

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Controlling Protected Designation of Origin of wine by Raman Spectroscopy

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1636999> since 2017-05-22T08:38:28Z

Published version:

DOI:<https://doi.org/10.1016/j.foodchem.2016.05.011>

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1 Controlling Wines' Protected Designation of Origin 2 through Raman Spectroscopy

3 Luisa Mandrile^{a,c}, Giuseppe Zeppa^b, Andrea Mario Giovannozzi^c and Andrea Mario Rossi^c

4 ^a *Department of Drug Science and Technology, Università degli Studi di Torino, Via Giuria 9, 10125, Torino, Italy*

5 *l.mandrile@inrim.it*

6 ^b *Dipartimento di Scienze Agrarie, Forestali e Alimentari (DISAFA) - Microbiologia agraria e Tecnologie alimentari, Largo*

7 *Paolo Braccini 2, 10095 Grugliasco (TO), Italy giuseppe.zeppa@unito.it*

8 ^c *Quality of life Division, Food Metrology program, Istituto Nazionale di Ricerca Metrologica, Strada delle Cacce, 91 10135,*

9 *Torino, Italy a.rossi@inrim.it, a.giovannozzi@inrim.it*

10
11 ^{*} *Corresponding author: Andrea Mario Rossi, tel +39 011 3919342; fax +39 011 346384 e-mail a.rossi@inrim.it*

12
13 **Keywords:** wine, raman spectroscopy, food traceability, chemometrics, fingerprint

14 **Abstract**

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

23 classification. The present methodology can be successfully applied on raw materials without any preliminary
24 treatment of the sample, providing a response in a very short time.

25

26 **Main text**

27 **1. Introduction**

28 In order to preserve the quality of food products coming from particular geographical areas and to protect
29 consumers against imitations and false information, the European Commission defined, via Regulations
30 1151/2012, the labels Traditional Specialty Guaranteed (TSG), Protected Designation of Origin (PDO) and
31 Protected Geographical Indication (PGI) (Regulation (Eu) No 1151/2012 Of The European Parliament And Of
32 The Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs). Quality labels
33 play an important role in consumer behavior and give confidence about the origins and the quality of food.
34 Labels assignment is an important market claim and it represents a valuable weapon to attest and justify the
35 economic value of alimentary products. Traceability has become a very relevant concept when it is associated to
36 edible products and it represents an essential tool to enhance traders and consumers' confidence in the safety,
37 quality, and authenticity of food.

38 Unfortunately, food traceability procedures mainly involve tedious administrative documents, while scientific
39 methodologies that objectively identify the authenticity of food would be preferable. According to this, scientific
40 research is now focused on developing analytical methods to authenticate the geographical origin of aliments in
41 a metrological way (Peres, Barlett, Loiseau and Montet, 2007) with the aim of linking food products with its
42 distinctive features, such as ingredients, physical properties and production methods. Food traceability analysis
43 are usually performed by means of several analytical techniques such as mass spectrometry for isotope ratio
44 determination (Durante, Baschieri, Bertacchini, 2015), DNA based techniques, such as Polymerase Chain
45 Reaction (PCR) (Pardo, 2014) and Nuclear Magnetic Resonance spectrometry (NMR) (Mazzei, Francesca,
46 Moschetti, Piccolo, 2010).

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

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

58 DNA based technologies were also exploited in this field due to their specificity of analysis because they are
59 strictly connected to the genotype (the inherited instructions that an organism carries within its genetic code) but
60 they inevitably miss the stochastic significant epigenetic differences accumulating over time across cells
61 (Petronis, 2010). Dordevic N et. al (2013) highlighted the need of new methods for a better geographical
62 discrimination between samples, demonstrating that multivariate methods are superior to univariate approaches.

63 Interesting alternatives or even complementary methods are represented by the NMR and vibrational
64 spectroscopy techniques. Godelmann et al. (2013) analyzed about 600 German wines and demonstrated that
65 ^1H NMR coupled with statistical data treatment can provide individual “fingerprint” of a wine sample, which
66 includes information about variety, origin, vintage, physiological state, technological treatment, and other. The
67 data fusion of NMR profiling and stable isotope data for wine analysis was also investigated and reported in
68 literature with good results (Monakhova et. al. 2014). However, the main drawbacks of all the cited techniques,
69 i.e. MS, NMR and DNA based techniques, are related to the elevated cost of the instruments, extensive sample
70 pre-treatments and long time of analysis which often reduce the accuracy and precision of measurements. Since
71 simple and rapid analytical methods are needed to meet the developing European labeling legislation, vibrational
72 spectroscopy is emerging as a new powerful tool in authenticating food provenance. Vibrational spectroscopy
73 techniques usually provide non-destructive analysis of the samples, fast collection times with none or minimal
74 sample pre-treatments that reduce the total time of analysis and could support the development of reliable control
75 procedures and screening methods for food traceability. Moreover, new modern and portable instruments with
76 smart accessories were developed in the last years making these techniques more suitable for *in line* process
77 monitoring and *in situ* analysis (Gallego, Guesalaga, Bordeu and González, 2011). These methods encompass
78 absorption spectroscopy in the mid-infrared (MIR) and the near-infrared (NIR) for studying fundamental
79 molecular vibrations and their harmonics (Bauer et al., 2008; Cozzolino, Damberg, 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 Boyacıa,
Topcu, Kadırlar, Tamerc, 2013; Paradkar and Irudayaraj, 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

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

110

111 **2. Material and Methods**

112 *2.1 Samples*

113 Study was performed on 315 samples of commercial wines obtained from Nebbiolo, Barbera and Dolcetto grape
114 by different winemakers. For each grape variety, wines of different area and ageing time were used (Table 1).
115 Among the total sample pool, more than 10 Protected Designation of Origin (PDO) wines were examined. The
116 number of samples for each PDO wine was different according to winemakers and diffusion and inevitably
117 limited by the availability of samples. All samples were furnished directly by producers and stored at +4°C until
118 analysis.

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

120

121 *2.2 Raman measurements*

122 Raman spectroscopy was performed with a Thermo Scientific NXR FT-Raman Module Nicolet Series™
123 equipped with an InGaAs detector (ThermoFisher Scientific, Waltham, USA), a CaF₂ beamsplitter and a 1064
124 nm laser line. Raman spectra were collected using a laser power of 0.9 W in a spectral range from 200 to 4000
125 cm⁻¹ with a resolution of 4 cm⁻¹. 256 scans were collected to obtain S/N ratio higher than 15. Samples were
126 analyzed in 4 ml glass vials positioned vertically on a powered stage.

127 *2.3 Multivariate Analysis.*

128 The raw Raman spectra were subjected to Discriminant Analysis using TQ Analyst™ 8.0 software
129 (ThermoFisher Scientific, Waltham, USA). Spectra were preprocessed using Savitzky-Golay smoothing filter
130 (Savitzky, Golay, 1964) in order to remove of as much noise as possible without unduly degrading the spectral
131 information. The spectral range to be analyzed was selected in such a way that the interference of the random
132 variability of spectra is minimized and it does not provide spurious information to the classification model.
133 Seven restricted spectral regions around Raman peaks are selected in order to optimize the classification result.

134 The frequency regions of spectra which do not contain any Raman peak (e.g. 800-600 cm⁻¹ and 2800-1800 cm-
135 1) were excluded. In this way the worthless information is ignored and the best classes' separation is obtained.
136 The suitable number of PCs to be considered is chosen as the best compromise between the explained variance
137 of each PC and the predictive capability of the model: when the cumulative variance reaches the *plateau*, further
138 components could not provide useful information and they should be excluded, so that variables that just
139 represent noise are not considered for classification. Otherwise, some variables that explain a little variability
140 shall not be excluded if they significantly improve the classification capability of the model (% of samples
141 correctly classified). The presented chemometric models for wine classification were first validated through a
142 leave-one-out cross-validation procedure during model optimization (mathematical pretreatment choice,
143 significative PCs selection etc.). Finally, the optimized models were all validated through a cross validation
144 procedure using exclusions sets made of five samples randomly chosen, the number of exclusion set is
145 proportional to the total number of calibration samples. After calibration, this classification technique permits to
146 calculate the unknown's distance to a class center in terms of Mahalanobis distance (Mahalanobis, 1936) and to
147 assign each unknown samples to the correct class. The basic idea on which Md is based is the fact that it contains
148 an autoscaling process in itself and it overcomes the assumption of spherical distribution of sample points about
149 the center of mass; therefore non-spherical distributions can be described as well. In the generalized formula for
150 Md reported below, the observation are represented by $x=(x_1, x_2, \dots, x_n)$ while $\mu=(\mu_1, \mu_2, \dots, \mu_n)$ represents
151 the observations' mean. The apex ^T indicates the transposed matrix ($x-\mu$). S^{-1} is the inverse of the covariance
152 matrix of the observations.

$$153 \quad Md(x)=\sqrt{(x-\mu)^T S^{-1} (x-\mu)}$$

154 If an ellipsoidal distribution is considered, then we would expect that the probability of the test point to belong to
155 the set does not only depend on the distance from the center of mass, but also on the direction. (De
156 Maesschalck, Jouan-Rimbaud, Massart, 2000).

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

162

163 3. Results and Discussion

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

171 As Fig.1 shows, the large band ascribed to OH stretching at 3350 cm^{-1} is clearly visible in all the analyzed
172 spectra. Besides, a minor band related to OH bending at $1700\text{-}1500\text{ cm}^{-1}$ can be noticed. The group of peaks
173 between $3000\text{-}2800\text{ cm}^{-1}$ is due to the symmetric and asymmetric stretching of CH_x bonds. Several other
174 characteristic peaks of ethanol are present at a lower than 1500 cm^{-1} frequency. They are associated to several
175 deformation modes of CH_x as reported in Tab 1S available in supplementary information. (Mammone, Sharma,
176 Nicol, 1980). All peaks of wine are slightly shifted in comparison with the pure ethanol peaks; this is due to the
177 simultaneous presence of different organic species such as glycerol, acetaldehyde, organic acids, and
178 polyphenols including flavonoids and non-flavonoids. At 1630 cm^{-1} a low intensity band is present in wine
179 spectra. This band is characteristic of C=O stretching, a not very active Raman vibration. The C=O peak could
180 be attributed to several species present in the matrix (e.g. organic acids and flavonoids) whose carbonyl groups
181 are characterized by slightly different vibration frequencies. Therefore, a quite broad signal is registered in this
182 region of the spectra.

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

190 **Fig. 1**

191 From an oenological point of view the specific features of wines are the result of a synergic effect of several
192 factors. The wine composition is very complex and the final organoleptic features are produced by the
193 interaction of many chemicals, such as sugar, alcohol, acids and tannins that provide the bitter taste; for example
194 the expression *total acidity* refers to the rough, tart and sour attributes of the wine which are evaluated in relation
195 to how well the acidity balances out the sweetness and bitter components of the wine. During the course of
196 winemaking and in the finished wines, acetic, butyric, lactic and succinic acid can play significant roles and all
197 together define the characteristic acidity of the wine (Bellman, Gallander, 1979). In the same way, from a
198 spectroscopic point of view, the final wine spectrum is the result of a synergic interaction of many factors and no
199 one of them can be commented singularly. Literature is poor of interpretative analysis regarding Raman spectra
200 of wine because of the complexity of the issue and only a chemometric analysis permits to extract the more
201 interesting information and selective parameters to distinguish and attest the authenticity of different wine
202 products. The chemometric approach used for the classification is a supervised classification method whose task
203 is grouping a set of objects in such a way that objects in the same group (called a class) are more similar to each
204 other than those belonging to other classes. In particular, the training data are given in the form of sets of spectra
205 with their desired partitioning as a supervised method would suggest (Finley and Joachims, 2005). Different
206 distance functions can be used in order to evaluate the distance between objects of the same class or the
207 assignment to the correct class for an unknown object. In this case the Mahalanobis distance (Md) is used, as
208 described in detail in Materials and Methods.

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

211 *3.1 Discrimination in accordance with grape*

212 At first three classes of grapes (Nebbiolo, Barbera, Dolcetto) were defined. 185 Nebbiolo, 75 Barbera, 45
213 Dolcetto wine samples were subjected to Raman analysis in order to constitute a substantial training set. The
214 *eigenanalysis* attested that the selected 305 calibration standards contain sufficient variability for the method
215 calibration. The spectral range was optimized as reported in Materials and Methods section. The optimized
216 chemometric model shows a total explained variability of 99.34 % using 20 principal components (PCs); the
217 number of principal components was optimized by considering the classification capability % (the number of
218 correctly classified samples during cross-validation) as a function of the PCs number. In particular, leave-one-

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

226 As Figure 2 shows, the best optimized method misclassifies 13.1 % of 305 standards during leave-one-out cross
227 validation process. The clouds of points representing the three classes are dense, meaning a high homogeneity
228 within each class. The three clouds are also very close to each other and they partially overlap which is the cause
229 of a misclassified percentage greater than 10 %. However, it should be taken into account that the disciplinary of
230 production of some wines allows a small percentage of other wines to be introduced (for example, Barbera wine
231 can contain up to 15 % of Nebbiolo grape in accordance with its disciplinary); this could likely represent one of
232 the reasons of the closeness of sample classes which causes the misclassification. A cross validation test was
233 performed (and repeated 5 times) in order to attest the real capability of the calibrated model to distinguish wines
234 according to the grape. 100 spectra (1/3 of the number of calibration standards per each class randomly chosen)
235 were used by groups of five for the cross validation of the model. During this leave-five-out validation 86 ± 2 %
236 of unknown samples provided a correct answer. Among the misclassified samples, 9 % belong to Barbera class,
237 2 % belong to Dolcetto class and 3 % belong to Nebbiolo class on average. It shall also be noticed that the
238 percentage of misclassified samples during leave-five-out cross validation method is comparable with leave-one-
239 out cross validation results (14 % of misclassified with 20 PCs) reached during model optimization.
240 Subsequently 10 new Nebbiolo samples were provided and they were used as a little external test set which
241 provided 90 % of correct answers.

242 **Fig. 2**

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

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

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

270 **Fig. 3**

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

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

276 3.2 Discrimination in accordance with production area

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

283 The study was focused on two wines in particular, Dolcetto and Nebbiolo. Within Dolcetto class (i) Dolcetto
284 d'Alba Doc and (ii) Dolcetto di Dogliani Docg were chosen for experiments. The production area of Dolcetto di
285 Dogliani is situated in the southernmost part of Piedmont whereas the Dolcetto d'Alba region is situated in a
286 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
287 the Langhe territory (from 250 to 700 m above sea level) and it characterized by a fresh climate because of the
288 proximity to Appennino Ligure and Alpi Marittime mountains chain. This represents the best climate condition
289 for Dolcetto wine production because it makes the grape maturation process slower. In this geographic area the
290 soil varies from generous red soil to sandy and dry soil (regione.piemonte.it); the best soil type for the Dolcetto
291 production is white, deep, clayey and calcareous. Dolcetto di Dogliani and Dolcetto d'Alba wines are produced
292 according to a strict disciplinary that declare in a very precise way the mandatory geographical area and the
293 variety of grape permitted. Also, the winemaking procedure and the final organoleptic features are usually
294 controlled through a qualified panel test. Dolcetto d'Alba and Dolcetto di Dogliani wines have very similar optic
295 and organoleptic features and even for expert sommelier it could be very difficult to distinguish the geographic
296 origin of the twos at taste. The Raman analysis coupled with chemometric provided a good identification method
297 for the classification of the samples according to the area of production as shown in the Cooman's plot in Fig. 4
298 a.

299 For Nebbiolo wine two classes were set as well: (i) Langhe (including Nebbiolo d'Alba, Barolo, Barbaresco); (ii)
300 Novara&Carema (including Colline Novaresi, Coste della Sesia, Ghemme, Gattinara and Carema). The
301 geographical areas involved are represented in the Piedmont map in the inlet of Fig 4 b. Nebbiolo wine is an
302 ancient red mono-vine wine. His tracks in Piedmont predate the seventeenth century and it has always thrived
303 here because of its adaptability to cold climates (langhevini.it). It is a noble Italian vine par excellence, from
304 which derive the majority quality red wines for long ageing in the north-west of Italy. This wine reaches his best

305 after few years from the production because of the territory in which it is produced. The geographic area
306 designated for the production of Nebbiolo is well specified in its own disciplinary. The soil should be clayey,
307 calcareous and acidic or a combination of the three; the territory must be hilly (at least 650 m above sea level)
308 and sunny (regione.piemonte.it). The chemometric analysis of Nebbiolo spectra provided satisfying results for
309 the classification of Nebbiolo from Langhe and from Novara&Carema territory as it can be noticed observing the
310 Cooman's plot in Fig. 4b. As commented before, the whole structure of the spectra of different wines is
311 responsible of the class separation. The number of considered PCs (6 for Dolcetto classification and 14 for
312 Nebbiolo classification) represents the best compromise between explained variance and classification
313 capability, as commented in paragraph 3.1 (Fig 1S b, c available in supplementary information). Also in this case
314 the only way to achieve significant results consists in a multivariate approach. The appreciable classification
315 capability higher than 90 % are obtained for the two classification models and the low number of misclassified
316 standards permits to conclude that Raman spectroscopy is able to discriminate wine provenance when a
317 consistent calibration is previously performed.

318 **Fig. 4**

319 The cross validation test provided satisfying results for both calibrated models. 10 samples were randomly
320 chosen (about 30 % of the number of calibration samples from each class) and they were used by couples for
321 validating the Dolcetto model with an error of 8 %, all of the misclassified samples belong to "Dolcetto d'Alba"
322 class. The leave-five-out cross-validation for Nebbiolo was performed using 65 spectra, five by five randomly
323 chosen with respect to the total number of calibration samples in each class. In this case, 7 % of them were
324 misclassified. In particular, 1 of them is from Alba, while 5 are from the northern part of Piedmont
325 (Novara&Carema class). Validation procedure was repeated 5 times for both DA methods attesting a standard
326 deviation of classification capability of 1 % and 2 % respectively.

327 *3.3 Discrimination in accordance with age*

328 As a third step, it was investigated the possibility to recognize aged from non-aged oenological products. Many
329 wines improve in quality during barrel and bottle storage. Such wines eventually reach their best features, and
330 with further ageing begin to decline. During the ageing period, the acidity decreases, additional clarification and
331 stabilization occur as well as the precipitation of undesirable substances, and complex compounds affecting
332 flavor and aroma are formed. Wines are usually aged in wooden barrel made of oak, allowing oxygen to enter

333 and water and alcohol not to escape. Wine simple phenols are further transformed during wine ageing into
334 complex molecules formed by the condensation of proanthocyanidins and anthocyanins, which explains the
335 change of color of aged wines. As the wine ages, anthocyanins react with other acids and compounds such as
336 tannins, pyruvic acid and acetaldehyde which change the color of the wine in "brick red" hues.

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

344 **Fig. 5**

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

348 **4. Conclusions**

349 In this paper it was demonstrated that Raman spectroscopy coupled with chemometric analysis can play a
350 relevant role in the authenticity of wine, providing positive results in the recognition of mono-vine wines in
351 terms of grape (validation test provided reliability of 93%), geographical provenance (reliability higher than
352 90%) and ageing time (reliability higher than 80%). One of the biggest advantage of the proposed method is the
353 direct analysis of wine through the glass container without any pretreatment and purification process. These
354 advantages, together with the rapidity of data collection, make Raman Spectroscopy particularly interesting for
355 the prevention of wines fraud and for the control procedures necessary to the assignment of quality labels. The
356 common drawback of Raman spectroscopy analysis of food matrices, such as the difficult spectra interpretation
357 are overcome thanks to user-friendly software which allow sophisticated chemometric methods to be elaborated
358 by treating a large amount of data. The chemometric identification of variability between the different classes
359 hits the target: wine differentiation in accordance with grape, geographical origin, and ageing time was
360 successfully performed using a Raman spectrometer. Even if a dedicated test set constituted by external samples
361 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) Qualitätskontrolle 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., Damberg, 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 chemistry*, 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 ¹³C Content of Glycerol Samples of Different Origin. *Agric. Food Chem*, 46, 477–
410 480.

411 Gallego A. L., Guesalaga A. R., Bordeu E. and González 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 ¹H NMR spectroscopic and stable
430 isotope (SNIF-NMR, ¹⁸O, ¹³C) 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 Boyacı İ., Topcu A., Kadırlar 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 Irudayaraj 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/viticultura/dwd/disciplinari/dolcettoalba.pdf (last access
483 18/12/2015)

484 http://www.regione.piemonte.it/agri/politiche_agricole/viticultura/dwd/vitigni/varietà_cloni/nebbiolo_descr.pdf
485 (last access 18/12/2015)

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

487 FIGURE CAPTIONS

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

490 **Figure 2**– Cooman’s plot for Nebbiolo, Barbera, Dolcetto classification model calculated using Discriminant
491 Analysis.

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

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

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

498

499

500