Controlling Protected Designation of Origin of wine by Raman Spectroscopy

Controlling Wines' Protected Designation of Origin through 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 wine provenience 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. More than 300 samples from tens of different producers were analyzed in order to create a consistent training set 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% in validation for grape analysis and geographical area provenance, whereas a classification capability of 84% was obtained for ageing time

classification. The present methodology can be successfully applied on 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 coming from particular geographical areas and to protect consumers against imitations and false information, the European Commission defined, via Regulations 1151/2012, the labels 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 play an important role in consumer behavior and give confidence about the origins and the quality of food. Labels assignment is an important market claim and it represents a valuable weapon to attest and justify the economic value of alimentary products. Traceability has become a very relevant concept when it is associated to edible products and it represents an essential tool to enhance traders and consumers' confidence in the safety, quality, and authenticity of food.

Unfortunately, food traceability procedures mainly involve tedious administrative documents, while scientific methodologies that objectively identify the authenticity of food would be preferable. According to this, scientific research is now focused on developing analytical methods to authenticate the geographical origin of aliments in a metrological way (Peres, Barlett, Loiseau and Montet, 2007) with the aim of linking food products with its 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) were successfully applied in the field of wine quality control due to 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 done by means of $^{13}C/^{12}C$ 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 can be used for routine analysis of wines, it is not always simple to find a physical, chemical, or biochemical explanation for variations of the isotope ratios in natural substances and to establish a relevant database for statistical evaluation.

DNA based technologies were also exploited in this field due to their specificity of analysis because they are strictly connected to the genotype (the inherited instructions that an organism carries within its genetic code) but they inevitably miss the stochastic significant epigenetic differences accumulating over time across cells (Petronis, 2010). Dordevic N et. al (2013) highlighted the need of new methods for a better geographical discrimination between samples, demonstrating that multivariate methods are superior to univariate approaches. Interesting alternatives or even complementary methods are represented by the NMR and vibrational spectroscopy techniques. Godelmann et al. (2013) analyzed about 600 German wines and demonstrated that ^{1}HNMR coupled with statistical data treatment can provide individual "fingerprint" of a wine sample, which includes information about variety, origin, vintage, physiological state, technological treatment, and other. The data fusion of NMR profiling and stable isotope data for wine analysis was also investigated and reported in literature with good results (Monakhova et. al. 2014). However, the main drawbacks of all the cited techniques, i.e. MS, NMR and DNA based techniques, are related to the elevated cost of the instruments, extensive sample pre-treatments and long time of analysis which often reduce the accuracy and precision of measurements. Since simple and rapid analytical methods are needed to meet the developing European labeling legislation, vibrational spectroscopy is emerging as a new powerful tool in authenticating food provenance. Vibrational spectroscopy techniques usually provide non-destructive analysis of the samples, fast collection times with none or minimal sample pre-treatments that reduce the total time of analysis and could support the development of reliable control procedures and screening methods for food traceability. Moreover, new modern and portable instruments with smart accessories were developed in the last years 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) 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). Besides, Raman spectroscopy, which is based on the inelastic scattering of a monochromatic light, provides a characteristic spectroscopic pattern, i.e. "molecular fingerprint", of the analyzed 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 both water and glass signals are very weak in the Raman spectrum (Schulz and Baranska, 2007; Yang, Irudayaraj 2001) and do not overlap signals of macro food components, such as proteins (Li-Cha, Nakai, Hirotsuka, 1994), lipids (Yang, Irudayaraj and Paradkar, 2005) and carbohydrates (Mathlouthi, Koenig, 1986), which can be revealed in a sensitive and specific way.

Raman spectroscopy has recently 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 beers authenticity (Downey, 2009). As regards alcoholic beverages, Raman spectroscopy was 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 was also investigated in the last years (Meneghini et. al., 2008). In particular, a very recent work from Coralie et.al (2014) demonstrated that the resonance condition of some chemical species present in wine, such as phenolic compounds, hydroxycinnamic acids and sugars, can be reached using lasers with different wavelengths in order to detect and analyze these species selectively.

In this work we evaluated the possibility of using Raman spectroscopy coupled with a chemometric data treatment to discriminate different wines from 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 that were chosen for their diffusion and their productive/economic relevancy on Italian wine market. The purpose of the work is to provide a statistically substantial classification methods based on a set of known responses (training set) through the chemometric treatment of data. The work scheme was structured on three levels: at first, the classification of wines in accordance with the used grapes; secondly, the classification of

wines in accordance with the production area and finally the classification of wines in accordance with ageing time.

2. Material and Methods

2.1 Samples

Study was performed on 315 samples of commercial wines obtained from Nebbiolo, Barbera and Dolcetto grape by different winemakers. For each grape variety, wines of different area and ageing time were used (Table 1). Among the total sample pool, more than 10 Protected Designation of Origin (PDO) wines were examined. The number of samples for each PDO wine was different according to winemakers and diffusion and inevitably limited by the availability of samples. All samples were furnished directly by 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 of 0.9 W in a spectral range from 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 powered 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 preprocessed using Savitzky-Golay smoothing filter (Savitzky, Golay, 1964) in order 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 the interference of the random variability of spectra is minimized and it does not provide spurious information to the classification model. Seven restricted spectral regions around Raman peaks are selected in order to optimize the classification result.

The frequency regions of spectra which do not contain any Raman peak (e.g. 800-600 cm-1 and 2800-1800 cm-1) were excluded. In this way the worthless information is ignored and the best classes' separation is obtained. The suitable number of PCs to be considered is choosen as the best compromise between the explained variance of each PC and the predictive capability of the model: when the cumulative variance reaches the *plateau*, further components could not provide useful information and they should be excluded, so that variables that just represent noise are not considered for classification. Otherwise, some variables that explain a little variability shall not be excluded if they significantly improve the classification capability of the model (% of samples correctly classified). The presented chemometric models for wine classification were first validated through a leave-one-out cross-validation procedure during model optimization (mathematical pretreatment choise, significative PCs selection etc.). Finally, the optimized models were all validated through a cross validation procedure using exclusions sets made of five samples randomly chosen, the number of exclusion set is proportional to the total number of calibration samples. After calibration, this classification technique permits to calculate the unknown's distance to a class center in terms of Mahalanobis distance (Mahalanobis, 1936) and to assign each unknown samples to the correct class. The basic idea on which Md is based is the fact that it contains an autoscaling process in itself and it overcomes the assumption of spherical distribution of sample points about the center of mass; therefore non-spherical distributions can be described as well. In the generalized formula for Md reported below, 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, then we would expect that the probability of the test point to belong to the set does not only depend on the distance from the center of mass, but also on the direction. (De Maesschalck,Jouan-Rimbaud, Massart, 2000).

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

3. Results and Discussion

Food systems are dynamic, chemically complex and generally heterogeneous matrices of large numbers of biological molecules. The chemical specificity, ease of sampling, rapidity of measurements and nondestructive nature of FT-Raman spectroscopy make it an attractive tool for food analysis. The 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 the incident light (Fig.1). For this reason, the very first step of compositional analysis of wine using FT-Raman is the attribution of characteristic frequency shifts observed in spectra to vibrational modes of molecular bonds (Table 1S in supplementary information).

As Fig.1 shows, the large band ascribed to OH stretching at 3350 cm^{-1} is clearly visible in all the analyzed spectra. Besides, a minor band related to OH bending at 1700-1500 cm^{-1} can be noticed. The group of peaks between 3000-2800 cm^{-1} is due to the symmetric and asymmetric stretching of CH_x bonds. Several other characteristic peaks of ethanol are present at a lower than 1500 cm^{-1} frequency. They are associated to several deformation modes of CH_x as reported in Tab 1S available in supplementary information. (Mammone, Sharma, Nicol, 1980). All peaks of wine are slightly shifted 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 is present in wine spectra. This band is characteristic of C=O stretching, a not very active Raman vibration. The C=O peak could be attributed to several species present in the matrix (e.g. organic acids and flavonoids) whose carbonyl groups are characterized by slightly different vibration frequencies. Therefore, a quite broad signal is registered in this region of the spectra.

The analyzed samples were chosen with the aim of representing a wide selection of the three selected wines and they were purchased by tens of different producers. Many samples are requested in order to represent the total variability of the system and to obtain a representative data set for the multivariate calibration. The Raman spectra of different wines are very similar to each other as it can be noticed in Fig. 1 where the spectra of Dolcetto, Barbera and Nebbiolo are compared. This explains why a univariate analysis would not be effective. It was decided to employ a multivariate approach to have a more complete interpretation of the characteristic pattern of spectroscopic signals.

Fig. 1

From an oenological point of view the specific features of wines are the result of a synergic effect of 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 that provide the bitter taste; for example the expression *total acidity* refers to the rough, tart and sour attributes of the wine which are evaluated in relation to how well the acidity balances out the sweetness and bitter components of the wine. During the course of winemaking and in the finished wines, acetic, butyric, lactic and succinic acid can play significant roles and all together define the characteristic acidity of the wine (Bellman, Gallander, 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 no one of them can be commented singularly. Literature is poor of interpretative analysis regarding Raman spectra of wine because of the complexity of the issue and only a chemometric analysis permits to extract the more interesting information and selective parameters to distinguish and attest the authenticity of different wine products. The chemometric approach used for the classification is a supervised classification method whose task is grouping 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. In particular, the training data are given in the form of sets of spectra with their desired partitioning as a supervised method would suggest (Finley and Joachims, 2005). Different distance functions can be used in order to evaluate the distance between objects of the same class or the assignation to the correct class for an unknown object. In this case the Mahalanobis distance (Md) is used, as described in detail in Materials and Methods.

Applying this concept to spectral data of wine, several classification models with good classification capability were obtained, as described in details in the following paragraphs (3.1; 3.2; 3.3).

3.1 Discrimination in accordance with grape

At first three classes of grapes (Nebbiolo, Barbera, Dolcetto) were defined. 185 Nebbiolo, 75 Barbera, 45 Dolcetto wine samples were subjected to Raman analysis in order to constitute a substantial training set. The *eigenanalysis* 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 explained variability of 99.34 % using 20 principal components (PCs); the number of principal components was optimized by considering the classification capability % (the number of correctly classified samples during cross-validation) as a function of the PCs number. In particular, leave-one-

out cross validation was reiteratively performed rising the number of considered PCs at each run and the percentage of correctly classified samples is plotted as a function of PCs number in figure 1S (see supplementary information), as well as the explained variance corresponding to each PC. The plot reported in figure 1S in supplementary information, is used to determine the best number of PCs, which corresponds to 20 in this case. In order to avoid the overfitting of data, the components that do not contribute significantly to cumulative explained variance and that not provide useful information for classification were excluded because they deal exclusively with experimental noise.

As Figure 2 shows, the best optimized method misclassifies 13.1 % of 305 standards during leave-one-out cross validation process. The clouds of points representing the three classes are dense, meaning a high homogeneity within each class. The three clouds are also very close to each other and they partially overlap which is the cause of a misclassified percentage greater than 10 %. However, it should be taken into account that the disciplinary of production of some wines allows a small percentage of other wines to be introduced (for example, Barbera wine can contain up to 15 % of Nebbiolo grape in accordance with its disciplinary); this could likely represent one of the reasons of the closeness of sample classes which causes the misclassification. A cross validation test was performed (and repeated 5 times) in order to attest the real capability of the calibrated model to distinguish wines according to the grape. 100 spectra (1/3 of the number of calibration standards per each class randomly chosen) were used by groups of five for the 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 class, 2 % belong to Dolcetto class and 3 % belong to Nebbiolo class on average. It shall also be noticed that the percentage of misclassified samples during leave-five-out cross validation method is comparable with leave-one-out cross validation results (14 % of misclassified with 20 PCs) reached during model optimization. Subsequently 10 new Nebbiolo samples were provided and they were used as a little external test set which provided 90 % of correct answers.

Fig. 2

The loadings profiles corresponding to principal component from 1 to 10, which are the most interesting for a qualitative description, are shown in fig. 3. From a careful analysis of them it is possible to determine which organoleptic and compositional features are responsible for the classification. However, it must be taken into account that a synergic cooperation of variables lead to the class separation and none of them can be considered separately from the others. For example, alcohol content of a wine is a key parameter for its oenological

characterization and it also plays an important role in the spectroscopic analysis in order to depict a faithful portrait of each sample. The ethanol Raman peaks are the most easy to be individuated in a Raman spectrum of wine and they can be identified in most of the calculated PCs as well. It is possible to affirm that this aspect plays a crucial role in wine classification. Another important feature that could help in classification is the sugar content. Since the sugar content of a wine depends on the advancement of the alcoholic fermentation, a founded hypothesis is the anti-correlation between the sugar and the alcohol content depending variables. PC8 and PC9 reveal that a significant variability of data is recognized during the statistical calculation in the spectral region around 3500 cm^{-1} and 500 cm^{-1} where the typical carbohydrates peaks can be found. The scores plot built in accordance with the above mentioned PCs reveals that the carbohydrate content varies from sample to sample without any correlation with the Dolcetto, Barbera or Nebbiolo belonging class. The difficulty of defining a coherent variability in this case lies in the fact that all the considered wines are dry wines.

Another important parameter in the Raman characterization of a biological matrix is the fluorescence effect. The colored substances contained in wine, such as anthocyanins and polyphenols in general, are directly related with the fluorescence effect observed during spectra acquisition. Fluorescence is generally an undesired effect in Raman analysis because of the risk of covering the interesting signals in the spectrum. It can also influence the statistical analysis of wine spectra in the classification process. Indeed, the slope of the baseline of PC1, and the wide band around 2000 and 1200 cm^{-1} of PC6 and PC7 attest that the fluorescence represents a significant variable for the examined system. This behavior is even more evident by looking at the disposition of data clouds as a function of PCs influenced by fluorescence, where it can be clearly seen that the fluorescence effect does not represent a negligible variable. However, the classification of wines is not impaired by the fluorescence contribution, whose success is witnessed not only by the satisfying modeling of training set but also by the above mentioned external validation set.

Fig. 3

The present data reveal that synergic interaction among variables represents the key to solve an apparently very complicated problem. Considering variables singularly it would not be possible to describe the huge amount of data in a significant way, but taking them all together a good separation of the three selected classes is produced.

Also dual class models were optimized and it turned out that the most difficult wines to separate are Barbera and Dolcetto wines whereas Nebbiolo sets oneself up a specific well distinguished class.

3.2 Discrimination in accordance with production area

After that the capability of Raman spectroscopy to separate wines according to the grape was demonstrated, a method to attest the geographical provenance of wine within the same grape class was developed. In order to understand the importance of the geographical area where a wine is produced, it is good to know that it exists in oenology a technical expression to intend the particular combination of elements such as climate, soil and regional knowhow of winemaker, which defines the uniqueness and unrepeatability that characterize a labeled wine, this is the French word *Terroir*.

The study was focused on two wines in particular, Dolcetto and Nebbiolo. Within Dolcetto class (i) Dolcetto d'Alba Doc and (ii) Dolcetto di Dogliani Docg were chosen for experiments. 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 Langhe territory as it can be seen in the map inlet of Fig. 4 a. The Dolcetto area is the highest of the Langhe territory (from 250 to 700 m above sea level) and it characterized by a fresh climate because of the proximity to Appenino Ligure and Alpi Marittime mountains chain. This represents the best climate condition for Dolcetto wine production because it makes the grape maturation process slower. In this geographic area the soil varies from generous red soil to sandy and dry soil (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 disciplinary that declare 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 very similar optic and organoleptic features and even for expert sommelier it could be very difficult to distinguish the geographic origin of the twos at taste. The Raman analysis coupled with chemometric provided a good identification method for the classification of the samples according to the area of production as shown in the Cooman's plot in Fig. 4 a.

For Nebbiolo wine two classes were set as well: (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 represented in the Piedmont map in the inlet of Fig 4 b. Nebbiolo wine is an ancient red mono-vine wine. His tracks in Piedmont predate the seventeenth century and it has always thrived here because of its adaptability to cold climates (langhevini.it). It is a noble Italian vine par excellence, from which derive the majority quality red wines for long ageing in the north-west of Italy. This wine reaches his best

after few years from the production because of the territory in which it is produced. The geographic area designated for the production of Nebbiolo is well specified in its own disciplinary. 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 (regione.piemonte.it). The chemometric analysis of Nebbiolo spectra provided satisfying results for the classification of Nebbiolo from Langhe and from Novara&Carema territory as it can be noticed observing the Cooman's plot in Fig. 4b. As commented before, the whole structure of the spectra of different wines is responsible of the class separation. The number of considered PCs (6 for Dolcetto classification and 14 for Nebbiolo classification) represents the best compromise between explained variance and classification capability, as commented in paragraph 3.1 (Fig 1S b, c available in supplementary information). Also in this case the only way to achieve significant results consists in a multivariate approach. The appreciable classification capability higher than 90 % are obtained for the two classification models and the low number of misclassified standards permits to conclude that Raman spectroscopy is able to discriminate wine provenance when a consistent calibration is previously performed.

Fig. 4

The cross validation test provided satisfying results for both calibrated models. 10 samples were randomly chosen (about 30 % of the number of calibration samples from each class) and they were used by couples for validating the Dolcetto model with an error of 8 %, all of the misclassified samples belong to "Dolcetto d'Alba" class. The leave-five-out cross-validation for Nebbiolo was performed using 65 spectra, five by five randomly chosen with respect to the total number of calibration samples in each class. In this case, 7 % of them were misclassified. In particular, 1 of them is from Alba, while 5 are from the northern part of Piedmont (Novara&Carema class). Validation procedure was repeated 5 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, it was investigated the possibility to recognize aged from non-aged oenological products. Many wines improve in quality during barrel and bottle storage. Such wines eventually reach their best features, and with further ageing begin to decline. During the ageing period, the acidity decreases, additional 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 barrel made of oak, allowing oxygen to enter

and water and alcohol not to escape. Wine simple phenols are further transformed during wine ageing into complex molecules formed by the condensation of proanthocyanidins and anthocyanins, which 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 in "brick red" hues.

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

Fig. 5

Also in this case a cross validation of the calibrated model was performed. 30 spectra of unknown samples were subjected to analysis by groups of five. The validation procedure was repeated 5 times and it provided 84 ± 4 % of correct answer on average. Among the 16 % wrongly classified, 80 % was Barolo and 20 % was Barbaresco.

4. Conclusions

In this paper it was demonstrated that Raman spectroscopy coupled with chemometric analysis can play a relevant role in the authenticity 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 advantage 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 rapidity of data collection, make Raman Spectroscopy particularly interesting for the prevention of wines fraud and for the control procedures necessary to the assignment of quality labels. The common drawback of Raman spectroscopy analysis of food matrices, such as the difficult spectra interpretation are overcome thanks to user-friendly software which allow sophisticated chemometric methods to be elaborated by treating a large amount of data. The chemometric identification of variability between the different classes hits the target: wine differentiation in accordance with grape, geographical origin, and ageing time was successfully performed using a Raman spectrometer. Even if a dedicated test set constituted by external samples should be subjected to analysis in order to attest classification capability of the proposed method in a real case,

362 this proof of principle aims at demonstrating that a multivariate calibration procedure can provide consistent
363 classification results when a substantial calibration set is subjected to spectroscopic analysis, even if the matrix is
364 complex as wine samples are. The more specific and sensible Raman analysis of wine is, the more Raman would
365 be exploitable for the single wine producer certification. The application of Raman spectroscopy to distinguish a
366 single wine producer will be the next challenge, with a higher impact in commercial field.

367 **Acknowledgements**

368 This research was funded by the European Union through the European Metrology Research Programme. The
369 European Metrology Research Program (EMRP) is jointly funded by the EMRP participating countries within
370 EURAMET and the European Union.

371 **References**

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

375 Adam L., Bartels W., Christoph N., Stempfl W. (1995) Qualita¨tskontrolle und Analytik im Fachlabor.
376 *Brennereianalytik Band 2,* 1, 27.

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

379 Bellman R. B., Gallander J. F. (1979). Wine Deacidification. In Chichester C. O., Mrak E. M., Stewart G. F.,
380 *Advances in Food Research,* 25.

381 Bernuy B., Meurens M., Mignolet E. and Larondelle Y. (2008). Performance Comparison of UV and FT-Raman
382 Spectroscopy in the Determination of Conjugated Linoleic Acids in Cow Milk Fat. *J. Agric. Food Chem.*, 56 (4),
383 1159–1163.

384 Bréas O., Reniero F., Serrini G., Martin G. J. and Rossmann A. (1994). Isotope ratio mass spectrometry:
385 Analysis of wines from different European Countries. *Rapid Communications in Mass Spectrometry*, 8, 12, 967–
386 970.

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

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

392 Cozzolino, D., Dambergs, R., Janik, L., Cynkar, W., & Gishen, M. (2006). Review: Analysis of grapes and wine
393 by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 14(1), 279.

394 Cozzolino, D., McCarthy, J., & Bartowsky, E. (2012). Comparison of near infrared and mid infrared
395 spectroscopy to discriminate between wines produced by different *Oenococcus oeni* strains after malolactic
396 fermentation: A feasibility study. *Food Control*, 26(1), 81–87.

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

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

402 Dordevic N. , Wehrens R. , Postma G.J. , Buydens L.M. , Camin F. (2012). Statistical methods for improving
403 verification of claims of origin for Italian wines based on stable isotope ratios. *Anal Chim Acta*. Dec 13;757:19-
404 25. DOI: 10.1016/j.aca.2012.10.046.

405 Durante C., Baschieri C., Bertacchini L. (2015). An analytical approach to Sr isotope ratio determination in
406 Lambrusco wines for geographical traceability purposes. *Food chem*istry, 173, 557-63.

407 Finley T. and Joachims T (2005). Supervised clustering with support vector machines. In Thorsten J. *ICML*.

408 Fronza G, Fuganti C., Graselli P., Reniero F., Guillou G., Breas O., Sada E., Rossmann A., Hermann A. J.
409 (1998), Determination of the ^{13}C Content of Glycerol Samples of Different Origin**.** *Agric. Food Chem*, 46, 477–
410 480.

411 Gallego A. L., Guesalaga A. R., Bordeu E. and Gonzàlez A. S. (2011). Rapid measurement of phenolic
412 compounds in red wine using Raman spectroscopy. *Instrumentation and Measurement*, 60, 2, 507 – 512.

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

415 Rolf Godelmann, Fang Fang, Eberhard Humpfer, Birk Schütz, Melanie Bansbach, Hartmut Schäfer, and
416 Manfred Spraul, (2014). Targeted and Nontargeted Wine Analysis by 1H NMR Spectroscopy Combined with
417 Multivariate Statistical Analysis. Differentiation of Important Parameters: Grape Variety, Geographical Origin,
418 Year of Vintage. *J. Agric. Food Chem.*, 61 (23), pp 5610–5619. DOI: 10.1021/jf400800d

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

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

424 Li-Chan E.C.Y., Griffiths P. R. and Chalmers J. M. (2010). Applications of Vibrational Spectroscopy in Food
425 Science. WILEY

426 Li-Chan E., Nakai S., Hirotsuka M. (1994). Raman Spectroscopy as a Probe of Protein Structure in Food
427 Systems in *Protein Structure-Function Relationships in Foods*. 163-197.

428 Monakhova Y.B. , Godelmann R. , Hermann A. , Kuballa T. , Cannet C. , Schäfer H. , Spraul M. , Rutledge
429 D.N.. (2014). Synergistic effect of the simultaneous chemometric analysis of ^{1}H NMR spectroscopic and stable
430 isotope (SNIF-NMR, ^{18}O, ^{13}C) data: application to wine analysis. *Anal Chim Acta*. 2014 Jun 23;833:29-39. doi:
431 10.1016/j.aca.2014.05.005.

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

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

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

438 Mazzei P., Francesca N., Moschetti G., Piccolo A. (2010). NMR spectroscopy evaluation of direct relationship
439 between soils and molecular composition of red wines from Aglianico grapes. *Analytica Chimica Acta*, 673,
440 167–172.

441 Meneghini, C., Caron, S., Proulx, A., Emond, F., Paradis, P., Pare, C., et al. (2008). Determination of ethanol
442 concentration by Raman spectroscopy in liquid-core microstructured optical fiber. *IEEE Sensors Journal*, 8(7),
443 1250–1255.

444 Misselhorn K., Grafahrend W. (1990). Rohstoffnachweis bei hochgereinigtem Alkohol. *Branntweinwirtschaft*,
445 130, 70–73.

446 Nordon A., Mills A., Burn R. T., Cusick F. M. and Littlejohn D. (2005). Comparison of non-invasive NIR and
447 Raman spectrometries for determination of alcohol content of spirits. *Analytica Chimica Acta*, 548, 1-2, 148 –
448 158.

449 Özbalcia B., Hakkı Boyacia İ., Topcua A., Kadılarb C., Tamerc U. (2013). Rapid analysis of sugars in honey by
450 processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chemistry*, 136,
451 (3–4), 1444–1452.

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

454 Pardo M. A. (2014). Evaluation of a dual-probe real time PCR system for detection of mandarin in commercial
455 orange juice. *Food chemistry* 172C, 377-84.

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

458 Petronis A. (2010). Epigenetics as a unifying principle in the etiology of complex traits and diseases. *Nature*,
459 465, 7299, 721-727.

460 Roig, B., & Thomas, O. (2003). UV monitoring of sugars during wine making. *Carbohydrate Research*, 338(1),
461 79–83.

462 Rossmann A. (2001). Determination of Stable Isotope Ratios in Food Analysis. *Food Reviews International*, 17,
463 3, 347-381.

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

466 Schulz H. and Baranska M. (2007). Identification and quantification of valuable plant substances by IR and
467 Raman spectroscopy. *Vibrational Spectroscopy*, 43, 13 – 25.

468 Socrates G., *Infrared and Raman Characteristic Group Frequencies: Tables and Charts*, 3rd Edition, Wiley
469 (2004) ISBN: 978-0-470-09307-8

470 Thygesen L. G., Løkke M. M E., Micklander and Engelsen S. B. (2003). Vibrational microspectroscopy of food.
471 Raman vs. FT-IR. *Trends in Food Science & Technology*, 14, 50.

472 Urbano, M., Luque de Castro, M. D., Pérez, P. M., García-Olmo, J., & Gómez-Nieto, M. A. (2006). Ultraviolet
473 visible spectroscopy and pattern recognition methods for differentiation and classification of wines. *Food*
474 *Chemistry*, 97(1), 166–175.

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

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

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

481 Regione.piemonte.it
482 http://www.regione.piemonte.it/agri/politiche_agricole/viticoltura/dwd/disciplinari/dolcettoalba.pdf (last access
483 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.