to distinguish and attest to the authenticity of wine products.

The chemometric approach used for the classification was a supervised classification method, which groups a set of objects in such a way that objects in the same group (called a class) are more similar to each other than those belonging to other classes. Training data were given as sets of spectra partitioned as suggested by the supervising method (Finley and Joachims, 2005). Different distance functions were used to evaluate distances between objects in the same class or the assignment of an unknown object to the correct class. In this case, Md was used, as described in detail in Materials and Methods. Applying this concept to the spectral data of wine, several classification models with good classification capability were obtained.

3.1 Discrimination in accordance with grape

Three classes of grapes (Nebbiolo, Barbera, Dolcetto) were selected. 185 Nebbiolo, 75 Barbera and 45 Dolcetto wine samples were subjected to Raman analysis to create a substantial training set. The *eigen analysis* attested that the selected 305 calibration standards contain sufficient variability for the method calibration. The spectral range was optimized as reported in Materials and Methods section. The optimized chemometric model shows a total variability of 99.34 % explained using 20 principal components (PCs); the number of principal components was optimized by considering the classification capability (%) (number of correctly classified samples during cross-validation) as a function of the PCs number. In particular, leave-one-out cross validation was performed reiteratively raising the number of PCs considered during each run, and the percentage of correctly classified samples was plotted as a function of PCs number (Figure 1S) as well as the variance explained corresponding to each PC. The plot reported in Figure 1S was used to determine the ideal number of PCs, which corresponded to 20. In order to avoid the over-fitting of data, components that did not contribute significantly to cumulative variance, and did not provide useful information for classification, were excluded because they dealt exclusively with experimental noise.

As Figure 2 shows, the best optimized method misclassified 13.1 % of 305 standards during leave-one-out cross validation. The clouds of points representing these three classes were dense, suggesting high homogeneity within each class. The three clouds were also very close to each other and overlapped partially, which was the cause of a misclassification percentage greater than 10 %. However, it should be taken into account that the discipline of some wine production allows a small percentage of other wines to be introduced (e.g. Barbera wine can contain up to 15 % of Nebbiolo grape); this might explain the closeness of sample classes, which also caused

misclassification. A cross validation test was performed (and repeated five times) to attest the real ability of the calibrated model to distinguish wines according to grape. 100 spectra (one third of the number of calibration standards per each class chosen randomly) were used in groups of five for cross validation of the model. During this leave-five-out validation, 86 ± 2 % of unknown samples provided a correct answer. Among the misclassified samples, 9 % belong to Barbera, 2 % belong to Dolcetto and 3 % belong to Nebbiolo, on average. It should also be noted that the percentage of misclassified samples during leave-five-out cross validation method was comparable with leave-one-out cross validation results (14 % of misclassified with 20 PCs) achieved during model optimization. Subsequently, 10 unknown Nebbiolo samples were used as a small test set that provided 90 % correct answers.

Fig. 2

The loadings profiles corresponding to PCs 1 to 10, which were the most interesting for a qualitative description, are shown in Figure 3. From careful analysis of them, it is possible to determine which organoleptic and compositional features were responsible for classification. However, it must be taken into account that a synergic interaction of variables led to the class separation and none can be considered separately. For example, the alcohol content of a wine is a key parameter for its oenological characterization and plays an important role in the spectroscopic analysis in order to depict a faithful portrait of each sample. The ethanol Raman peaks are the easiest to be identified in Raman spectra and these can be identified in most calculated PCs as well. This aspect plays a crucial role in wine classification.

Sugar content is another important feature that can help in classification. Since the sugar content of a wine depends on the advancement of the alcoholic fermentation, a well-founded hypothesis is the negative correlation between sugar and the alcohol contents. PC8 and PC9 revealed a significant variability in data observed around 3500-500 cm^{-1}, where carbohydrates peaks are typically found. The scores plot, built in accordance with these PCs, revealed the carbohydrate content varied from sample to sample without any correlation with the Dolcetto, Barbera or Nebbiolo classes. The difficulty of defining coherent variability in this case lies in the fact that all the samples considered were dry wines.

Another important parameter in the Raman characterization of a biological matrix is the effect of fluorescence. The colored substances in wine, such as anthocyanins and polyphenols in general, are directly related to the fluorescence effect observed during spectra acquisition. Fluorescence is, generally, an undesirable effect in Raman analysis because of the risk of disguising interesting signals in the spectrum. It can also influence the

statistical analysis of wine spectra during classification. Indeed, the baseline slope of PC1, and the wide band around 2000 and 1200 cm^{-1} in PC6 and PC7, attest to the fact that fluorescence represents a significant variable in the system examined. This behavior is even more evident when looking at the disposition of data clouds as a function of PCs influenced by fluorescence, where it can be seen clearly that fluorescence is a significant variable. However, the classification of wines is not impaired by fluorescence, the success of which is not only in satisfactory modeling of training sets but also by external validation sets.

Fig. 3

Our data revealed that synergic interactions among variables represented the key to solving an apparently complicated problem. It was not possible to describe the data if the variables were considered independently but taken together good separation was achieved.

Also, dual class models were optimized and, as it turned out, the most difficult wines to separate were Barbera and Dolcetto whereas Nebbiolo sets formed a specific well separated class.

3.2 Discrimination in accordance with production area

We also demonstrated the capability of Raman spectroscopy to separate wines according to grape, and developed a method to attest to geographical provenance within the same grape class. In order to understand the importance of geographical area of produced, it is good to know that if a technical expression describing particular combination of elements, such as climate, soil and regional knowhow of winemaker, which defines the uniqueness and unrepeatability that characterize a labeled wine (*Terroir*) exists. The study focused on two wines in particular, Dolcetto and Nebbiolo. Within the Dolcetto class (i) Dolcetto d'Alba Doc and (ii) Dolcetto di Dogliani Docg were chosen. The production area of Dolcetto di Dogliani is situated in the southernmost part of Piedmont whereas the Dolcetto d'Alba region is situated in a northern part of the Langhe territory as shown in Figure 4 a. Dolcetto is highest in the Langhe territory (from 250 to 700 m above sea level) and characterized by a fresh climate because of its proximity to the Appenino Ligure and Alpi Marittime mountains ranges. This represents the best climate condition for Dolcetto wine production because it slows the grape maturation process.

In this geographic area, the soil varies from generous red soil to sandy and dry soil (www.regione.piemonte.it); the best soil type for the Dolcetto production is white, deep, clayey and calcareous. Dolcetto di Dogliani and

Dolcetto d'Alba wines are produced according to a strict discipline that declares, in a very precise way, the mandatory geographical area and the variety of grape permitted. Also, the winemaking procedure and the final organoleptic features are usually controlled through a qualified panel test. Dolcetto d'Alba and Dolcetto di Dogliani wines have strong sensory features and even an expert sommelier might find it difficult to distinguish the geographic origins of the two by taste. The Raman analysis coupled with chemometrics provided a good identification method for classification of the wines according to the area of production, as shown in the Cooman's plot in Figure 4 a.

For Nebbiolo wine, two classes were also set: (i) Langhe (including Nebbiolo d'Alba, Barolo, Barbaresco); (ii) Novara&Carema (including Colline Novaresi, Coste della Sesia, Ghemme, Gattinara and Carema). The geographical areas involved are shown in the Piedmont map in the inset of Fig 4 b. Nebbiolo wine is an ancient red mono-vine wine. Its history in Piedmont region predates the 17^{th} century, and it has thrived because of adaptation to cold climates (www.langhevini.it). The geographic area designated for production of Nebbiolo is also clearly specified. The soil should be clayey, calcareous and acidic or a combination of the three; the territory must by hilly (at least 650 m above sea level) and sunny (www.regione.piemonte.it). The chemometric analysis of Nebbiolo spectra enables classification of Nebbiolo from Langhe and from Novara & Carema, as shown in the Cooman's plot in Figure 4b.

As stated previously, the whole spectra for the different wines are responsible for class separation. The number of PCs considered (6 for Dolcetto classification and 14 for Nebbiolo classification) represented the best compromise between explained variance and classification capability, as discussed in Section 3.1 (Figure 1S b and Figure 1S c, available in supplementary information). Again, the only way to achieve the desired results was to use multivariate approach. Appreciable classification capability (> 90 %) was obtained for the two classification models, and the low number of misclassified standards suggests Raman spectroscopy is able to discriminate wine provenance when a consistent calibration is performed.

Fig. 4

The cross validation test provided satisfactory results for both models. Ten samples were chosen randomly (ca. 30 % of the calibration samples from each class) and used in pairs to validate the Dolcetto model with an error of 8 %; all of the misclassified samples belonged to "Dolcetto d'Alba". The leave-five-out cross-validation for Nebbiolo was performed using 65 spectra, five-by-five chosen randomly with respect to the total in each class. In

this case, 7 % were misclassified. In particular, one of them was from Alba, while five were from the northern part of Piedmont (Novara&Carema class). The validation procedure was repeated five times for both DA methods attesting a standard deviation of classification capability of 1 % and 2 % respectively.

3.3 Discrimination in accordance with age

As a third step, the potential to 'recognize' aged from non-aged oenological products was investigated. Many wines improve in quality during barrel and bottle storage. Left too long, however, such wines begin to deteriorate. During the ageing period, acidity decreases, and further clarification and stabilization occur as well as the precipitation of undesirable substances, and complex compounds affecting flavor and aroma are formed. Wines are usually aged in wooden barrels made of oak, allowing oxygen to enter but preventing water and alcohol from escaping. Simple phenols are transformed during ageing into complex molecules formed by the condensation of proanthocyanidins and anthocyanins, which also explains the change of color of aged wines. As the wine ages, anthocyanins react with other acids and compounds, such as tannins, pyruvic acid and acetaldehyde, which change the color of the wine to "brick red" hues.

One of the most interesting comparisons that can be performed on Piedmont's wines concerns Barolo and Barbaresco wine. They are both produced with the Nebbiolo grape and follow a mono-grape strict production protocol. What makes a Barolo wine different from a Barbaresco wine is essentially the ageing time: Barbaresco is aged for at least 26 months whereas Barolo is aged for at least 38 months. In this study, 56 samples of Barolo and 24 samples of Barbaresco were analyzed using Raman spectroscopy and the data collected were processed by discriminant analysis, as previously described. The statistical separation of the two wines produced positive results when 9 PCs were considered, as shown in Figure 5.

Fig. 5

A cross validation of the calibrated model was performed. Spectra from unknown samples (30) were subjected to analysis in groups of five. The validation procedure was repeated five times and provided 84 ± 4 % correct answers, on average. Among the 16 % wrongly classified, 80 % were Barolo and 20 % were Barbaresco.

4. Conclusions

In this paper, it was shown that Raman spectroscopy coupled with chemometric analysis can play a role in the authentication of wine, providing positive results in the recognition of mono-vine wines in terms of grape

(validation test provided reliability of 93%), geographical provenance (reliability higher than 90%) and ageing time (reliability higher than 80%). One of the biggest advantages of the proposed method is the direct analysis of wine, through the glass container, without any pretreatment and purification process. These advantages, together with the speed of data collection, make Raman Spectroscopy particularly interesting for the prevention fraud and control of quality labels. The common drawbacks of Raman spectroscopy in analysis of food matrices, such as problems with interpretation, were overcome with user-friendly software that allow sophisticated chemometric methods to be elaborated using large amounts of data. The chemometric identification of variability between the different classes meant wines could be differentiated in accordance with grape, geographical origin, and ageing time using Raman spectrometry. A dedicated test set consisting of external samples was subjected to analysis in order to demonstrate the classification capability of the proposed method; this proof of principle aimed to show that a multivariate calibration procedure could provide consistent classification results when a substantial calibration set was subjected to spectroscopic analysis, even in a complex matrix. The more specific and user-friendly Raman analysis is, the more likely it is to be exploited by wine producers for certification. The application of Raman spectroscopy to distinguish a single producer will be the next challenge, with a higher impact in commercial field.

Acknowledgements

The present work has been supported by EMRP project "NEW02 Raman". EMRP is jointly founded by EMRP participating countries within EURAMET and the European Union.

References

Acevedo, F. J., Jiménez, J., Maldonado, S., Domínguez, E., & Narváez, A. (2007). Classification of wines produced in specific regions by UVvisible spectroscopy combined with support vector machines. *Journal of Agricultural and Food Chemistry*, 55(17), 6842–6849.

Adam L., Bartels W., Christoph N., Stempfl W. Band 2: Qualita¨tskontrolle und Analytik im Fachlabor. *Brennereianalytik*. Behr's, Hamburg (1995)

Bauer, R., Nieuwoudt, H., Bauer, F. F., Kossmann, J., Koch, K. R., & Esbensen, K. H. (2008). FTIR spectroscopy for grape and wine analysis. *Analytical Chemistry*, 80(5), 1371–1379.

Bellman R. B., Gallander J. F. (1979). Wine Deacidification. In Chichester C. O., Mrak E. M., Stewart G. F., *Advances in Food Research*, 25. Academic Press. p. 3. ISBN 0-12-016425-6. Retrieved 2009-08-04.

Bernuy B., Meurens M., Mignolet E. and Larondelle Y. (2008). Performance Comparison of UV and FT-Raman Spectroscopy in the Determination of Conjugated Linoleic Acids in Cow Milk Fat. *J. Agric. Food Chem.*, 56(4), 1159–1163.

Bréas O., Reniero F., Serrini G., Martin G. J. and Rossmann A. (1994). Isotope ratio mass spectrometry: Analysis of wines from different European Countries. *Rapid Communications in Mass Spectrometry*, 8(12) 967–970.

Coralie Martina, Jean-Luc Bruneela, François Guyonc, Bernard Médinac, Michael Jourdesd, Pierre-Louis Teissedred, François Guillaumea, (2014). Raman spectroscopy of white wines. *Analytical Methods*, DOI:10.1016/j.foodchem.2015.02.076

Cozzolino D., (2014). Sample preparation, sources of errors and future perspectives on the application of vibrational spectroscopy in the wine industry. *Journal of the Science of Food and Agriculture* DOI: 10.1002/jsfa.6733

Cozzolino, D., Dambergs, R., Janik, L., Cynkar, W., & Gishen, M. (2006). Review: Analysis of grapes and wine by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, DOI: 10.1255/jnirs.679

375 Cozzolino, D., McCarthy, J., & Bartowsky, E. (2012). Comparison of near infrared and mid infrared
376 spectroscopy to discriminate between wines produced by different *Oenococcus oeni* strains after malolactic
377 fermentation: A feasibility study. *Food Control*, DOI: 10.1016/j.foodcont.2012.01.003

378 De Maesschalck R., Jouan-Rimbaud D., Massart D.L. (2000). The Mahalanobis distance. *Chemometrics and*
379 *Intelligent Laboratory Systems*, 50, 1–18.

380 Downey G. (2009). Identity Confirmation of a Beer by Fingerprint and Profiling Techniques, Lecture in 5th
381 Annual Meeting of trace TRACE: New Methods and Systems for Confirming the origin of Food, Freising
382 (Munich), 1-3- April

383 Dordevic N., Wehrens R., Postma G.J., Buydens L.M., Camin F. (2012). Statistical methods for improving
384 verification of claims of origin for Italian wines based on stable isotope ratios. *Anal Chim Acta*. DOI:
385 10.1016/j.aca.2012.10.046.

386 Durante C., Baschieri C., Bertacchini L. (2015). An analytical approach to Sr isotope ratio determination in
387 Lambrusco wines for geographical traceability purposes. *Food chem*istry, DOI: 10.1016/j.foodchem.2014.10.086

388 T. Finley and T. Joachims, Supervised Clustering with Support Vector Machines, Proceedings of the
389 International Conference on Machine Learning (ICML), 2005.

390 Fronza G, Fuganti C., Graselli P., Reniero F., Guillou G., Breas O., Sada E., Rossmann A., Hermann A. J.
391 (1998), Determination of the ^{13}C Content of Glycerol Samples of Different Origin**.** *Agric. Food Chem*, 46, 477–
392 480.

393 Gallego A. L., Guesalaga A. R., Bordeu E. and Gonzàlez A. S. (2011). Rapid measurement of phenolic
394 compounds in red wine using Raman spectroscopy. *Instrumentation and Measurement*, DOI:
395 10.1109/TIM.2010.2051611

396 García-Jares, C., & Médina, B. (1995). Prediction of some physico-chemical parameters in red wines from
397 ultraviolet–visible spectra using a partial least squares model in latent variables. *Analyst*, 120(7), 1891–1896.

398 Rolf Godelmann, Fang Fang, Eberhard Humpfer, Birk Schütz, Melanie Bansbach, Hartmut Schäfer, and
399 Manfred Spraul, (2014). Targeted and Nontargeted Wine Analysis by 1H NMR Spectroscopy Combined with

400 Multivariate Statistical Analysis. Differentiation of Important Parameters: Grape Variety, Geographical Origin,
401 Year of Vintage. *J. Agric. Food Chem.* DOI: 10.1021/jf400800d

402 Yang H., Irudayaraj J. and Paradkar M. M. (2005). Discriminant analysis of edible oils fats by FTIR, FT – NIR
403 and FT Raman spectroscopy. *Food Chemistry*, 93(1), 25- 32.

404 Yang H., Irudayaraj J. (2001). Comparison of near-infrared, fourier transform-infrared, and fourier transform-
405 raman methods for determining olive pomace oil adulteration in extra virgin olive oil. *Journal of the American*
406 *Oil Chemists'Society*, 78(9) 889 – 895.

407 Li-Chan E.C.Y., Griffiths P. R. and Chalmers J. M. (2010). Applications of Vibrational Spectroscopy in Food
408 Science. John Wiley & Sons ISBN 978-0-470-74299-0

409 Li-Chan E., Nakai S., Hirotsuka M. (1994). Raman Spectroscopy as a probe of protein structure in food systems.
410 In *Protein Structure-Function Relationships in Foods*; Yada, R.Y., Jackman, R.L., Smith J.L., Eds, Blackie
411 Academic & Professional, Chapman & Hall Inc. London, England; pp163-197.

412 Monakhova Y.B. , Godelmann R. , Hermann A. , Kuballa T. , Cannet C. , Schäfer H. , Spraul M. , Rutledge
413 D.N.. (2014). Synergistic effect of the simultaneous chemometric analysis of ^{1}H NMR spectroscopic and stable
414 isotope (SNIF-NMR, ^{18}O, ^{13}C) data: application to wine analysis. *Anal Chim Acta*. DOI:
415 10.1016/j.aca.2014.05.005.

416 Mahalanobis P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of*
417 *Sciences of India*, 2(1), 49–55.

418 Mammone J.F., Sharma S.K., Nicol M. (1980). Raman spectra of methanol and ethanol at pressures up to 100
419 kbar. *Journal of Physical Chemistry*, 84(23), 3130 – 3134.

420 Mathlouthi M., Koenig J. L. (1986). Vibrational spectra of carbohydrates. *Advanced Carbohydrates Chemistry*
421 *and Biochemistry*, 44**,** 7-89.

422 Mazzei P., Francesca N., Moschetti G., Piccolo A. (2010). NMR spectroscopy evaluation of direct relationship
423 between soils and molecular composition of red wines from Aglianico grapes. *Analytica Chimica Acta*
424 DOI:10.1016/j.aca.2010.06.003

Meneghini, C., Caron, S., Proulx, A., Emond, F., Paradis, P., Pare, C., et al. (2008). Determination of ethanol concentration by Raman spectroscopy in liquid-core microstructured optical fiber. *IEEE Sensors Journal*, DOI: 10.1109/JSEN.2008.926172

Misselhorn K., Grafahrend W. (1990). Rohstoffnachweis bei hochgereinigtem Alkohol. *Branntweinwirtschaft*, 130, 70–73.

Nordon A., Mills A., Burn R. T., Cusick F. M. and Littlejohn D. (2005). Comparison of non-invasive NIR and Raman spectrometries for determination of alcohol content of spirits. *Analytica Chimica Acta*, 548(1-2), 148 – 158.

Özbalcia B., Hakkı Boyacia İ., Topcua A., Kadılarb C., Tamerc U. (2013). Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry*, DOI: 10.1016/j.foodchem.2012.09.064

Paradkar M. and Irudayaray J. (2001). Discrimination and classification of beet and cane sugars and their inverts in maple syrup by FT-Raman. *Applied Engineering in Agriculture*, 18, 379-383.

Pardo M. A. (2014). Evaluation of a dual-probe real time PCR system for detection of mandarin in commercial orange juice. *Food chemistry*, DOI: 10.1016/j.foodchem.2014.09.096

Peres, B., Barlett N., Loiseau G. and Montet D. (2007). Review of the current methods of analytical traceability allowing determination of the origin of foodstuffs. *Food Control*, 18(3), 228-235.

Petronis A. (2010). Epigenetics as a unifying principle in the etiology of complex traits and diseases. *Nature*, DOI:10.1038/nature09230

Roig, B., & Thomas, O. (2003). UV monitoring of sugars during wine making. *Carbohydrate Research*, DOI:10.1016/S0008-6215(02)00396-8

Rossmann A. (2001). Determination of Stable Isotope Ratios in Food Analysis. *Food Reviews International*, DOI:10.1081/FRI-100104704

Savitzky A., Golay M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), 1627–39.

Schulz H. and Baranska M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy*, DOI: 10.1016/j.vibspec.2006.06.001.

Socrates G., *Infrared and Raman Characteristic Group Frequencies: Tables and Charts*, 3rd Edition, Wiley & Sons (2004) ISBN: 978-0-470-09307-8

Thygesen L. G., Løkke M. M E., Micklander and Engelsen S. B. (2003). Vibrational microspectroscopy of food. Raman vs. FT-IR. *Trends in Food Science & Technology*, DOI:10.1016/S0924-2244(02)00243-1

Urbano, M., Luque de Castro, M. D., Pérez, P. M., García-Olmo, J., & Gómez-Nieto, M. A. (2006). Ultraviolet visible spectroscopy and pattern recognition methods for differentiation and classification of wines. *Food Chemistry*, DOI:10.1016/j.foodchem.2005.05.001

European Commission, Directorate-General for Agriculture Food Quality Policy in the European Union, "Protection of geographical indication, Designation of Origins and certificates of Specific Character for Agricultural products and Food-stuffs", Working document of the commission services, Guide to community regulation. 2nd edition, August 2004, 46 (2004) http://ec.europa.eu/agriculture/publi/gi/broch_en.pdf (15/02/10)

EU Agricultural Product Quality Policy (2010) http://ec.europa.eu/agriculture/quality/ (15/02/10)

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:343:0001:0029:en:PDF

Regione.piemonte.it
http://www.regione.piemonte.it/agri/politiche_agricole/viticoltura/dwd/disciplinari/dolcettoalba.pdf (last access 18/12/2015)

http://www.regione.piemonte.it/agri/politiche_agricole/viticoltura/dwd/vitigni/varieta_cloni/nebbiolo_descr.pdf
(last access 18/12/2015)

langhevini.it http://www.langhevini.it/pagine/ita/vitigni/nebbiolo.lasso (last access 18/12/2015)

FIGURE CAPTIONS

Figure 1–: Dolcetto d'Alba PDO (100% Dolcetto grape) (green spectrum), Barbera d'Alba PDO (minimum 85% Barbera grape) (red spectrum) and Barolo PDO (100% Nebbiolo grape) (black spectrum).

Figure 2– Cooman's plot for Nebbiolo, Barbera, Dolcetto classification model calculated using Discriminant Analysis.

Figure 3– Loadings profiles of the first 10 PCs of the Nebbiolo, Barbera, Dolcetto classification model calculated through discriminant analysis.

Figure 4– a) Geographical representation of Dolcetto d'Alba and Dolcetto di Dogliani wine production areas. Cooman's plot and statistical data of DA calibration. b) Geographical representation of Nebbiolo d'Alba and Nebbiolo di Novara & Carema wine production areas. Cooman's plot and statistical data of DA calibration.

Figure 5– Cooman's plot of Barolo and Barbaresco classification model and statistical results of calibration.

Table 1

Grape	Denomination	Ampelographic origin	Production Area	Number of samples
Nebbiolo	Barbaresco	100% Nebbiolo	Langhe	24
Nebbiolo	Barolo	100% Nebbiolo		56
Nebbiolo	Nebbiolo d'Alba	100% Nebbiolo		27
Nebbiolo	Nebbiolo Langhe	100% Nebbiolo		
Nebbiolo	Colline Novaresi Nebbiolo	100% Nebbiolo	North Piedmont (Novara)	33
Nebbiolo	Coste della Sesia Nebbiolo	100% Nebbiolo		2
Nebbiolo	Ghemme	100% Nebbiolo		10
Nebbiolo	Gattinara	100% Nebbiolo		12
Nebbiolo	Carema	100% Nebbiolo		25
Nebbiolo	Lessona	100% Nebbiolo		3
Nebbiolo	Canavese	100% Nebbiolo	Canavese	3
Barbera	Barbera d'Alba	85-100% Barbera 0-15% Nebbiolo	Langhe	50
Barbera	Barbera d'Alba Superiore	85-100% Barbera 0-15% Nebbiolo	Langhe	14
Barbera	Various (Asti, Pinerolo, Novara)	85-100% Barbera 0-15% Nebbiolo	North Piedmont	11
Dolcetto	Dolcetto d'Alba	100% Dolcetto	Langhe	16
Dolcetto	Dolcetto di Dogliani	100% Dolcetto	Dogliani	11
Dolcetto	Dolcetto di Diano d'Alba	100% Dolcetto	Langhe	18

Figure 1

Click here to download high resolution image

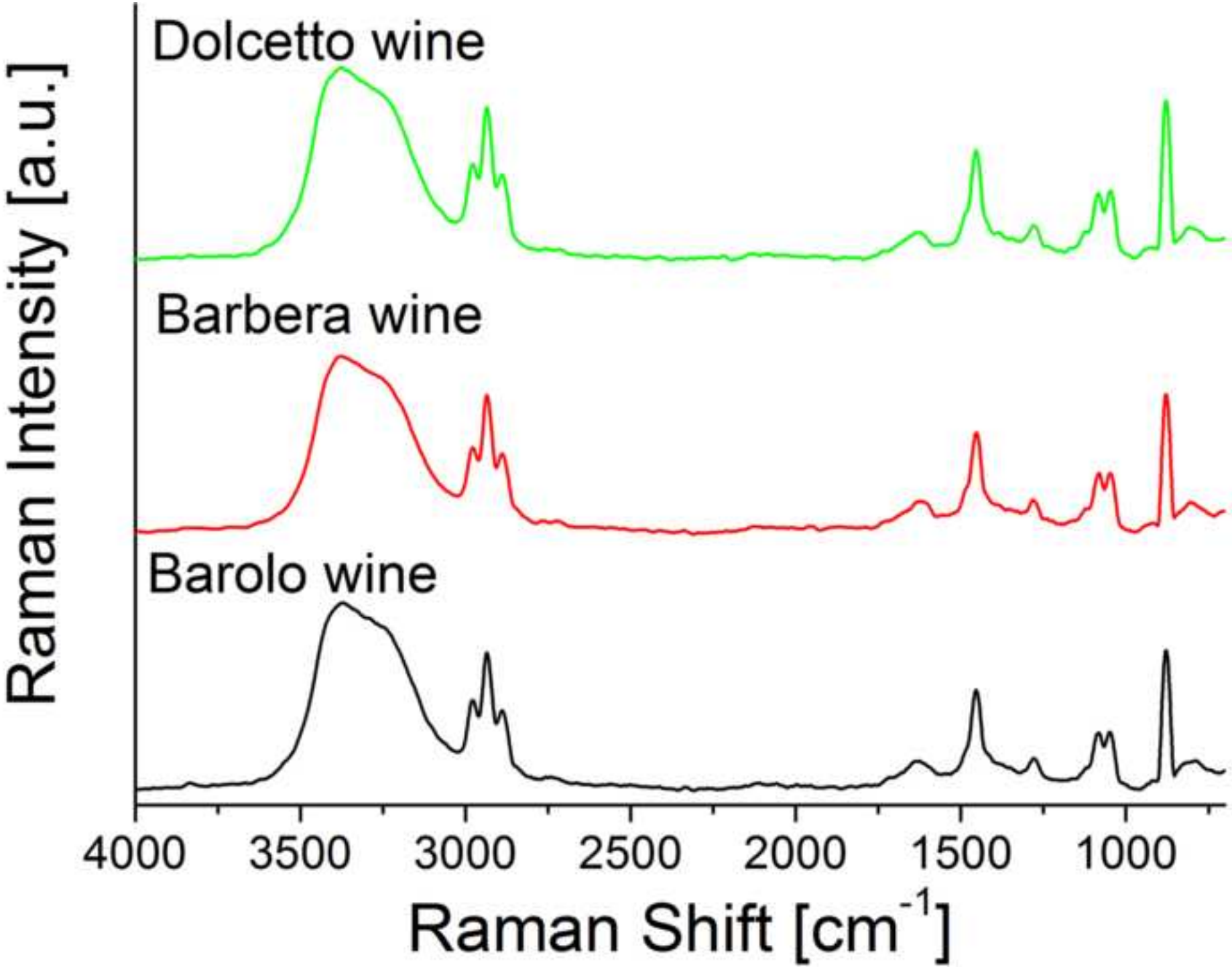
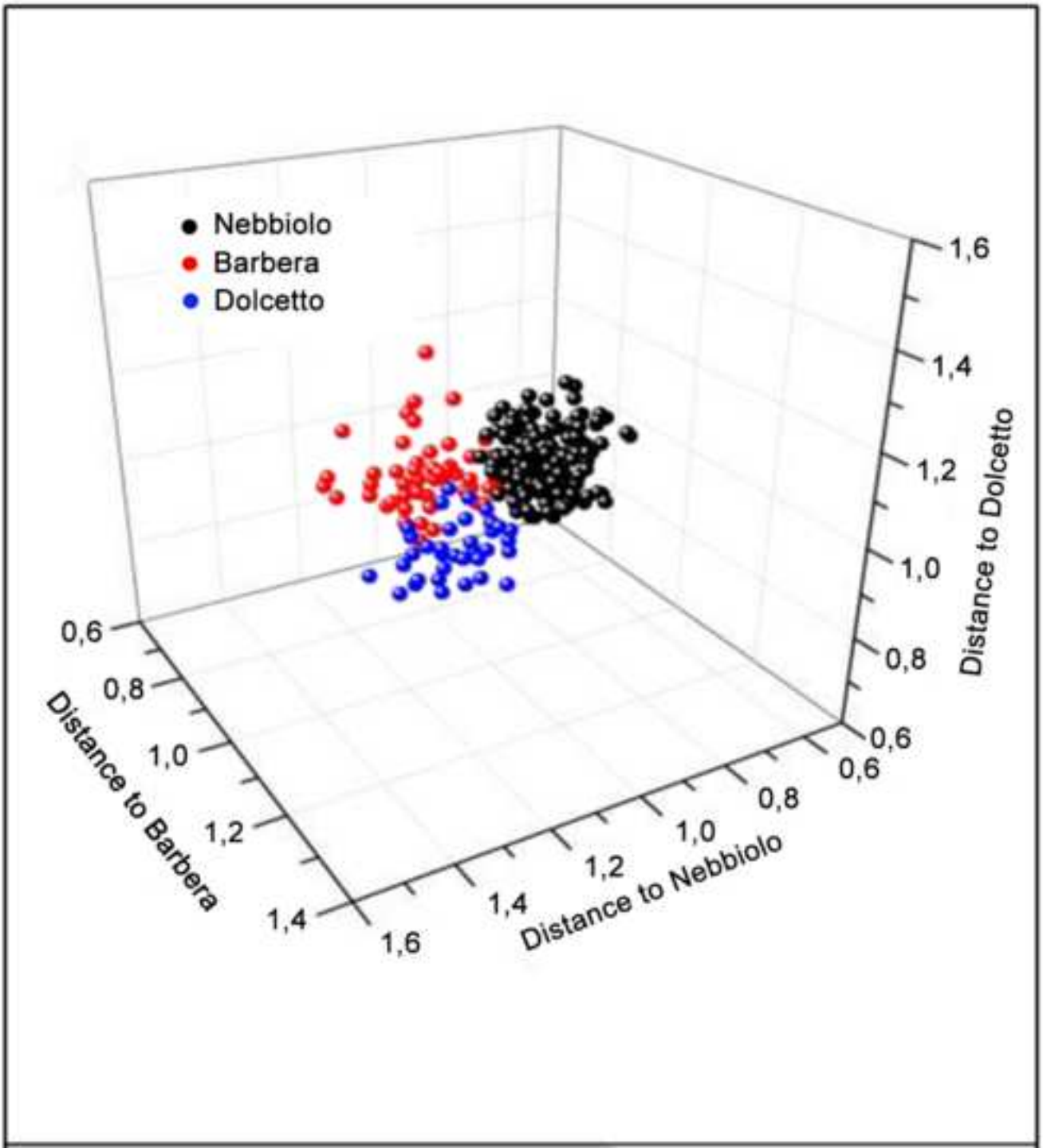


Figure 2

Click here to download high resolution image



N° of samples per class	Nebbiolo: 185 Barbera: 75 Dolcetto: 45
N° PCs	20
Cumulative expl. Var.	99.3 %
Predictive capability	86 ± 2 %

Figure 3

Click here to download high resolution image

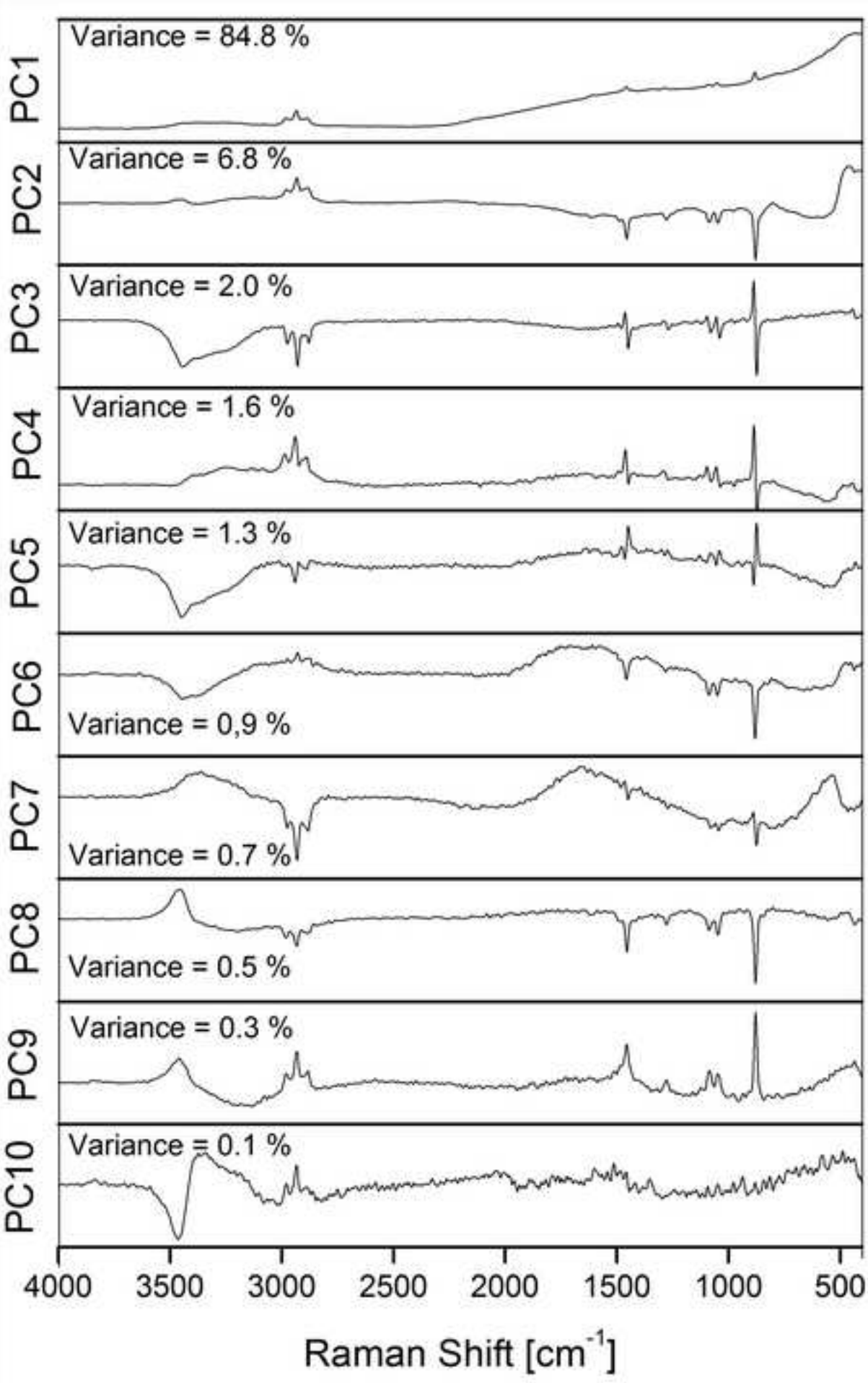
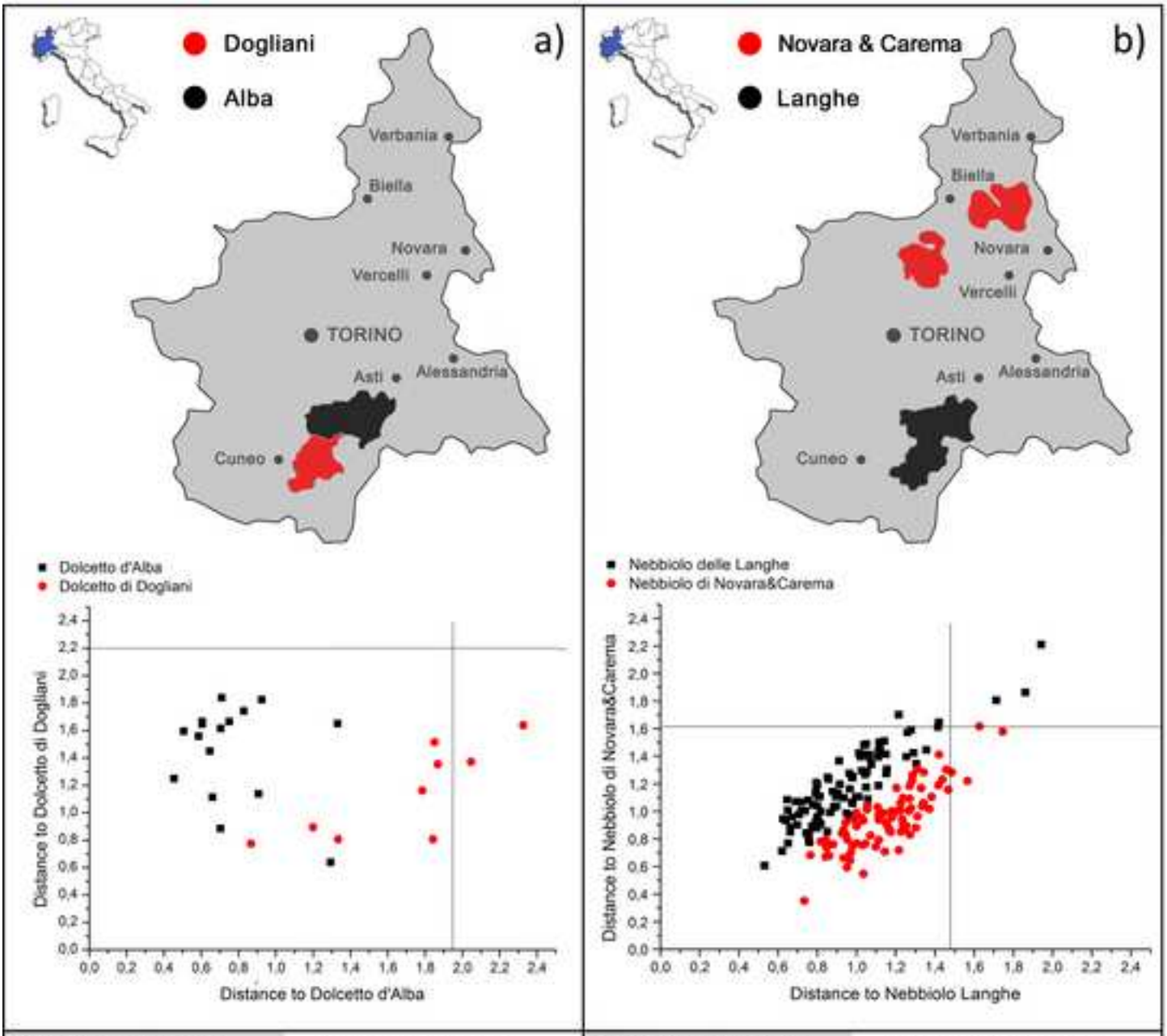


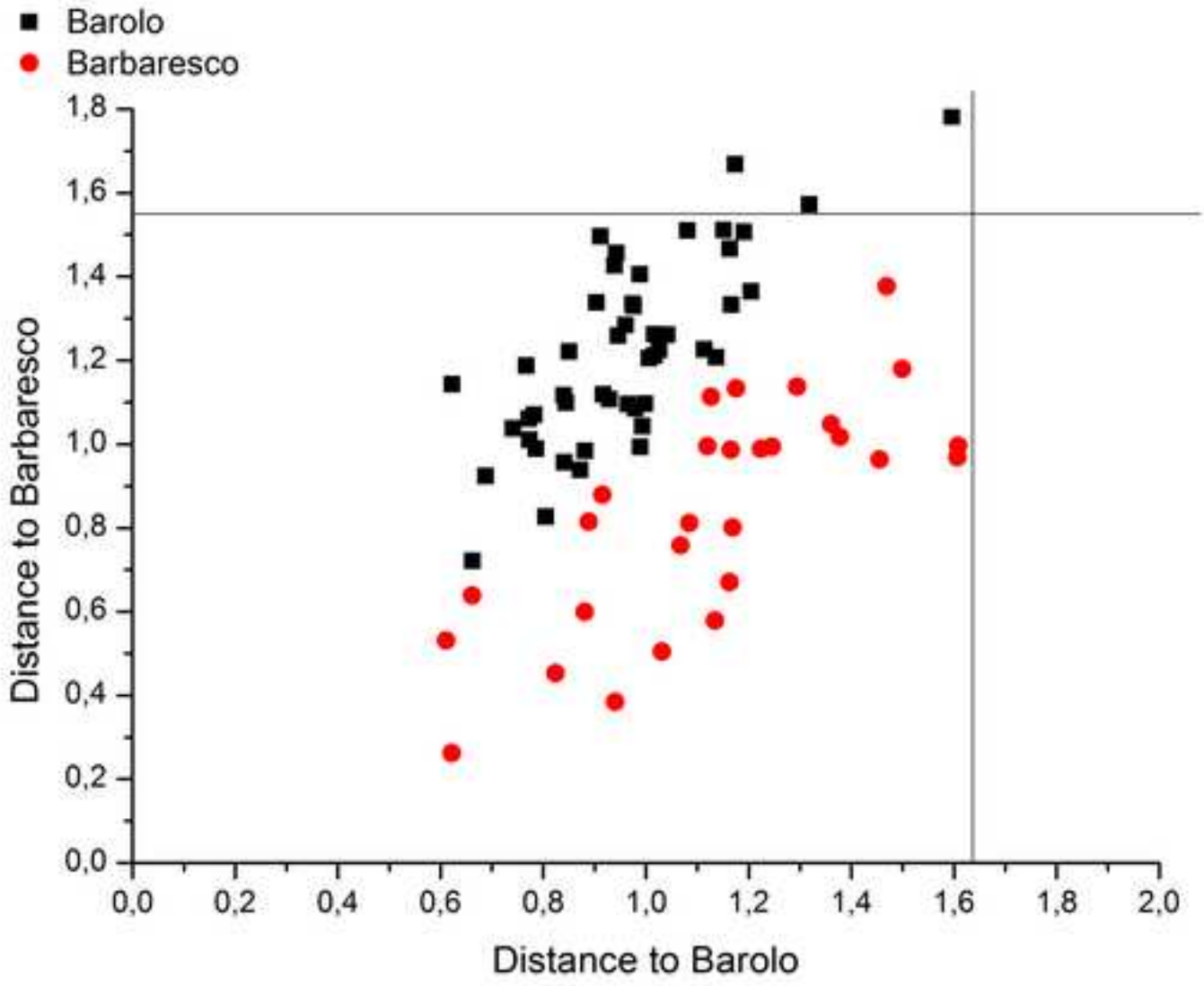
Figure 4

Click here to download high resolution image



N° of samples per class	Dolcetto Alba: 16 Dolcetto Dogliani: 11	**N° of samples per class**	Nebbiolo Alba: 107 Carema&Novara: 85
N° PCs	6	**N° PCs**	14
Cumulative expl. Var.	99.5 %	**Cumulative expl. Var.**	99.0 %
Predictive capability	92 ± 1 %	**Predictive capability**	93 ± 2 %

Figure 5
Click here to download high resolution image



N° of samples per class	Barolo: 56 Barbaresco: 24
N° PCs	9
Cumulative expl. Var.	97.7 %
Predictive capability	84 ± 4 %