1   **Geographical identification of Chianti red wine based on ICP-MS element composition**

2   Benedetta BRONZI[a], Claudio BRILLI[a], Gian Maria. BEONE[b], Maria Chiara FONTANELLA[b],

3   Davide BALLABIO[c], Roberto TODESCHINI[c], Viviana CONSONNI[c], Francesca GRISONI[c],

4   Francesco Parri[a], Massimo BUSCEMA[d,e]

5

6   [a]RUFFINO Srl, 50065 Pontassieve, Firenze, Italy

7   [b]Department for Sustainable Process, Agricultural Faculty, Università Cattolica del Sacro Cuore of

8   Piacenza, 29212 Piacenza, Italy

9   [C]Department of Earth and Environmental Sciences, University of Milano Bicocca, 20126 Milano,

10   Italy

11   [d]SEMEION, 00128 Roma, Italy

12   [e]Dept. Mathematical and Statistical Sciences, University of Colorado, 80217-3364 Denver, CO,

13   USA

14

15

16   **ABSTRACT**

17   Chianti is a precious red wine and enjoys a high reputation for its high quality in the world wine

18   market. Despite this, the production region is small and product needs efficient tools to protect its

19   brands and prevent adulterations. In this sense, ICP-MS combined with chemometrics has

20   demonstrated its usefulness in food authentication. In this study, Chianti/Chianti Classico, authentic

21   wines from vineyard of Toscana region (Italy), together samples from 18 different geographical

22   regions, were analyzed for major and trace elements with the objective of differentiate them from

23   other Italian wines. Partial Least Squares-Discriminant Analysis (PLS-DA) identified variables to

24   discriminate wine geographical origin. REE, major and trace elements all contributed to the

25   discrimination of Chianti samples. General model was not suited to distinguish PDO red wines from

26   samples, with similar chemical fingerprints, collected in some regions. Specific classification

27 models enhanced the capability of discrimination, emphasizing the discriminant role of some

28 elements.

29

30 **1. INTRODUCTION**

31 The quality and the reliability of the origin of food products has become a field of increasing

32 importance in Europe. Since 1992 the European Union created systems known as PDO (Protected

33 Designation of Origin) but with Council Regulation (EC) No 479/2008 brought the prerogative for

34 the approval of the PDO at Community level, while before the Ministerial Decree was the means to

35 proceed in the individual countries of the European Union. PDO covers agricultural products that

36 are produced, processed and prepared in a given geographical area using recognized know-how.

37 Products, registered under this scheme, could be protected against frauds, including adulteration,

38 false declaration of age, and geographical origin (Medina, 1996, Ortega, Gonzalez-San Jose, &

39 Beltrán, 1999). Typical foodstuffs with recognized PDO are cheeses, meats, oils, honey, beers and

40 wines.

41 The EU is a leading producer of wine, as it accounts for 45% of wine-growing areas, 65% of

42 production, 57% of global consumption and 70% of exports in global terms (Regulation, 1992). In

43 2016, the total Italian wine production was 16% higher than the historical average, while PDO wine

44 production increased by 21% compared to the historical average. The total volume of wine, linked

45 to its geographical origin, was represented by 56% of white wine and 44% of red wine (ISTAT,

46 2016/17).

47 Wine is often an adulterated product for its complex production chain and for its commercial value

48 (Di Paola-Naranjo et al., 2011). The identification of the geographical origin of wines is of great

49 interest for wine consumers and producers and it is of fundamental relevance providing criteria for

50 tracing the production chain by chemical markers (Aceto et al., 2013).

51 Chianti area is a valley in central-western Italy between Florence and Siena. It is mostly known for

52 the production of high-quality red wines, which are distinguished with a PDO label at European

53 level (Regulation (EU) No 1308/2013) and CGDO at national level (DPR 02.07.1984). Nowadays,

54 the production of Chianti and Chianti Classico wines is estimated in 835,000 hl/year and 286,400

55 hl/year in 2017 respectively, with a real value of these productions of €87 million for Chianti and

56 €67 million for Chianti Classico in 2016/2017 (ISTAT, 2017).

57 It has been observed that the elemental composition of wines depends on several factors, including

58 soil characteristics, type of grape, area of production and environmental conditions, allowing the

59 definition of a representative ''fingerprint" which is especially important for quality wines

60 produced in specific regions (Gonzálvez, Llorens, Cervera, Armenta, & De la Guardia, 2009).

61 The application of relationship between elements present in the wine and soil composition based on

62 multi-element data, with varying degrees of success has been previously studied in different wine-

63 producing regions or countries like Argentina (Di Paola-Naranjo et al., 2011), South Africa (P P

64 Coetzee, Van Jaarsveld, & Vanhaecke, 2014; Paul P Coetzee et al., 2005; Van der Linde, Fischer, &

65 Coetzee, 2010), Canada (Greenough, Mallory-Greenough, & Fryer, 2005; Taylor, Longerich, &

66 Greenough, 2003), New Zealand (Angus, O'keeffe, Stuart, & Miskelly, 2006), Slovenia (Šelih,

67 Šala, & Drgan, 2014), Romania (Geana et al., 2013), Czech Republic (Kment et al., 2005),

68 Germany (Thiel, Geisler, Blechschmidt, & Danzer, 2004), Switzerland (Gremaud, Quaile, Piantini,

69 Pfammatter, & Corvi, 2004), Portugal (Almeida & Vasconcelos, 2003), Spain (Barbaste, Medina,

70 Sarabia, Ortiz, & Pérez-Trujillo, 2002; Gonzálvez et al., 2009; Iglesias, Besalú, & Anticó, 2007)

71 and Italy (Aceto et al., 2013; Galgano, Favati, Caruso, Scarpa, & Palma, 2008; Marengo & Aceto,

72 2003).

73 The aim of the present study is to investigate the REEs, major, trace and ultratrace element content

74 to characterise samples of Chianti/Chianti Classico authentic red wine from vineyards of Toscana

75 region (Italy) and differentiate them from wines of other Italian regions of production. ICP-MS is a

76 suitable technique for accurate and fast determination of major and trace elements in the same

77 sample. In order to measure REEs and ultra-trace elements we used a desolvating nebulizer coupled

78 with ICP-MS (Spalla et al., 2009; Bentlin et al., 2012; Aceto et al. 2013) to improve the sensitivity

79   and reduce the intensity ratios of REE oxide/REE.

80   We further statistically analysed the chemical element composition of wine samples to design the

81   best model able to recognize the classes to which each sample belonged, including samples of

82   Chianti/Chianti Classico micro-vinified, that expressed the most natural, unaltered characteristics of

83   a single vineyard block.

84   In particular, analytical fingerprints based on the concentrations of some major, trace, ultratrace

85   elements and REEs were used to calibrate multivariate classification models able to distinguish

86   Chianti from non-Chianti samples. The capability of classification models to correctly indicate the

87   origin of production was evaluated by means of validation protocols. The relevance of each element

88   for the characterisation of Chianti samples was investigated and, finally, further studies were

89   conducted in order to analyse which Italian regions show major overlaps in the chemical

90   fingerprints of wine samples with respect to Chianti.

91

92   **2.   MATERIALS AND METHODS**

93

94   2.1.   Wine samples

95   The wine samples were obtained by microvinification of grapes harvested directly in the vineyard.

96   The sampling team collected grape samples from different regions of Italy, which were selected

97   among those that are well-suited to red wine production (Table S1). The sampling campaign was

98   carried out in 2013 (246 collected grapes) and 2014 (393 collected grapes). From all 639 red wine

99   samples, 122 were Sangiovese, 95 Merlot, 68 Montepulciano, 42 Cabernet Sauvignon, 33 Barbera,

100   32 Aglianico, 28 Syrah, 21 Nebbiolo, 20 Nero d'Avola, 17 Petit Verdot, 15 Cannonau, 13 Croatina,

101   10 Bovale Sardo and 123 of 34 other varieties.

102

103   2.2.   Major and trace element determinations

104   The analytical measurements of major and trace elements investigated (As, Ba, Co, Fe, Li, Mg, Mn,

4

105 Na, Ni, Rb, Sr, Ti, Tl and V) were carried out using an ICP-MS instrument (7700x model; Agilent

106 Technologies, Santa Clara, CA, USA). The analysis of wine can be performed directly after 1:20

107 dilution with nitric acid at 2% (obtained by dilution of HNO_3 67% suprapure, Merck -Darmstadt,

108 Germany- with Milli-Q Water - electrical resistivity >18MΩ cm). Standard solutions were prepared

109 from 1000 mg L^{-1} ICP-MS stock solutions (ICPMS calibration standard, ULTRA Scientific, North

110 Kingstown, Rhode Island). The operating conditions of the ICP-MS are listed in Table S2.

111

112 2.3.   REE and ultratrace element determinations

113 Wine samples (10 mL) were digested with 6 mL of HNO_3 67% (suprapure, Merck) and 4 mL of

114 H_2O_2 30% (Merck) in a heating block system (DIGIPREP, SCP Science, Champlain, NY, USA) in

115 polypropylene tubes (digiTUBES, SCP Science) at 130°C for 3 hours. The digested solutions were

116 diluted to 50 mL with Milli-Q Water and filtered by using 0.45 μm filter (digiFILTER, SCP

117 Science) before the determination of Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Re, Sm, Ta, Tb, Tm,

118 V, Y, and Yb by ICP-MS. The ICP-MS instrument was equipped with a nebulization/desolvation

119 sample introduction system (Apex; Elemental Scientific, Omaha, NE, USA). The solution is sucked

120 inside a concentric nebulizer micro-PFA and is completely vaporized within a spray chamber

121 maintained at 140 °C. The aerosol passes through a condenser Peltier cooled to a temperature of 2

122 °C and is desolvated by the heated macro-porous Teflon Membrane Desolvation (Spiro TMD,

123 Elemental Scientific) that further reduces the amount of the solvent. In this way the sensitivity

124 increases of at least ten times and the oxides are reduced to 0.1% (CeO^+/Ce). Standard solutions

125 were prepared from a 10 mg L^{-1} REE multi-element ICPMS stock solution (ICP-MS calibration

126 standard, CPI International). The operating conditions for the instrument are listed in Table S3.

127

128 2.4.   Statistical analysis

129 Experimental data were organized in a numerical matrix with 639 rows (wine samples) and 31

130 columns (element determinations), where each x_{ij} element of the matrix represents the concentration

131 of the *j*-th element in the *i*-th wine sample. In order to reduce the variability scale of concentrations,

132 the data matrix was log-scaled before the multivariate statistical analysis.

133 The dataset was preliminarily investigated by means of K-Contractive Maps (K-CM) method

134 (Buscema, 2014) to analyze data patterns and evaluate the relationships among the different regions

135 of wine production. K-CM is a machine learning method that combines an artificial neural network

136 (ANN) for the non-linear optimization of the variable relationships and a fuzzy profiling of the

137 samples. In particular, K-CM exploits the non-linear variable relationships provided by the Auto-

138 CM neural network (Buscema, 2008) to obtain a fuzzy profiling of the samples and then applies

139 local similarity analysis to evaluate significant patterns in the data. K-CM is capable of supporting

140 its decisions by using a weighted semantic map of samples and a variable semantic graph, which

141 shows the relationships among all the variables of the training samples.

142 Supervised classification was successively applied in order to distinguish Chianti *vs*. non-Chianti

143 samples by means of predictive multivariate models. Classification methods aim to develop

144 mathematical models to recognize the membership of each sample to its proper class on the basis of

145 a selected set of variables (i.e., chemical element concentrations). The classification model can be

146 then utilized to predict the membership of unknown samples to the Chianti or non-Chianti class.

147 In this work, we used Partial Least Squares - Discriminant Analysis (PLS-DA), which is a linear

148 classification method that combines the properties of Partial Least Squares (PLS) regression with

149 the discrimination power of a classification technique. The underlying PLS regression algorithm

150 identifies latent variables with a maximum covariance with the classes (Höskuldsson, 1988). When

151 dealing with PLS-DA, the class vector is transformed into a dummy matrix Y, which represents the

152 membership of each sample in a binary form. The PLS2 model is then calibrated on the Y matrix

153 (Barker, 2003) and then the probability that a sample belongs to a specific class can be calculated

154 on the basis of the estimated class values (Pérez, 2009). Thus, each modelled class can be described

155 by a classification function reporting the coefficients that determine the linear combination of the

156 original variables to define the classification score. Before the PLS-DA calculation, data were

157 autoscaled.

158 Classification models were validated through a strict protocol based on random data splitting. This

159 ensured a reliable assessment of the predictive performance, that is, the capability to correctly

160 predict unknown sample as Chianti or non-Chianti. Samples were randomly assigned to a training

161 or a test set, which included 479 and 160 samples, respectively (i.e., 75% and 25% splitting). The

162 so-called stratified splitting was performed, by maintaining the class proportion of the whole set in

163 the training and test sets. The training set thus included 94 Chianti and 385 non-Chianti samples,

164 while the test set 31 and 129 Chianti and non-Chianti samples, respectively. The 479 training

165 samples were used to calibrate the classification models. To this end, an internal cross-validation

166 procedure with 5 cancellation groups was used to estimate the optimal number of PLS-DA latent

167 variables. Test samples were used to evaluate the predictive ability of the trained models only.

168 The performance of the classification models was encoded within the so-called confusion matrix,

169 that is, a square matrix having as many rows and columns as the number of classes. To each

170 column/row there corresponds one class, and rows and columns represent experimental and

171 predicted classes, respectively. Each entry c_{gk} of this matrix is the number of samples belonging to

172 g-th class and predicted as the k-th class. From the confusion matrix, sensitivity (Sn) and specificity

173 (Sp) can be calculated: these accuracy metrics are used to evaluate the model ability to correctly

174 predict Chianti and non-Chianti samples, respectively, and were calculated as follows (Ballabio,

175 2018):

$$Sn = \frac{TP}{TP + FN}$$

176

$$Sp = \frac{TN}{FP + TN}$$

177

178 where TP and FN indicate the number of Chianti samples correctly predicted (True Positive) and

179 the number of Chianti samples classified as non-Chianti (False Negative), respectively; TN and FP

180 represent the number of non-Chianti samples correctly classified (True Negative) and the number of

181 non-Chianti samples predicted as Chianti (False Positive), respectively. Finally, the Non-Error Rate

182 (NER) was calculated as the arithmetic mean of specificity and sensitivity.

183

184 2.5.   Software

185 Exploratory analysis by K-CM method was carried out by a dedicated software released by

186 SEMEION (Buscema, 1999). Classification models were calculated with the Classification Toolbox

187 for MATLAB (Ballabio, 2013).

188

189 **3.   RESULTS AND DISCUSSION**

190 3.1. ICP-MS analysis

191 The calibration, quality control of the multi-element determinations by ICP-MS system were

192 performed in agreement with the USEPA 6020B (2014) and UNI EN ISO 17294-2 (2005). The

193 method performance with respect to precision was aligned with Horwitz and Albert model (2006).

194 The Method Detection Limits (MDL) and results of the selected elements in wines in 2013 and

195 2014 are reported in Table S4.

196

197 3.2. Regional compositional patterns

198 The K-CM analysis was carried out on the data matrix (639 x 31) by using the 18 geographical

199 provenance regions as the targets. Prior to the K-CM analysis, all the variables were range scaled on

200 the minimum and maximum values so that each scaled variable was allowed to vary in the range 0–

201 1. In addition, the class variable was unfolded into a binary vector of dimension 18 (i.e., the number

202 of considered geographical regions), where 1 indicates the class membership.

203 In the first stage of the procedure, the Auto-CM neural network was run with the following

204 parameter settings: contraction parameter $C = \sqrt{p} = \sqrt{31} = 5.568$ and learning rate constant

205 $\alpha = \dfrac{1}{n} = \dfrac{1}{639} = 1.565 \times 10^{-3}$, where p is the number of variables and n the number of samples

206 characterising the dataset. During the learning phase, the 31 mono-dedicated connections between

207 the input and hidden layer were updated at the end of each epoch on the basis of the difference

208 between the values of the corresponding input and hidden nodes and further modulated for the input

209 node itself. All the variable weights converged to the same value (i.e., the contraction parameter C)

210 after 600 epochs. After convergence, all the linear and non-linear relationships between variables

211 were encoded into the connection weight matrix, which is the final output of Auto-CM. In the

212 second phase of K-CM procedure, the connection weight matrix, which encodes the full set of

213 relationships among the input variables, was used to transform the original dataset into a new fuzzy

214 dataset, more informative than the previous one. In the final step, the class memberships of the wine

215 samples were determined according to the same principle as the k-nearest neighbor (kNN) classifier

216 (Kowalski, 1972).

217 A graphical visualization of the wine samples and their distribution in the 18 considered Italian

218 regions is achieved by the semantic map of Figure 1, which is based on a Minimum Spanning Tree

219 (MST) calculated on the meta-distances between samples. The meta-distance is a measure of

220 dissimilarity between two samples, which takes into account the distances they have from all the

221 training samples. The MST is often used as a filter on the distance matrix to highlight the most

222 significant relationships between samples through the concepts of the graph theory (Kruskal, 1956).

223 The MST is what we might call the nervous system of any dataset; it is an undirected graph with

224 many vertices as the samples being evaluated, completely linked each other and colored according

225 to the class membership.

226 From the semantic map, it is apparent how the majority of wine samples taken from the same region

227 tend to cluster forming well-defined branches of the tree. In particular, samples from Sardinia

228 (white), Basilicata (light green), Sicily (turquoise), Apulia (bright green) and Piedmont (yellow) are

229 well-separated meaning that the wine samples from these Italian regions have quite characteristic

230 elemental composition. Chianti samples (black) are mainly concentrated in the terminal upper part

231 of the tree, but they are distributed into different branches along with several samples from other

232 regions. In addition, this map clearly shows some anomalous Chianti samples that are spread

233   overall the tree. Since the aim of this work was the discrimination of Chianti wines (to protect the

234   DPO label), supervised classification models were then specifically calibrated in order to better

235   distinguish Chianti from samples of all other regions.

236

237   3.3. Multivariate classification models

238   PLS-DA classification models were calibrated on the 479 training samples with the aim of

239   discriminating between Chianti and non-Chianti wine samples on the basis of their chemical

240   profiles related to the concentrations of 31 chemical elements (major, trace and ultratrace elements).

241   The optimal number of latent variables was selected by an internal cross-validation procedure. The

242   model predictivity was assessed on the 160 test samples that were not used in the model calibration

243   process. Classification results are summarised in Tables 3 and 4.

244   The obtained PLS-DA model demonstrated a satisfactory capability to discriminate Chianti and

245   non-Chianti samples. In fact, when evaluating results achieved on the training set, 75 Chianti

246   samples out of 94 were correctly classified, leading to sensitivity equal to 0.80. Predictions

247   achieved on test samples confirmed this discrimination ability. In fact, 22 out of 31 Chianti test

248   samples were correctly classified (sensitivity equal to 0.71). Similar performance was obaserved for

249   non-Chianti samples: 304 (out of 385) training samples and 102 (out of 129) test samples were

250   correctly predicted as non-Chianti, thus leading to specificity values equal to 0.79 in both cases.

251   NER values equal to 0.79 and 0.75 for training and test sets (Table 4), respectively, indicate a

252   similar model performance in fitting and prediction, as well as the absence of overfitting.

253   The discriminant power of each chemical element in the PLS-DA model was investigated by the

254   coefficients related to the Chianti class (Figure 2). Such coefficients are used to linearly combine

255   the chemical concentrations to calculate the classification score for the Chianti class. The higher the

256   score, the higher the probability to predict a sample as belonging to the Chianti class. Positive large

257   coefficients are thus associated with those chemical elements that contribute more to determine the

258   Chianti membership, while the opposite happens for elements with highly negative coefficients.

259 Close-to-zero coefficients are related to elements with no relevant contribution in the discrimination

260 of Chianti wine samples. REEs, major, trace and ultratrace elements all contribute to the

261 discrimination of Chianti samples. Major and trace elements are mainly associated to large positive

262 or negative coefficients and thus appear to have a higher influence in the identification of Chianti

263 samples; in particular, Mg, Ti, and Co have large positive coefficients, while V, Li, As negative

264 ones. The contribution of REEs and ultratrace elements is not completely irrelevant, despite

265 associated to lower weights. In particular, Tb, Gd, Dy and Re have positive contribution to

266 determine the Chianti fingerprint, while Ta, Lu, Yb and Tm have negative coefficients.

267 Since non-Chianti wines were sampled over different Italian regions (Table S1) and some overlap

268 among regions arose from the exploratory analysis (Figure 1), further investigation was carried out

269 to evaluate which regions have higher or significant overlaps with respect to Chianti, that is, which

270 regions provide wines with chemical fingerprints similar to that of Chianti samples. Such regional

271 analysis was enhanced by taking into considerations prediction errors on non-Chianti training and

272 test samples. In particular, 81 training and 27 test non-Chianti samples were incorrectly classified

273 by the PLS-DA model, they being assigned to the Chianti class (Table 1 and Table 2, respectively).

274 These misclassifications can be further detailed in terms of the region of origin (Table S5, where the

275 total number of samples, the number of misclassified samples and the percentage of

276 misclassifications (number of misclassified samples over the total number of samples) are reported

277 for each region in both training and test sets. Most of the regions (i.e. Abruzzo, Sardinia, Apulia and

278 Piedmont) with a large number of samples are characterised by a small degree of overlap with

279 respect to Chianti. Only 6 training (out of 47, 13%) and 1 test (out of 10, 10%) samples from

280 Abruzzo were incorrectly classified as Chianti; similar results were achieved for Sardinia, Apulia

281 and Piedmont. On the other hand, Sicily (29% and 31% errors on the training and the test set,

282 respectively), Marches (30% and 50% on the training and test sets, respectively) and Umbria (41%

283 errors on the training set), were characterised by a remarkably higher percentage of errors. This

284 indicates that the calibrated PLS-DA model is not particularly suited to distinguish Chianti wines

285 from samples from such regions.

286

287 3.4. Specific classification models to distinguish critical regions

288 The unsatisfactory performance of the PLS-DA model on some regions can be due to the calibration

289 of a "general" model, which makes it hard to find a unique linear combination of variables that

290 grasp all the regional differences between wines. On this basis, specific classification models were

291 calibrated to discriminate Chianti samples from the regions characterized by critical error

292 percentages. Sicily, Marches and Umbria were, thus, considered, as they not only are characterised

293 by high percentages of errors (29%, 30% and 41%, respectively) but also by a high number of

294 samples. Other regions demonstrated a significant overlap with Chianti (such as Emilia Romagna,

295 Friuli-Venezia Giulia, Trentino-South Tyrol), but, due to their low number of samples, they were

296 not taken into account.

297 Three region-based distinct models were calibrated, as follows:

298 1. *Sicily vs*. *Chianti wine samples*. The model was calibrated on 149 samples (94 from Chianti and

299   55 from Sicily) and validated on 44 samples (31 from Chianti and 13 from Sicily);

300 2. *Marches vs*. *Chianti wine samples*, with 140 training samples (94 from Chianti and 46 from

301   Marches) and 43 test samples (31 from Chianti and 12 from Marches);

302 3. *Umbria vs*. *Chianti wine samples*, with 128 training samples (94 from Chianti and 34 from

303   Abruzzo) and 38 test samples (31 from Chianti and 7 from Abruzzo).

304 The classification performance achieved on training and test samples for each regional model is

305 reported in Table 2. It is remarkable how the capability of discrimination of critical regions was

306 enhanced when modelled against Chianti one at a time. For example, the samples from Sicily are

307 well-distinguished from Chianti ones by the regional model: 93% and 85% of Sicily samples are

308 correctly predicted in the training and test set, respectively (Table 2), while around 30% of Sicilian

309 wine samples were misclassified in the general model (Table 1). Similar results, even if with a

310 lower performance on test samples, were achieved for Chianti *vs*. Marche and *vs*. Umbria samples,

311 with NER values larger than or equal to 0.72 for both training and test sets (Table 2).

312 The discriminating role of each chemical element was again assessed by looking at the

313 classification coefficients of the PLS-DA models (Table S5). Compared to the general classification

314 model, the discriminating capability of some elements for specific regions emerged. For example,

315 Ni was previously associated to a close-to-zero coefficient (Figure 2). However, when looking at

316 the discrimination between Chianti and non-Chianti samples from Sicily, Ni plays a role in

317 characterising Chianti wines. Similar conclusions can be extended to Na and Rb when

318 characterising Marches´s wines against Chianti. Finally, Sr emerged as a relevant chemical marker

319 when separating Chianti from Sicily, Marches and Umbria, while again it was associated a small

320 weight in the global classification model (Figure 2).

321

322 **4. CONCLUSIONS**

323 In this work, the combination of ICP-MS with chemometrics has demonstrated to be useful for

324 distinguish Chianti red wines from samples of other Italian region. 31 chemical elements (major,

325 trace and ultratrace elements), measured with direct dilution and acid digestion, represented the

326 chemical profile of each wine samples and the calibration of PLS-DA models. This general model

327 demonstrated a satisfactory capability to discriminate Chianti and non-Chianti samples. Accuracy

328 metrics, like sensitivity and specificity, improved with specific regional PLS-DA model. These

329 specific models introduced elements with high discriminant potentiality between Chianti and other

330 regions, compared with general model, like Sr.

331

335

336

13

337   Appendix A. Supplementary data

338

339   REFERENCES

340

341   Aceto, M., Robotti, E., Oddone, M., Baldizzone, M., Bonifacino, G., Bezzo, G., Di Stefano, R.,

342   Gosetti, F., Mazzucco, E., Manfredi, M., Marengo E. (2013). A traceability study on the

343   Moscato wine chain. *Food Chemistry*, *138*(2), 1914–1922.

344   Almeida, C. M. R., & Vasconcelos, M. T. S. D. (2003). Multielement composition of wines and

345   their precursors including provenance soil and their potentialities as fingerprints of wine

346   origin. *Journal of Agricultural and Food Chemistry*, *51*(16), 4788–4798.

347   Angus, N. S., O'keeffe, T. J., Stuart, K. R., & Miskelly, G. M. (2006). Regional classification of

348   New Zealand red wines using inductively-coupled plasma-mass spectrometry (ICP-MS).

349   *Australian Journal of Grape and Wine Research*, *12*(2), 170–176.

350   Ballabio, D., Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-

351   DA. *Analytical methods*, *5*, 3790-3798.

352   Ballabio, D., Grisoni, F., Todeschini, R. (2018). Multivariate comparison of classification

353   performance measures. *Chemometrics and Intelligent Laboratory Systems*, *174*, 33-44.

354   Barbaste, M., Medina, B., Sarabia, L., Ortiz, M. C., & Pérez-Trujillo, J. P. (2002). Analysis and

355   comparison of SIMCA models for denominations of origin of wines from de Canary Islands

356   (Spain) builds by means of their trace and ultratrace metals content. *Analytica Chimica Acta*,

357   *472*(1), 161–174.

358   Barker, M., Rayens, W.S. (2003). Partial least squares for discrimination. *Journal of Chemometrics*.

359   *17*, 166-173.

360   Bentlin, F.R.S., dos Santos, C.M.M., Flores, É.M.M., Pozebon, D. (2012). Lanthanides

361   determination in red wine using ultrasound assisted extraction, flow injection, aerosol

362   desolvation and ICP-MS. *Analytica Chimica Acta*, *710*, 33-39.

363   Buscema, M. (1999). Supervised ANNs and Artificial Organisms, ver. 22.0, Semeion Software,

364   Rome (Italy).

365   Buscema, M., Grossi, E., Snowdon, D., Antuono, P. (2008). Auto-Contractive Maps: an artificial

366   adaptive system for data mining. An application to Alzheimer disease. *Current Alzheimer*

367   *Research*, *5*, 481-498.

368   Buscema, M., Consonni, V., Ballabio, D., Mauri, A., Massini, G., Breda, M., Todeschini, R. (2014).

369   K-CM: a new artificial neural network. Application to supervised pattern recognition.

370   *Chemometrics and Intelligent Laboratory Systems*, *138*, 110-119.

371   Coetzee, P. P., Steffens, F. E., Eiselen, R. J., Augustyn, O. P., Balcaen, L., & Vanhaecke, F. (2005).

372   Multi-element analysis of South African wines by ICP-MS and their classification according

373   to geographical origin. *Journal of Agricultural and Food Chemistry*, *53*(13), 5060–6.

374   http://doi.org/10.1021/jf048268n

375   Coetzee, P. P., Van Jaarsveld, F. P., & Vanhaecke, F. (2014). Intraregional classification of wine

376   via ICP-MS elemental fingerprinting. *Food Chemistry*, *164*, 485–492.

377   Di Paola-Naranjo, R. D., Baroni, M. V, Podio, N. S., Rubinstein, H. R., Fabani, M. P., Badini, R.

378   G., … Gallegos, E. (2011). Fingerprints for main varieties of Argentinean wines: terroir

379   differentiation by inorganic, organic, and stable isotopic analyses coupled to chemometrics.

380   *Journal of Agricultural and Food Chemistry*, *59*(14), 7854–7865.

381   D.P.R. 02.07.1984. Gazzetta Ufficiale della Repubblica Italiana n. 290, 20 October 1984.

382   EU (2013) Regulation (EU) No 1308/2013 Of The European Parliament And Of The Council of 17

383   December 2013 establishing a common organisation of the markets in agricultural products

384   and repealing Council Regulations (EEC) No 922/72, (EEC) No 234/79, (EC) No 1037/2001

385   and (EC) No 1234/2007 https://eur-lex.europa.eu/legal-

386   content/EN/ALL/?uri=celex%3A32013R1308

387   Galgano, F., Favati, F., Caruso, M., Scarpa, T., & Palma, A. (2008). Analysis of trace elements in

388   southern Italian wines and their classification according to provenance. *LWT - Food Science*

389   *and Technology*, *41*(10), 1808–1815. http://doi.org/10.1016/j.lwt.2008.01.015

390   Geana, I., Iordache, A., Ionete, R., Marinescu, A., Ranca, A., & Culea, M. (2013). Geographical

391   origin identification of Romanian wines by ICP-MS elemental analysis. *Food Chemistry*,

392   *138*(2), 1125–1134.

393   Gonzálvez, A., Llorens, A., Cervera, M. L., Armenta, S., & De la Guardia, M. (2009). Elemental

394   fingerprint of wines from the protected designation of origin Valencia. *Food Chemistry*,

395   *112*(1), 26–34.

396   Greenough, J. D., Mallory-Greenough, L. M., & Fryer, B. J. (2005). Geology and wine 9: Regional

397   trace element fingerprinting of Canadian wines. *Geoscience Canada*, *32*(3).

398   Gremaud, G., Quaile, S., Piantini, U., Pfammatter, E., & Corvi, C. (2004). Characterization of

399   Swiss vineyards using isotopic data in combination with trace elements and classical

400   parameters. *European Food Research and Technology*, *219*(1), 97–104.

401   http://doi.org/10.1007/s00217-004-0919-0.

402   Horwitz W. and Albert R.J. (2006). The Horwitz Ratio (HorRat): A Useful Index of Method

403   Performance with Respect to Precision. *Assoc*. *Off*. *Anal*. *Chem*., 89, 4, 1095- 1109.

404   Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, *2*, 211–228.

405   Iglesias, M., Besalú, E., & Anticó, E. (2007). Internal standardization− atomic spectrometry and

406   geographical pattern recognition techniques for the multielement analysis and classification

407   of Catalonian red wines. *Journal of Agricultural and Food Chemistry*, *55*(2), 219–225.

408   ISTAT.IT [database on the Internet] 2017 Italian wine production Istituto Nazionale di Statistica,

409   Rome − Italy. Available from: http://www.thenumbersofwine.com/2017/07/19/2016-italian-

410   wine-production/

411   Kment, P., Mihaljevič, M., Ettler, V., Šebek, O., Strnad, L., & Rohlová, L. (2005). Differentiation

412   of Czech wines using multielement composition−A comparison with vineyard soil. *Food*

413   *Chemistry*, *91*(1), 157–165.

414   Kowalski, B.R., Bender, C.F. (1972). The K-Nearest Neighbor classification rule (Pattern

415   Recognition) applied to nuclear magnetic resonance spectral interpretation. *Analytical*

416   *Chemistry*, *44*, 1405-1411.

417   Kruskal, J.B. (1956). On the shortest spanning subtree of a graph and the Traveling Salesman

418   Problem. *Proceeding of the American Mathematical Society*, *7*, 48-50.

419   Marengo, E., & Aceto, M. (2003). Statistical investigation of the differences in the distribution of

420   metals in Nebbiolo-based wines. *Food Chemistry*, *81*(4), 621–630.

421   Medina, B. (1996). Wine authenticity. In P. R. Ashurst & M. J. Dennis (Eds.), Food authentication

422   (pp. 60–107). London: Blackie.

423   Ortega, M., Gonzalez-San Jose, M. L., & Beltrán, S. (1999). Metal content of Spanish red wines

424   from certified denomination of origin. *Quimica Analitica*, *18*(1), 127–131.

425   Pérez, N.F., Ferré, J., Boqué, R. (2009). Calculation of the reliability of classification in

426   discriminant partial least-squares binary classification. *Chemometrics and Intelligent*

427   *Laboratory Systems*, *95*, 122–128.

428   Regulation, C. (1992). Council Regulation (EEC) No. 2081/92 of 14 July 1992 on the protection of

429   geographical indications and designations of origin for agricultural products and foodstuffs.

430   Off. *J. Eur. Union*, *208*, 1–8.

431   Šelih, V. S., Šala, M., & Drgan, V. (2014). Multi-element analysis of wines by ICP-MS and ICP-

432   OES and their classification according to geographical origin in Slovenia. *Food Chemistry*,

433   *153*, 414–423.

434   Spalla, S., Baffi, C., Barbante, C, Turretta, C., Cozzi, G., Beone, G.M., Bettinelli, M. (2009).

435   Determination of rare earth elements in tomato plants by inductively coupled plasma mass

436   spectrometry techniques. *Rapid Communications in Mass Spectrometry*, 23, 3285-3292.

437   Taylor, V. F., Longerich, H. P., & Greenough, J. D. (2003). Multielement analysis of Canadian

438   wines by inductively coupled plasma mass spectrometry (ICP-MS) and multivariate

439   statistics. *Journal of Agricultural and Food Chemistry*, *51*(4), 856–60.

440   http://doi.org/10.1021/jf025761v

441 Thiel, G., Geisler, G., Blechschmidt, I., & Danzer, K. (2004). Determination of trace elements in

442   wines and classification according to their provenance. *Analytical and Bioanalytical*

443   *Chemistry*, *378*(6), 1630–1636.

444 US Environmental Protection Agency, 2014. Method 6020B: Induc-tively coupled plasma-mass

445   spectrometry. SW-846.

446 UNICHIM, 2005. UNI EN ISO 17294-2:2005 Qualità dell acqua – Applicazione della spettrometria

447   di massa al plasma accoppiato induttivamente (ICP-MS) – Parte 2.

448 Van der Linde, G., Fischer, J. L., & Coetzee, P. P. (2010). Multi-element analysis of South African

449   wines and their provenance soils by ICP-MS and their classification according to

450   geographical origin using multivariate statistics. *South African Journal for Enology &*

451   *Viticulture*, *31*(2), 143.

452

TABLES

Table 1. Classification errors for non-Chianti training and test samples. For each region, the total number of samples, the number of misclassified samples and the percentage of errors are reported for both training and test sets. Regions are sorted according to the number of samples they contain, in descending order.

Table 2. Indices of classification performance obtained on training and test sets for the specific regional PLS-DA models discriminating Chianti and critical regions (Sicily, Marches, Umbria); Sn indicates Chianti sensitivity, Sp indicates Chianti specificity, NER is the average of Sn and Sp.

Table 3. Confusion matrices obtained on 479 training samples and 160 test samples by means of PLS-DA classification model.

Table 4. Classification performance on training and test sets: sensitivity (Sn), specificity (Sp) and Non-Error Rate (NER) on Chianti class are reported.

Table 1. Classification errors for non-Chianti training and test samples. For each region, the total number of samples, the number of misclassified samples and the percentage of errors are reported for both training and test sets. Regions are sorted according to the number of samples they contain, in descending order.

	Training			Test		
Region	No. samples	No. misclassified	% errors	No. samples	No. misclassified	% errors
Sicily	55	16	29%	13	4	31%
Abruzzo	47	6	13%	10	1	10%
Marches	46	14	30%	12	6	50%
Sardinia	42	1	2%	10	0	0%
Umbria	34	14	41%	7	0	0%
Piedmont	31	5	16%	12	1	8%
Apulia	28	1	4%	12	0	0%
Emilia Romagna	14	5	36%	8	4	50%
Basilicata	13	0	0%	5	1	20%
Lombardy	13	4	31%	4	0	0%
Friuli-Venezia Giulia	12	5	42%	5	2	40%
Lazio	10	1	10%	9	0	0%
Veneto	10	1	10%	5	4	80%
Calabria	9	1	11%	3	0	0%
Campania	8	2	25%	7	1	14%
toscana	8	2	25%	2	1	50%
Trentino-South Tyrol	5	3	60%	5	2	40%

Table 2. Indices of classification performance obtained on training and test sets for the specific regional PLS-DA models discriminating Chianti and critical regions (Sicily, Marches, Umbria); Sn indicates Chianti sensitivity, Sp indicates Chianti specificity, NER is the average of Sn and Sp.

Set	Sicily			Marches			Umbria		
	NER	Sn	Sp	NER	Sn	Sp	NER	Sn	Sp
Training	0.92	0.90	0.93	0.87	0.87	0.87	0.83	0.80	0.85
Test	0.89	0.94	0.85	0.72	0.77	0.67	0.74	0.77	0.71

Table 3. Confusion matrices obtained on 479 training samples and 160 test samples by means of PLS-DA classification model.

	Training		Test	
	Chianti	non-Chianti	Chianti	non-Chianti
Chianti	75	19	22	9
non-Chianti	81	304	27	102

Table 4. Classification performance on training and test sets: sensitivity (Sn), specificity (Sp) and Non-Error Rate (NER) on Chianti class are reported.

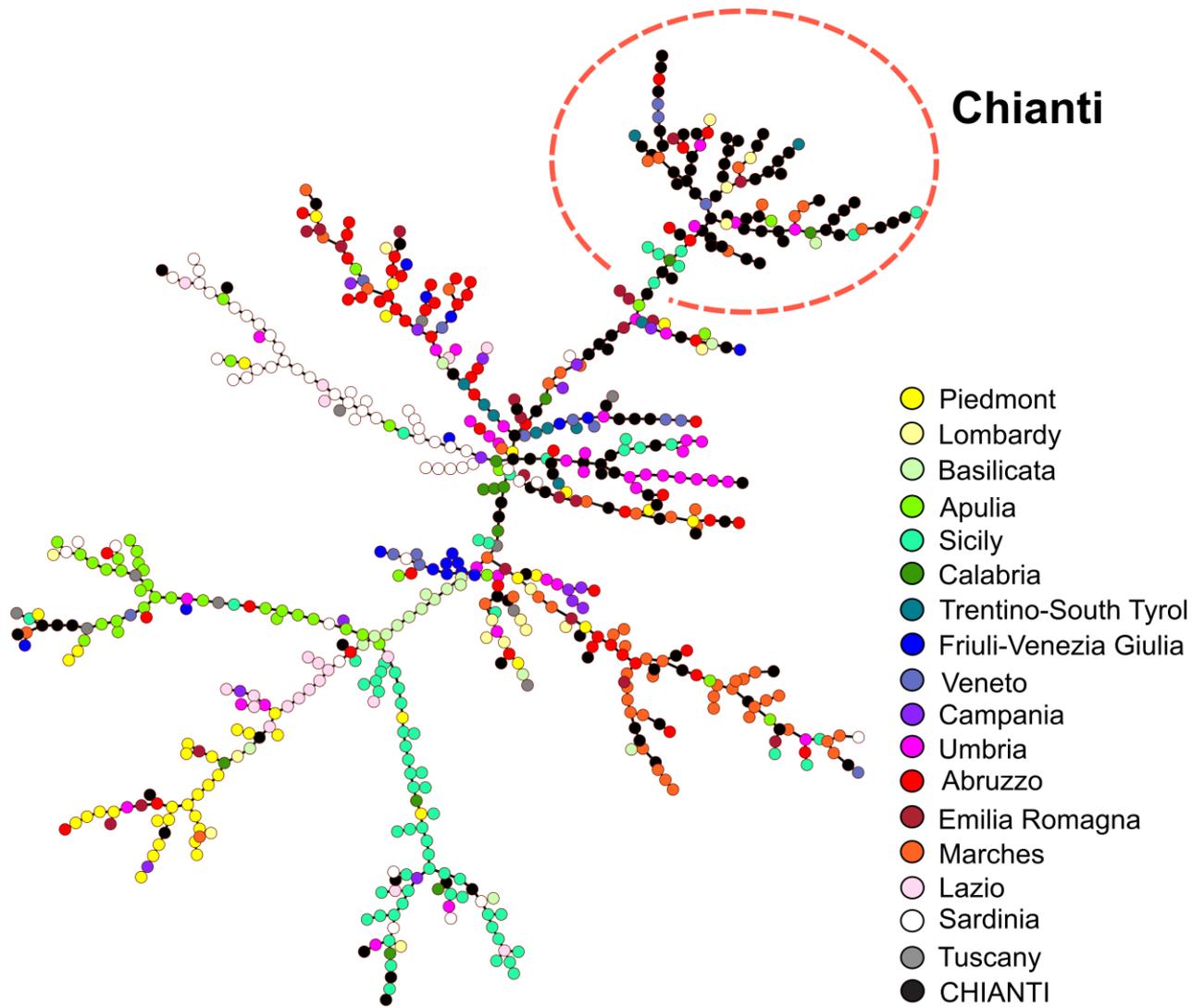|	**NER**	**Sn**	**Sp**
training set	0.79	0.80	0.79
test set	0.75	0.71	0.79

FIGURES

Figure 1. Semantic map of the wine dataset obtained as a Minimum Spanning Tree. The tree nodes represent the wine samples, colored according to the provenance regions.

Figure 2. PLS-DA coefficients for the Chianti class associated to the 31 REEs, major, trace and ultratrace elements.

Figure 1. Semantic map of the wine dataset obtained as a Minimum Spanning Tree. The tree nodes represent the wine samples, colored according to the provenance regions.
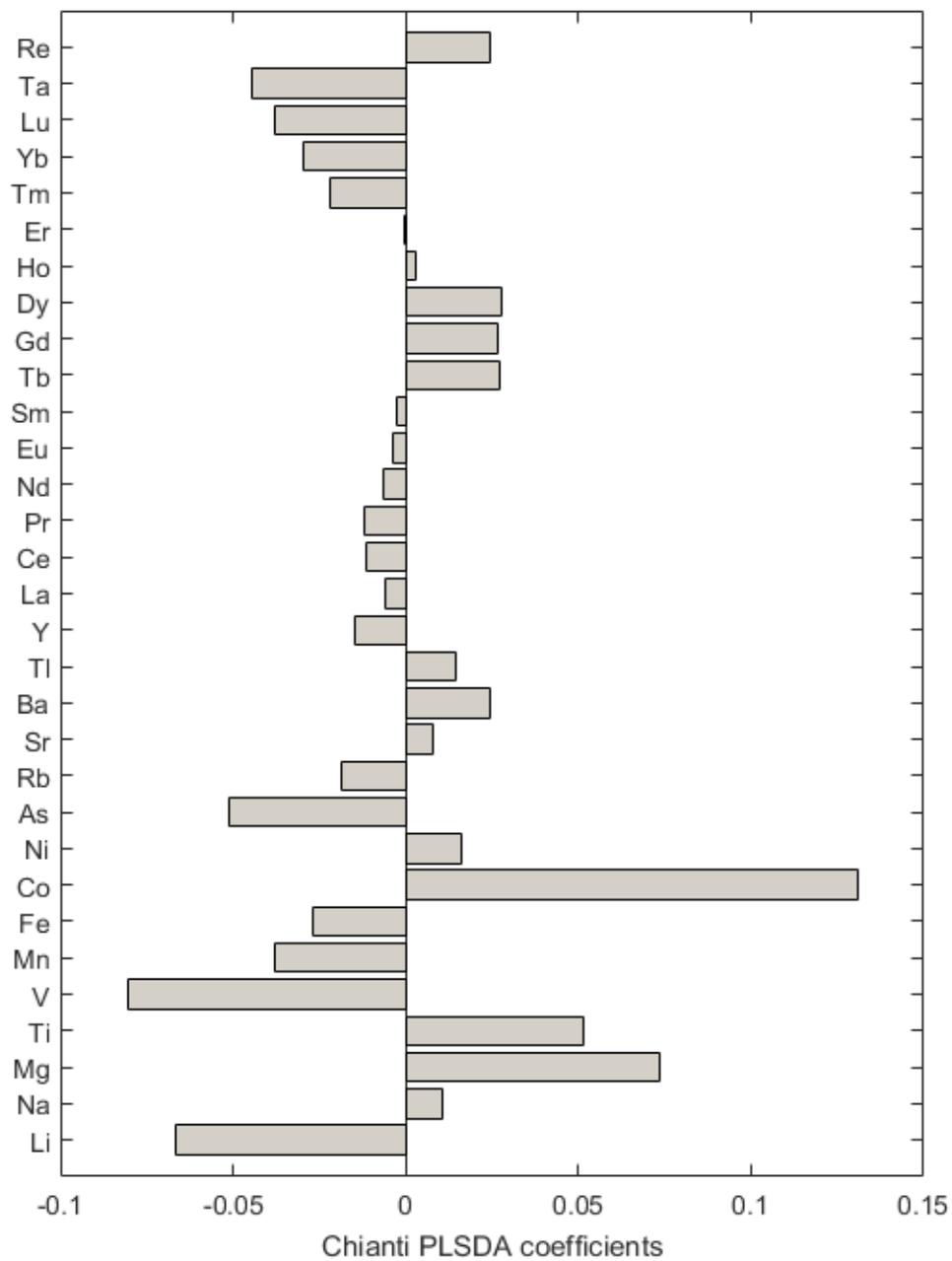
Figure 2. PLS-DA coefficients for the Chianti class associated to the 31 REEs, major, trace and ultratrace elements.

Supplementary Material