



HAL
open science

Enhancing wine authentication: leveraging 12,000+ international mineral wine profiles and artificial intelligence for accurate origin and variety prediction

Leticia Sarlo, Coraline Duroux, Yohann Clément, Pierre Lanteri, Fabien Rossetti, Olivier David, Augustin Tillement, Philippe Gillet, Agnès Hagège, L. David, et al.

► To cite this version:

Leticia Sarlo, Coraline Duroux, Yohann Clément, Pierre Lanteri, Fabien Rossetti, et al.. Enhancing wine authentication: leveraging 12,000+ international mineral wine profiles and artificial intelligence for accurate origin and variety prediction. *OENO One*, 2024, 58 (4), <10.20870/oeno-one.2024.58.4.8107>. <hal-04793165>

HAL Id: hal-04793165

<https://hal.science/hal-04793165v1>

Submitted on 20 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

1 **Enhancing Wine Authentication: Leveraging 12,000+** 2 **International Mineral Wine Profiles and Artificial** 3 **Intelligence for Accurate Origin and Variety Prediction**

4 Leticia Sarlo^{1,2}, Coraline Duroux², Yohann Clément³, Pierre Lanteri³, Fabien Rossetti¹, Olivier
5 David⁴, Augustin Tillement^{2,5}, Philippe Gillet², Agnès Hagège³, Laurent David⁵, Michel Dumoulin⁶,
6 Richard Marchal^{7,8}, Théodore Tillement², François Lux^{1,9,*}, Olivier Tillement¹

7

8 ¹ Institut Lumière-Matière, UMR 5306, Université Claude Bernard Lyon 1-CNRS, Université de
9 Lyon, Villeurbanne Cedex 69100, France

10 ² M&Wine, 305 rue des Fours, 69270 Fontaines Saint Martin, France

11 ³ Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, UMR 5280, -CNRS,
12 Villeurbanne Cedex 69100, France

13 ⁴ Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

14 ⁵ Université Claude Bernard Lyon 1, Institut National des Sciences Appliquées, Université Jean
15 Monnet, CNRS, UMR 5223, Ingénierie des Matériaux Polymères, 15 bd Lattarjet, 69622
16 Villeurbanne, France

17 ⁶ Agro Œno Conseil, Mâcon, France

18 ⁷ Université de Reims Champagne-Ardenne, Laboratoire d'Oenologie, BP-1039, 51687 Reims
19 Cedex 02, France

20 ⁸ Université de Haute-Alsace, LVBE, 68008 Colmar Cedex, France

21 ⁹ Institut Universitaire de France (IUF), Paris

22 *corresponding author: francois.lux@univ-lyon1.fr

23 **Abstract**

24 For the wine industry, ensuring quality and authenticity hinges on the precise determination of wine
25 origin. In our study, we developed a fast semi-quantitative method to analyse 41 chemical elements
26 in wine, employing Inductively Coupled Plasma Mass Spectrometry (ICP-MS). This methodology
27 characterises what we term the Mineral Wine Profile (MWP). In contrast to an organic molecular
28 profile, the mineral composition of a wine remains constant from the moment it is bottled. Mineral
29 elements play a crucial role in the terroir of wine: they pass primarily from soil to grape and are then

30 influenced by various vinification techniques. Indeed, it is widely recognised that the original soil
31 characteristics are altered by a multitude of winemaking procedures, presenting a considerable
32 challenge when endeavouring to extract origin-related information in a typical scenario. Our study
33 demonstrates that statistical analyses and artificial intelligence (AI) could be a tool for accurately
34 deciphering origin information within the MWP, provided sufficient mineral elements are measured
35 and a comprehensive database of wine samples is employed to establish effective learning. In this
36 study, a dataset comprising 12966 MWPs was created in just over a year. The first analysis revealed
37 correlations between the elements in wine, especially between rare earth elements, between
38 macronutrients and between micronutrients. A machine learning method was then developed to
39 assess wine origin and principal grape variety. Six models were tested by comparing the area under
40 the receiver operating characteristic curve (AUC), with Extreme Gradient Boosting as the chosen
41 model. Mean accuracies of 92% for country classification, 91% for the French wine region, and 85%
42 for the main grape variety were obtained, and mean AUC scores of 0.964 for country classification,
43 0.961 for the French wine region and 0.914 for the main grape variety. This study represents the first
44 comprehensive investigation at this scale on wine samples, and underscores the importance of using
45 a comprehensive MWP dataset for AI applications when verifying wine origin. The authentication
46 of a wine with over 99% specificity could be routinely achievable through this approach.

47

48 Wine, ICP-MS, Elemental composition, Geographical origin, Machine learning, Extreme Gradient
49 Boosting, Mineral Wine Profile

50

51 **Introduction**

52

53 Each wine possesses a distinct character, primarily shaped by the intricate interplay between its
54 terroir and the winemaking process. Terroir, closely linked to the unique combination of vine-soil
55 dynamics, climatic conditions and the topography of a wine-producing region, exerts a profound
56 influence on the final product (Leeuwen, 2020). The different stages of winemaking, from the
57 harvesting of the grapes to bottling, constitute another pivotal factor in defining a wine's identity
58 (Castiñeira, 2004). Today, wine authenticity and unique expression are matters of debate between
59 specialists, and the lack of chemical characterisation to ensure its origin can lead to counterfeiting.

60 Three approaches to addressing these challenges are frequently reported in the literature: DNA
61 analysis (Baleiras-Couto, 2006), determination of wine organic compounds (i.e., polyphenols and
62 volatile compounds) and mineral profiling (Popîrdă, 2021), which can all be considered the

63 fingerprint of a wine sample. The first technique is applied to the identification of grape variety,
64 relying on the recovery of DNA from the beverage following the identification of a suitable sequence
65 which characterises the species (Villano, 2017). However, this approach is only suitable for young
66 wines, because DNA degradation over time gradually hinders the identification of a wine sample
67 (Villano, 2017; Zambianchi, 2022).

68 Nuclear magnetic resonance (NMR), which has commonly been used in food sciences for several
69 decades (Hatzakis, 2019), has gained in popularity in recent years as a tool for wine organic
70 compound screening and analysis (Le Mao, 2023). Another technique that is used is the combination
71 of gas chromatography (GC) and mass spectrometry (MS) to establish the organic profile (Schartner,
72 2023). Both methodologies rely on organic component analysis, but because these molecules are
73 sensitive to natural evolution (ageing) or premature changes (oxidation, impacts of storage
74 conditions) to the wine (Zhang, 2023), it is challenging to compare the same sample over time.

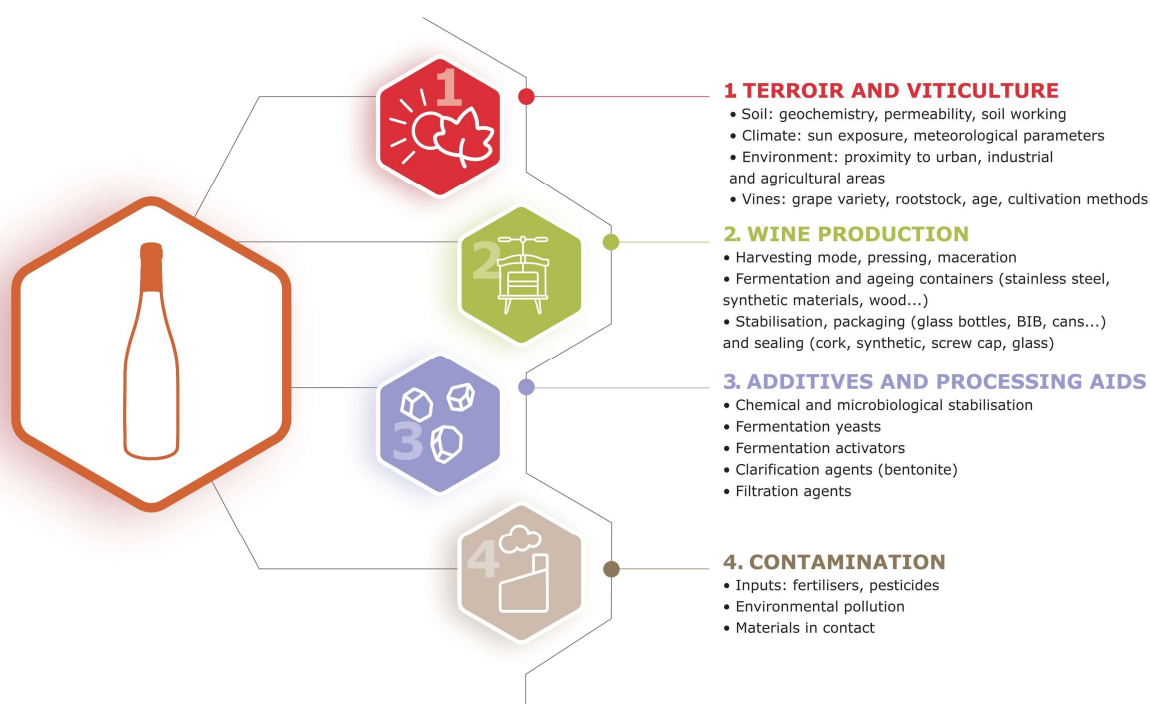
75 The measurement of isotopes can be carried out to quantify both organic and inorganic profiles,
76 playing a significant role in determining the origin of wines. Among the commonly used techniques
77 is the analysis of hydrogen and oxygen isotopes by isotope ratio mass spectrometry (IRMS) (Li,
78 2023) and the analysis of carbon isotopes by liquid chromatography coupled with IRMS (LC-IRMS)
79 (Perini & Bontempo, 2022). Isotopes of heavy elements, which can be analysed by Inductively
80 Coupled Plasma Mass Spectrometry (ICP-MS), such as lead or strontium, have proven to be suitable
81 for tracing the origin of food products (Drivelos, 2012), including wine (Cellier, 2021; Su, 2023).
82 However, sample preparation is time-consuming and costly, as it involves multiple steps, such as dry
83 evaporation, purification, and extraction, which acts as a barrier to constructing a rich and
84 comprehensive database.

85 To overcome the limitations posed by the evolution of organic compounds, determining elementary
86 inorganic content in wines is an alternative way of assessing the fingerprint of wines. This mineral
87 fingerprint is the Mineral Wine Profile (MWP). The concentration of different elements is influenced
88 by the terroir and winemaking processes, as schematised in Figure 1, and can be analysed using ICP-
89 MS, a robust and reliable technique (Lima, 2021). When coupled with multivariate statistical data
90 analysis methods, a classification of the origin of this food product has been shown to be possible
91 (Ellis, 2012; Giaccio, 2008). The most popular statistical methods are usually principal component
92 analysis (Bentlin, 2011; Lima, 2023) and discriminant analysis (Griboff, 2021; Pasvanka, 2021),
93 which can be employed alongside machine learning classification algorithms (Astray, 2021; Da
94 Costa, 2020).

95 Despite these numerous attempts to develop authentication methods, existing studies are constrained
96 by their focus on specific parameters, such as individual countries (Pasvanka, 2021), regions (Alonso

97 Gonzalez, 2021), wine appellations (Astray, 2021), and grape varieties (Da Costa, 2020; Tanabe,
98 2020). This often comes at the cost of restricted sample collection size, as the aforementioned studies
99 are based on a range of 14 to 113 samples, limiting the statistical significance and the broader
100 applicability of the findings.

101 In order to conciliate the need for a cost-efficient and reliable method of wine analysis with the need
102 for an extensive database, we developed a fast semi-quantitative (SQ) analytical method. This
103 method uses ICP-MS, which is capable of quantifying around forty mineral elements significantly
104 present in wines that constitute the MWP. Our approach stands out from existing solutions due to
105 the creation of an “oenothèque” and database comprising several thousand international wines.
106 Through statistical analysis and using the extreme gradient boosting algorithm, which was trained
107 on the thousands of MWP from our database, our goal was to establish timeless traceability of wine
108 blends and to be able to determine the origin of an unknown wine after ICP-MS analysis of a
109 collected 30 mL sample.



110

111 **Figure 1. Description of different factors that can influence the concentration of major, minor**
112 **and trace mineral elements in bottled wine, namely terroir and viticulture, wine production,**
113 **additives and contamination.**

114 **Materials and methods**

115 **1. Reagents and materials**

116 All utilised reagents were of analytical grade. Ultrapure water (MilliQ[®], 18.2 mΩ.cm) and nitric acid
117 Suprapur[®] grade (69% (v/v), Roth) were used for sample dilution and standard preparation. Certified

118 metal-free tubes (VWR[®]) were employed for collecting and preparing both samples and standards.
119 A semi-quantitative calibration standard was prepared by diluting the multi-element standard
120 (Reference 85006.186), purchased from VWR, with 100 mg/L of Al, Ag, As, B, Ba, Be, Bi, Ca, Cd,
121 Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Ti, Tl, V and Zn, in 5% HNO₃ (v/v). ICP-
122 MS tuning solution, containing 1 µg/L of Ce, Co, Li, Tl and Y in 2% HNO₃ (v/v) (Agilent
123 Technologies), was used to optimise the ICP-MS signal intensities. The use of these solutions is
124 described below. Moreover, a commercial red wine was used as a final control to ensure reproducible
125 results (i.e., not exceeding 15% variations) over time. An indium solution, used as an internal
126 standard to ensure good sample conservation, was prepared by diluting the 1000 mg/L indium
127 standard in 4% HNO₃ (v/v) (purchased from SCP Science) and then added to each sample.

128 **2. Sample preparation**

129 Wine samples were collected from wine contests organised in France and their descriptive data was
130 provided by the organisers. Their origin and grape variety are assured by the French decree NOR:
131 ESSC1303876A. This decree obliges contest organisers to verify the authenticity of wines entered
132 in the competition and winemakers to declare the varieties employed during vinification.

133 Samples of approximately 30 mL of wine were put in metal-free tubes. Direct wine dilution was
134 found to provide optimal balance in terms of user-friendliness, result accuracy and precision
135 (Godshaw, 2017). Samples were diluted 1:3 using 1% HNO₃ (v/v) and 10 µg/L of indium standard
136 solution. This initial dilution provides sample storage in acidic conditions, ensuring the preservation
137 of elemental composition over time and limiting mineral precipitation and adsorption onto the metal-
138 free tube walls. A second dilution of 1:5 using 1% HNO₃ (v/v) was performed just before the
139 analysis. The total dilution factor (1:15) was fine-tuned to minimise matrix effects, which can occur
140 due to the presence of alcohol or other organic matter when performing trace element quantification
141 (Catarino, 2006).

142 **3. ICP-MS analysis**

143 The ICP-MS measurements were made between June 2022 and October 2023 at the Université Lyon
144 1, Institut des Sciences Analytiques, using different quadrupole-ICP-MS equipment. The majority of
145 the multi-element determination was conducted using a simple quadrupole-ICP-MS 7850 from
146 Agilent Technologies, equipped with an integrated autosampler SPS 4. A micromist nebuliser was
147 used for all measurements. The collision cell was set to helium mode for all elements, at a flow rate
148 of 5 mL/min, to minimise polyatomic interferences. The operating conditions were as follows:
149 1550 W forward power, 15 L/min plasma gas flow, 1 L/min carrier gas flow and 1 L/min auxiliary

150 gas flow. The remaining parameters were adjusted daily using a tuning solution to optimise the
151 signal.

152 Elemental concentrations were obtained through SQ analysis using the 28-element standard at a
153 concentration of 20 µg/L and 1% HNO₃ as the blank. SQ approach was performed for 42 elements,
154 with 100 sweeps and one replicate: ¹¹B, ²³Na, ²⁴Mg, ²⁷Al, ²⁸Si, ³¹P, ³⁴S, ³⁵Cl, ³⁹K, ⁴³Ca, ⁴⁵Sc, ⁴⁷Ti,
155 ⁵¹V, ⁵²Cr, ⁵⁵Mn, ⁵⁶Fe, ⁵⁹Co, ⁶⁰Ni, ⁶³Cu, ⁶⁶Zn, ⁷⁵As, ⁷⁹Br, ⁸⁵Rb, ⁸⁸Sr, ⁸⁹Y, ⁹⁰Zr, ⁹³Nb, ¹¹¹Cd, ¹¹⁵In, ¹¹⁸Sn,
156 ¹²⁷I, ¹³³Cs, ¹³⁷Ba, ¹³⁹La, ¹⁴⁰Ce, ¹⁴¹Pr, ¹⁴⁶Nd, ¹⁴⁷Sm, ¹⁸²W, ²⁰⁵Tl, ²⁰⁸Pb and ²³⁸U. These 41 elements,
157 except for In used as internal standard, constitute the mineral wine profile (MWP). The following
158 elements were absent from the calibration standard: ²⁸Si, ³¹P, ³⁴S, ³⁵Cl, ⁴⁵Sc, ⁷⁹Br, ⁸⁵Rb, ⁸⁹Y, ⁹⁰Zr,
159 ⁹³Nb, ¹¹⁵In, ¹¹⁸Sn, ¹²⁷I, ¹³³Cs, ¹³⁹La, ¹⁴⁰Ce, ¹⁴¹Pr, ¹⁴⁶Nd, ¹⁴⁷Sm, ¹⁸²W and ²³⁸U. Their concentrations
160 are interpolated between elements present in the calibration standard, applying response factors,
161 which depend on their isotopic mass, isotopic abundance and ionisation energy.

162 The analytical procedure is summarised in Figure S1. The determined concentrations then served as
163 input for the machine learning algorithms.

164 4. Statistical analysis

165 Statistical analyses were performed with the libraries *scipy.stats*, (Virtanen, 2020) and *scikit-learn*
166 (Pedregosa, 2012) compatible to Python version 3.9.19.

167 Values below the limit of quantification, determined by the Agilent MassHunter 5.2 software version
168 D.01.02 during each analysis, were imputed as 10⁻⁴ (ppb). Elemental concentration results were
169 summarised using mean, median, and interquartile range (IQR). Histograms of the Box-Cox
170 transformed data were calculated.

171 The raw database was normalised via Z-score transformation and samples, taking out the under-
172 represented labels in colour and the category. Spearman correlation coefficients were computed and
173 correlations were considered significant when *p*-value < 0.05. Cluster analysis was conducted using
174 Ward's method and Euclidean distance.

175 An exploratory analysis employing Principal Component Analysis (PCA) was carried out to identify
176 underlying patterns in the dataset, and its first 10 principal components were visualised using the t-
177 Stochastic Neighbour Embedding (t-SNE) technique. t-SNE is utilised to reduce data dimensionality
178 to two dimensions, preserving both local and global structures, thus facilitating cluster visualisation
179 (van der Maaten, 2008).

180 **5. Sample classification**

181 **5.1. Selection of machine learning model**

182 For model selection, the dataset underwent 80:20 stratified random split, to compose the train and
183 the test set, respectively. The test set was used to verify the trained model performance, as it had not
184 been previously seen by the model. All samples with unknown label values were taken out of the
185 dataset before the stratified split was done. Both sets were composed of all 39 elements (B, Na, Mg,
186 Al, P, S, Cl, K, Ca, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Br, Rb, Sr, Y, Zr, Nb, Cd, Sn, I, Cs, Ba,
187 La, Ce, Pr, Nd, Sm, W, Tl, Pb and U).

188 Six machine learning models were benchmarked: Random Forest, k-nearest neighbours (k-NN),
189 support vector machine (SVM), Logistic Regression implemented in *scikit-learn*, extreme gradient
190 boosting (XGB) implemented in the *XGBoost* library (Chen, 2016) and an artificial neural network
191 model (ANN) created using the *TensorFlow* library (Abadi, 2015). No optimisation of the models'
192 functions was done for their comparison. The metric chosen for their comparison was the area under
193 the receiver operating characteristic (ROC) curve. This curve is a graphic representation of the trade-
194 off between specificity and sensitivity of a model (Fawcett, 2006). The area under this curve (AUC)
195 is an important metric when evaluating a model, it represents the probability that a classifier will
196 correctly rank a randomly selected positive instance higher than a randomly chosen negative
197 instance. An AUC of 0.5 reflects random guessing, while a perfect classifier achieves an AUC of 1.0
198 (Fawcett, 2006).

199 The chosen classes were country, French region and principal grape variety. Only labels with more
200 than 50 samples were classified. The model was trained and tested for 10 iterations. The AUC score
201 was computed for each iteration and its mean was then calculated.

202 **5.2. Application of XGB in Sample Classification**

203 The XGB was found to be the best performing technique (as explained the Results and Discussion
204 section). XGB is a boosting ensemble learning algorithm, which uses many decision trees whose
205 predictions are combined in order to obtain the final classification (Chen, 2016). This machine
206 learning method has been applied in several domains, such as disease and stock prediction (Chen,
207 2016; Ma, 2021), with very good results when distinguishing the geographical origin of food
208 products (Kang, 2023; Wen, 2023).

209 In order to improve the model predictions after selection, a grid search was employed to optimise
210 the model's parameters: number of estimators were set to 500, maximum tree depth to five, learning
211 rate to 0.1, gamma to zero and the regularisation parameter lambda to one. The model parameter
212 "objective" was *binary:logistic*. The metrics chosen to evaluate the classifier performance were

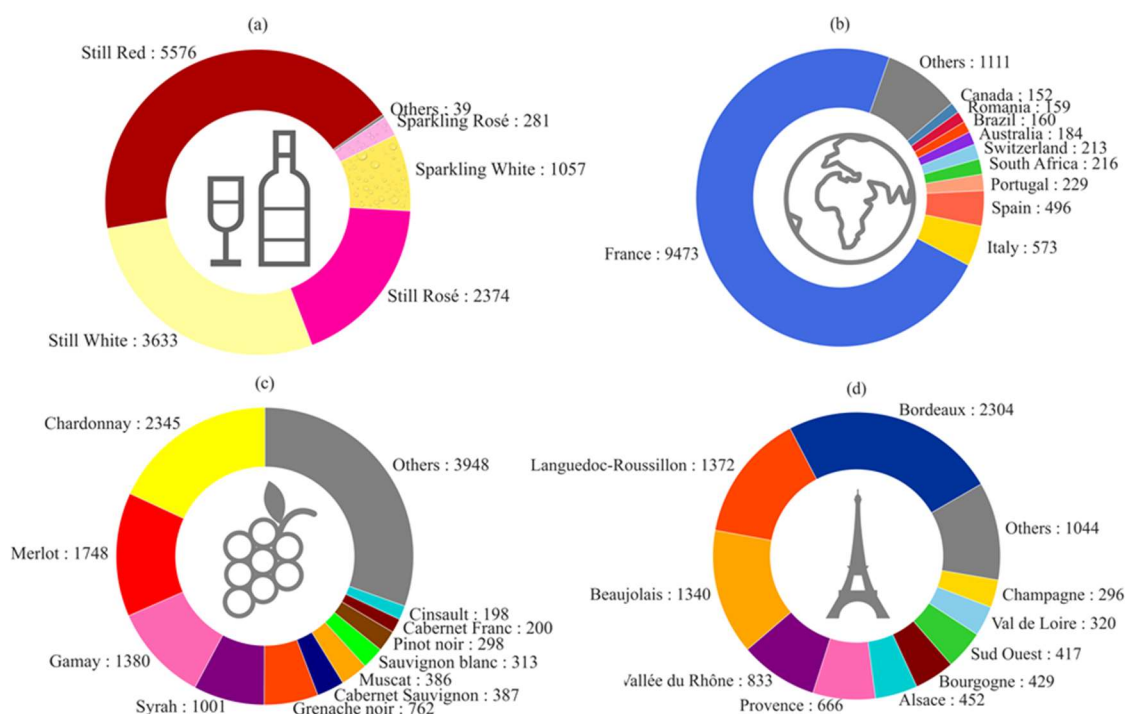
213 sensitivity, specificity, accuracy and the AUC. The first metric was the probability of a positive
 214 individual being correctly classified as positive and the second the probability of a negative
 215 individual being correctly classified as negative. Accuracy was the ratio of correctly classified
 216 samples to the total number of samples present in the evaluation dataset (Hicks, 2022).

217 The same classes and labels were classified in the optimised XGB model. It was repeated 10 times
 218 and the means of the metrics were calculated.

219 Results and discussion

220 1. Determining the elemental composition of the wine samples

221 For this study, the MWP of 12966 wines of commercial origin as well as from international
 222 competitions were obtained. The wines originated from a wide variety of countries (more than 45),
 223 with France being the most represented (9473 wines), as well as several regions (more than 200) and
 224 grape varieties (more than 200). Figure 2 illustrates the distribution of the analysed wines in the
 225 database. A more detailed description is given in Tables S1, S2 and S3.



227 **Figure 2. Distribution of the 12966 analysed wines based on the wine type (a), country (b),**
228 **grape variety (c), and French wine region (d). The category "Others" contains labels with**
229 **fewer than 150 samples. A more detailed description of each category is given in Tables S1, S2**
230 **and S3.**

231 Semi-quantitative analysis is an interesting alternative to full-quantitative analysis and is particularly
232 valuable for rapidly screening the elemental composition of samples, since it enables fast
233 determination of the approximate elemental composition in unknown samples. In semi-quantitative
234 mode, the entire mass range is scanned, thereby recording a signal for every possible element or
235 isotope. Moreover, SQ analysis involves estimating the relative concentrations of elements without
236 relying on the multi-standard calibration necessary for quantitative analysis. It also eliminates the
237 need for multiple calibration curves due to incompatibility of the simultaneous presence of certain
238 elements (i.e., Zr in the presence of an excess of Na). As a consequence, this mode has high
239 economical advantages both in time and reagents.

240 The accuracy was not determined, but several papers have reported high accuracy (bias < 20%)
241 (Catarino, 2006; Chen, 2008). While it is associated with lower accuracy compared to quantitative
242 analysis, this method is suitable for creating a comprehensive database of mineral profiles that can
243 be exploited in multivariate data analysis and machine learning algorithms. Indeed, the aim of the
244 analysis is not to ascertain the concentrations of the different elements present in wine but to provide
245 values that can be considered characteristic of a given wine. Reproducibility was determined via the
246 analysis of the control wine over a 15 day-period and was found to not exceed 15% for all the
247 elements.

248 The developed SQ method enabled analysis of more than 200 samples per day, with each sample
249 being analysed for seven minutes. To prevent any potential signal variation over time, several control
250 points were implemented: analysis of the control wine at the start, midpoint and end of each
251 sequence; analysis of blanks and 28-element standard every 40 analyses; monitoring of indium
252 concentration in each sample. It was used to determine the MWP of the wines of this study.

253 Comprehensive statistical summaries, including the mean, median and interquartile range (IQR) of
254 mineral element concentrations, are provided in Table 1. The concentration distributions across the
255 database using Box-cox transformation are depicted in Figure 3. Lambda values are given in Table
256 S4. Visual inspection of the histograms suggests multimodal distributions of data. Consequently, the
257 Spearman correlation coefficient was adopted to evaluate the relationships between variables, given
258 its non-parametric nature as a measure of rank correlation.

259 **Table 1. Mean, median and interquartile range (IQR) for the 39 concentrations of mineral**
 260 **elements measured by ICP-MS.**

	Mean (ppb)	Median (ppb)	IQR (ppb)
B	5075	4692	2768
Na	2542.10 ¹	2040.10 ¹	1937.10 ¹
Mg	1090.10 ²	1025.10 ²	4234.10 ¹
Al	646.8	504.6	476.6
P	5325.10 ²	4889.10 ²	2893.10 ²
S	1586.10 ²	1450.10 ²	7804.10 ¹
Cl	3102.10 ¹	2286.10 ¹	1968.10 ¹
K	1017.10 ³	9814.10 ²	5655.10 ²
Ca	7202.10 ¹	6784.10 ¹	2730.10 ¹
Ti	21.62	12.55	14.03
V	35.84	2.99	19.77
Cr	18.88	15.33	11.77
Mn	2458	1420	2323
Fe	2236	1675	2232
Co	5.30	4.08	3.47
Ni	34.86	27.44	23.97
Cu	142.6	69.66	123.1
Zn	954	856.8	558.4
As	5.57	3.49	4.69

Br	287.88	223.4	239.5
Rb	1665	1439	1159
Sr	516.3	404.4	364
Y	0.73	0.29	0.67
Zr	5.17	1.87	4.57
Nb	0.43	0.14	0.38
Cd	0.32	0.24	0.23
Sn	1.74	0.97	1.32
I	4.09	3.41	3.44
Cs	10.43	4.13	5.53
Ba	279.3	164	263.6
La	0.59	0.12	0.42
Ce	1.16	0.27	0.81
Pr	0.12	0.03	0.10
Nd	0.51	0.12	0.44
Sm	0.10	0.00	0.09
W	0.73	0.25	0.56
Tl	0.37	0.27	0.28
Pb	14.02	9.88	9.48
U	0.46	0.16	0.37

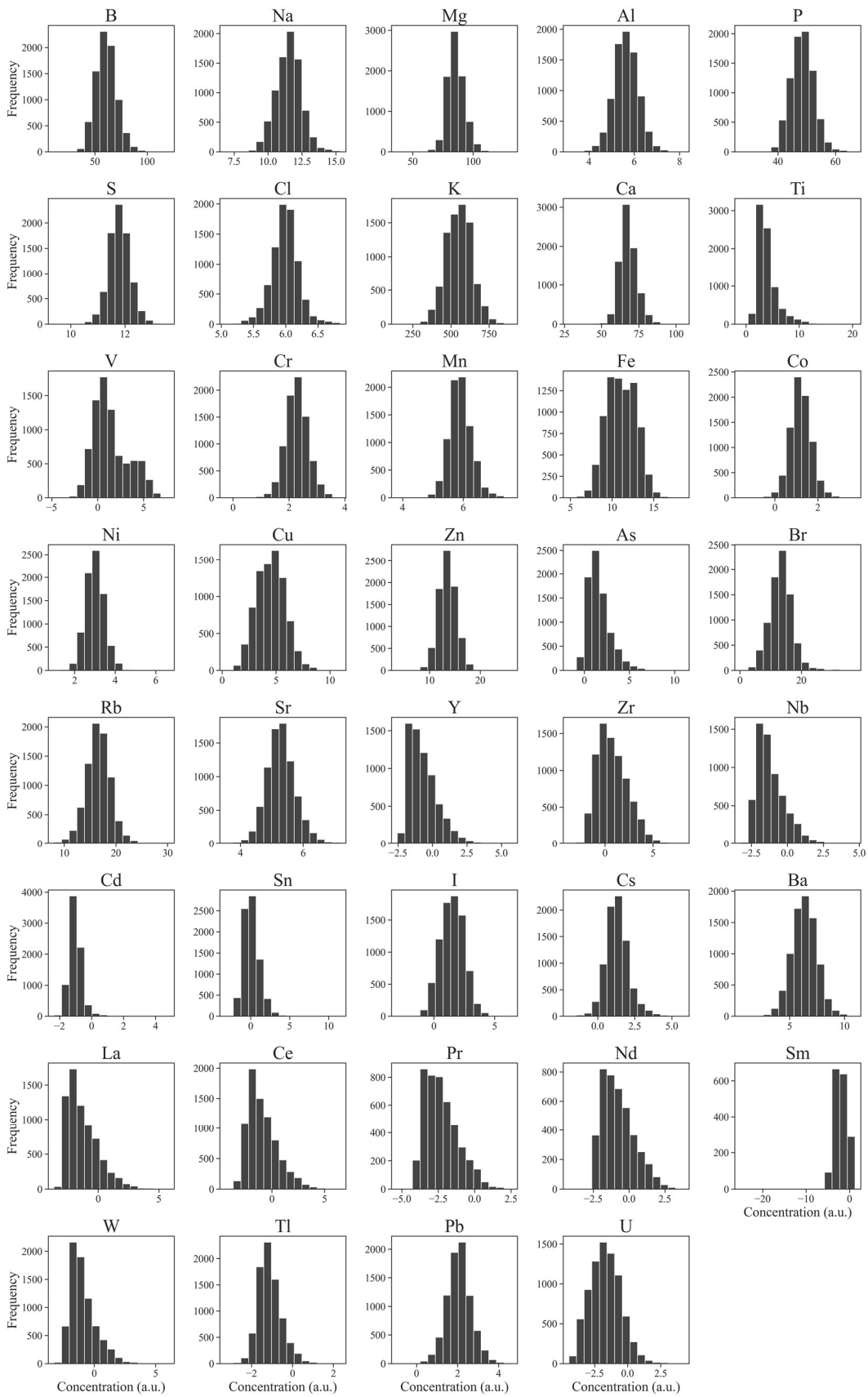
261

262 Figure 3 illustrates the noticeable dispersion of elemental content in wine. This phenomenon is
263 documented in existing literature across various contexts, including comparisons between wines

264 from different countries (Bentlin, 2011; Griboff, 2021), within the same country (Kment, 2005),
265 across different types (Griboff, 2021), and even when examining different vintages from the same
266 vineyard (Tanabe, 2020). This variability in elemental content underscores the intricate interplay of
267 factors shaping wine composition, such as geographical origin, vinification techniques and
268 environmental influences.

269 Given this complexity, establishing a comprehensive wine database that accounts for these diverse
270 factors is indispensable for advancing our understanding of the relationships between elements and
271 the broader context of wine production. In the subsequent sections, we delve more deeply into the
272 examination of elemental correlations and their potential applications in wine traceability.

273



275 **Figure 3. Element concentration distribution in the database. Concentration (x-axis) is**
276 **transformed using the Box-Cox power transformation. If the concentration value was under**
277 **the limit of quantification, the data was not plotted. The concentration and frequency scales**
278 **have been adapted for each element. The Lambda values for each element are given in Table**
279 **S4.**

280 **2. Statistical analysis**

281 **2.1. Correlation analysis**

282 To investigate potential relationships between element concentrations, a Spearman correlation test
283 was conducted at a 95% confidence level. The correlation coefficients (ρ) are depicted in the
284 correlation matrix plot, in Figure 4a. Positive (red shading) and negative (blue shading) correlations
285 were found amid the pairs of elements, with the most significant being positive correlations between
286 rare earth elements, Y, La, Ce, Pr, Nd and Sm (ρ varying from 0.93 to 0.79) and Zr-Y (ρ 0.71).

287 Rare earth elements are a group with similar chemical behaviour, thus their correlation is expected
288 and has been reported elsewhere (Alonso Gonzalez, 2021; Pasvanka, 2021). They are associated with
289 soil content and bentonite treatment (Catarino, 2008; Pohl, 2007). The correlation between Zr and Y
290 may also be explained by this clarification agent, as well as by the use of Yttria-stabilised zirconia
291 for wine stabilisation and/or filtration (Catarino, 2008; Salazar, 2006; Silva-Barbieri, 2022).

292

294 **Figure 4. (a) Correlation matrix plot illustrating the relationships between 39 pairs of element**
295 **concentrations across all samples of the database. The colour gradient reflects the Spearman**
296 **correlation coefficient (ρ), with statistically significant correlations (p -value < 0.05) indicated**
297 **in red (positive correlation) or blue (negative correlation). Non-significant correlations are in**
298 **white cells. (b) Hierarchical cluster dendrogram generated using Ward's method and**
299 **Euclidean distance for the 39 elements. Distances indicate the degree of correlation between**
300 **different elements. Three clusters are identifiable: C1 (Ba, Mn, Co, Fe, Cr, Zn, Pb, Cu, Ni, Cd,**
301 **Sn, Tl, Rb and Cs), C2 (K, Mg, P, I, Ca, S, Sr, B, Br, Cl and Na) and C3 (Zr, Al, U, As, V, W,**
302 **Nb, Ti, Ce, La, Nd, Pr, Sm and Y).**

303

304 **2.2. Cluster analysis**

305 To further elucidate metal concentration relationships, a cluster analysis was performed utilising the
306 Ward method and Euclidean distance. The resulting dendrogram, depicted in Figure 4b, reveals three
307 distinct clusters which are in agreement with the correlations found in Figure 4a. The first cluster
308 (C1) comprises Ba, Mn, Co, Fe, Cr, Zn, Pb, Cu, Ni, Cd, Sn, Tl, Rb and Cs, which are predominantly
309 micronutrients. These elements are essential to plant health and some (e.g., iron and copper) also
310 impact the colour and oxidative stability of wine (Pohl, 2007).

311 The second cluster (C2) comprises plant macronutrients K, Mg, P, Ca and S, which have major roles
312 in plant metabolism. Additionally, this cluster includes supplementary elements, such as I, Sr, B, Br,
313 Cl and Na, whose presence in wine warrants further exploration. The third cluster (C3) encompasses
314 rare earth elements alongside Zr, Al, U, As, V, W, Nb and Ti, their presence in wine originating from
315 various sources, including soil composition and winemaking techniques (Catarino, 2008; Pohl,
316 2007), as illustrated in Figure 1.

317

318 **2.3. Principal Component Analysis**

319 Building on the insights gleaned from the correlation coefficients and cluster analysis, principal
320 component analysis (PCA) was subsequently employed to further explore the intricate relationships
321 within the dataset. Due to its high dimensionality, visualisation was not possible. Therefore, a
322 dimensionality reduction technique t-SNE was applied to the first ten principal components, which
323 accounted for 70% of total variance.

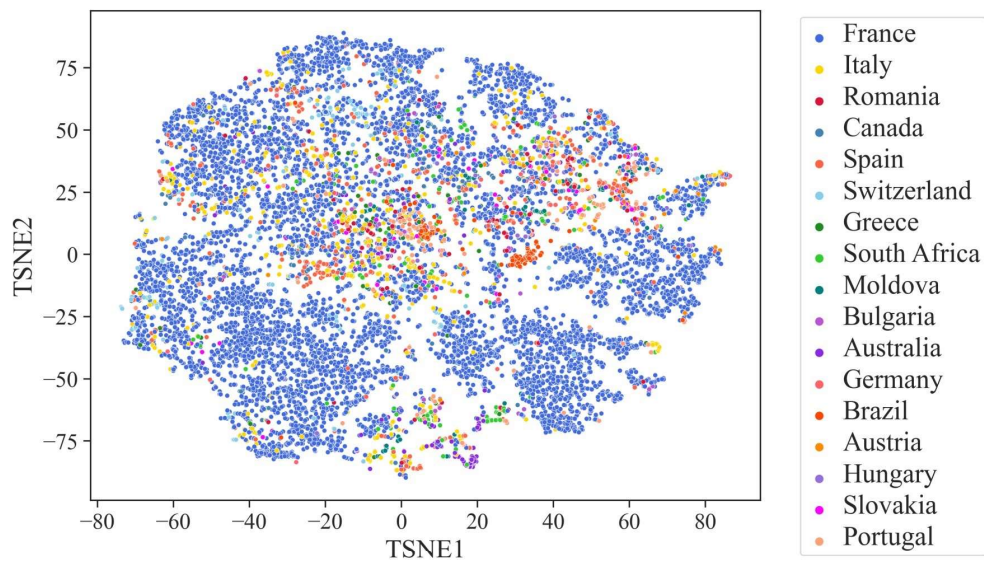
324 Initially, the grouping of all the samples was studied, as shown in Figure 5 and Figure S2. In terms
325 of wine type, red wines are well separated from rosé and white wines, which is expected due to the

326 differences in winemaking processes. This is consistent with other studies (M. Gajek, 2021), which
327 have also shown – albeit using a limited number of samples - differences in mineral wine profiles
328 for red, rosé and white wines. For the distribution of data based on the country of origin, French
329 wines constitute the majority and are represented in the large (blue) cluster. Smaller groups of
330 Spanish and Italian wines are visible. Lastly, grape variety grouping was also studied, for which the
331 samples containing unknown varieties were taken out in order to better illustrate the existing clusters.
332 These grape variety clusterings were attributed to the interaction of two different factors: the unique
333 composition of each variety and its region of origin. Therefore, the assessment of wine region clusters
334 was carried out for French wines, as most of their regions are well represented in the database, which
335 resulted in the t-SNE representations in Figure c and Figure S2b.

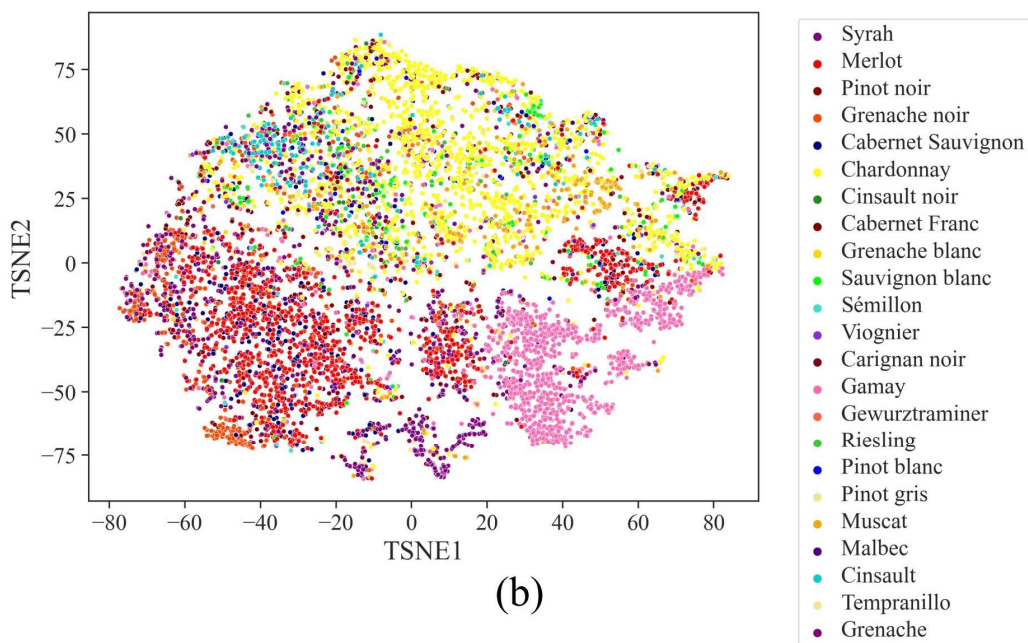
336 For French wine type, the separation of red wines from white and rosés is clear. For the French wine
337 regions and principal grape variety, all the categories with more than 50 samples have been
338 represented and only French samples with known categories have been plotted. Clusters are evident
339 for the Beaujolais, Vallée du Rhône, Bordeaux and Champagne regions. The separation of the former
340 three regions may be due to the typicality of their soils, as well as the typical varieties that are used in
341 the production of the respective wines. This is supported by the separation of the principal grape
342 varieties, as the clusters of Gamay, Grenache noir and Merlot correspond to those of Beaujolais,
343 Vallée du Rhône and Bordeaux, respectively.

344 For the Champagne region separation there is an additional explanation to that involving principal
345 variety (Chardonnay) and soil. Most of the sparkling wines in the dataset come from this region, thus
346 the chemical difference between a white still wine and a white sparkling wine may artificially play
347 a role in this difference.

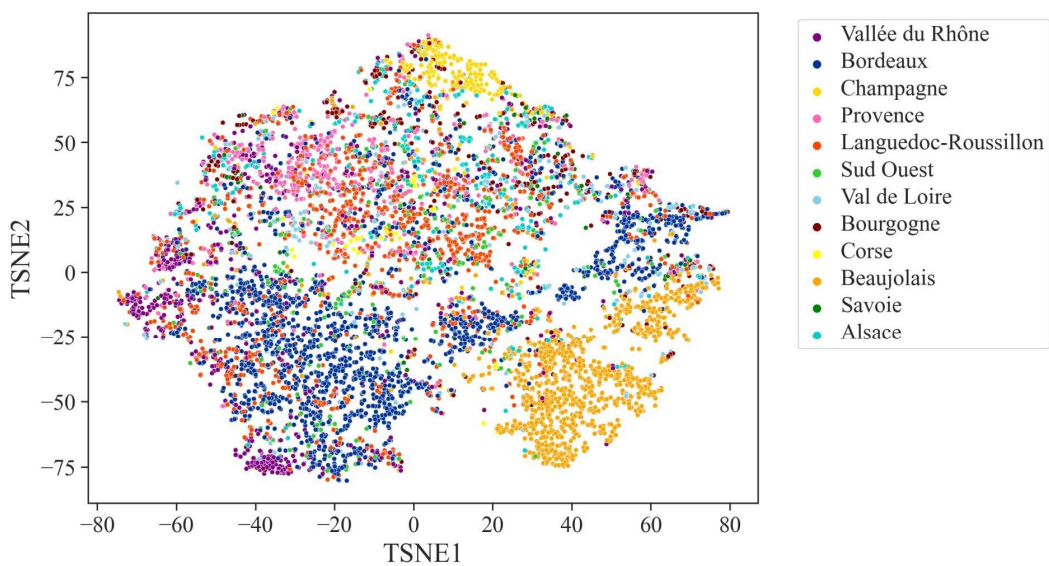
348 The promising patterns observed in both the data grouping and a rich database containing more than
349 12000 samples have motivated the development of machine learning techniques tailored to sample
350 classification.



(a)



(b)



(c)

352 **Figure 5. (a-c) t-SNE representation of international MWP samples according to country (a),**
353 **international main grape variety (b) and French region (c) of the 10 first principal components**
354 **(67% of total variance). Only samples with known labels are represented in the images. For**
355 **complementary images (type and French varieties), see Figure S2. Figure 5c and Figure S2b**
356 **have similarities, showing that the typicality of the wine producing region is related to the grape**
357 **variety. A clear separation for red and white/rosé is illustrated for international samples in**
358 **Figure S2a.**

359 3. Sample classification

360 3.1. Selecting the machine learning model

361 Various supervised learning algorithms have been employed in the literature to determine and
362 differentiate the origin of wines. These include the support vector machine (Astray, 2021; Da Costa,
363 2020), stepwise linear discriminant analyses (Pérez-Magariño, 2004), random forest (Astray, 2021;
364 Da Costa, 2020) and artificial neural networks (Astray, 2021; Da Costa, 2020; Pérez-Magariño,
365 2004; Wu, 2021). Renaweera *et al.* were also able to identify blending percentages using
366 spectrofluorimetric analysis with another machine learning algorithm, the extreme gradient boosting
367 discriminant analysis (Ranaweera, 2022).

368 Given the extensive array of machine learning techniques prevalent in the literature, an initial
369 performance assessment was conducted to determine the optimal model for sample classification.
370 The models were trained and tested ten times, using 80:20 random stratified split, and the mean AUC
371 score was computed after classification of test samples. The results are given in Table 2. The model
372 with the best performance was Extreme Gradient Boosting (XGB), and was thus chosen to be
373 developed in this study.

374 **Table 2. Mean AUC comparison for the six machine learning models tested in the classification**
375 **of wine origin and grape variety. The highest score achieved for each class is highlighted in**
376 **bold**

Model	Mean AUC		
	Country	French Wine Region	Grape Variety
Random Forest	0.952	0.953	0.872
k-NN	0.836	0.871	0.759
SVM	0.964	0.946	0.893
Logistic Regression	0.939	0.913	0.875
Extreme Gradient Boosting	0.977	0.967	0.919
ANN	0.925	0.897	0.851

377 **3.2.Application of XGB in Sample Classification**

378 The performance metrics of the classifier for countries, French region and principal grape varieties
379 of the wines are given in Table 3. When assessing the classifier performance in terms of country
380 prediction, it is evident that the models can accurately predict a wine's country. The AUC surpasses
381 0.9 across all of the countries, indicating the models' high reliability when distinguishing between
382 different samples. Furthermore, the accuracy metrics show remarkable values, with at least 83%
383 correctly classified samples across all countries. Particularly noteworthy are the results for Brazil
384 and Australia, with accuracy levels exceeding 96%.

385 French wines constitute the predominant country; this facilitates the comprehensive coverage of
386 various French wine regions within the database, which enables the development of classifiers to
387 predict the origin of these wines within their respective territory. The developed models showed high
388 predictive ability, with their AUCs varying from 0.906 to 0.996. Their best performance, as
389 illustrated by the AUC, was for Bordeaux, Beaujolais and Champagne regions. This is in agreement
390 with their natural separation in the set, as presented in Table 3 and led to accurate predictions, varying
391 from 95.2% to 97.8%.

392 These results are promising as they indicate that the MWP is a robust tool for verifying the origin of
393 a wine. Further avenues of research include the exploration of its use when tracing a wine to sub-
394 regional level, thereby providing insights into the unique terroir characteristics within a larger wine-
395 producing region. Additionally, this tool can be applied in further research to explore the differences
396 between the mineral signatures of wine-producing regions, as it translates not only the fingerprint of
397 the soil but also viticultural practices and winemaking techniques, as illustrated in Figure 1.

398 When distinguishing grape varieties, XGB demonstrates reliability when distinguishing the principal
399 wine varieties, with an AUC exceeding 0.8 for all of the labels. The classification of the principal
400 wine varieties proves more complex than country or region due to the prevalence of multivarietal
401 wines in certain regions, as well as the use of rootstocks. Moreover, the International Organisation
402 of Vine and Wine's labelling rules (International Organisation of Vine and Wine, 2024) do not
403 require varietal names and their percentages to be mentioned, which increases the difficulty of the
404 classification.

405 These factors may explain the separation shown in Table 3 with clustering remarkable only to the
406 Gamay variety. However, the model showed high overall performance for the classification of
407 principal wine varieties with accuracy ranging from 73.7% to 98.0%. Two possible avenues of
408 improvement are possible: training the model exclusively on monovarietal wines and enriching the
409 database with the rootstock used for each variety in a wine. Even though this is a possibility, the

410 models performed remarkably well without these options, showing their potential for grape variety
411 classification.

412 **Table 3. Mean ($N_{iteration} = 10$) performance metrics for predicting a wine’s country, French region and principal grape variety. Only labels with**
 413 **more than 50 samples were classified**

Country	Samples	AUC	Sensitivity	Specificity	Accuracy	Country	Samples	AUC	Sensitivity	Specificity	Accuracy
France	9454	0.981	0.938	0.932	0.936	Canada	152	0.985	0.937	0.959	0.958
Italy	568	0.968	0.914	0.903	0.904	Moldova	103	0.988	0.943	0.943	0.943
Spain	495	0.962	0.885	0.916	0.915	Greece	92	0.945	0.872	0.873	0.873
Portugal	228	0.968	0.893	0.928	0.928	Hungary	90	0.954	0.878	0.895	0.895
South Africa	216	0.983	0.916	0.947	0.946	Bulgaria	83	0.929	0.724	0.949	0.948
Switzerland	213	0.985	0.930	0.957	0.957	Austria	82	0.943	0.881	0.834	0.834
Australia	184	0.992	0.957	0.962	0.962	Slovakia	71	0.959	0.900	0.872	0.872
Brazil	160	0.996	0.950	0.982	0.982	Germany	68	0.907	0.707	0.899	0.898
Romania	158	0.948	0.863	0.925	0.924						
French region	Samples	AUC	Sensitivity	Specificity	Accuracy	French region	Samples	AUC	Sensitivity	Specificity	Accuracy
Bordeaux	2303	0.987	0.950	0.952	0.952	Bourgogne	429	0.959	0.891	0.908	0.907
Languedoc-	1372	0.957	0.895	0.895	0.895	Sud-Ouest	412	0.906	0.806	0.852	0.850

Roussillon											
Beaujolais	1340	0.996	0.969	0.980	0.978	Vallée de la Loire	318	0.930	0.864	0.845	0.846
Vallée du Rhône	833	0.946	0.880	0.872	0.873	Champagne	295	0.982	0.934	0.972	0.971
Provence	665	0.957	0.899	0.890	0.891	Savoie	69	0.968	0.871	0.936	0.936
Alsace	447	0.980	0.921	0.949	0.947	Corse	62	0.964	0.858	0.907	0.907
Principal grape variety	Samples	AUC	Sensitivity	Specificity	Accuracy	Principal grape variety	Samples	AUC	Sensitivity	Specificity	Accuracy
Chardonnay	2344	0.967	0.908	0.893	0.896	Riesling	98	0.932	0.835	0.878	0.877
Merlot	1747	0.960	0.903	0.918	0.916	Malbec	94	0.847	0.716	0.833	0.832
Gamay	1379	0.992	0.956	0.983	0.980	Tempranillo	89	0.941	0.828	0.926	0.925
Syrah	1000	0.934	0.871	0.863	0.864	Pinot gris	83	0.910	0.812	0.830	0.830
Grenache noir	762	0.952	0.911	0.866	0.867	Viognier	81	0.874	0.819	0.785	0.786
Cabernet Sauvignon	422	0.828	0.777	0.743	0.744	Grenache blanc	81	0.851	0.769	0.780	0.780

Muscat	384	0.959	0.892	0.883	0.883	Sémillon	75	0.918	0.840	0.822	0.822
Sauvignon blanc	311	0.938	0.868	0.869	0.869	Gewurztramine r	71	0.952	0.907	0.871	0.871
Pinot noir	297	0.860	0.785	0.801	0.800	Cinsault noir	65	0.807	0.731	0.737	0.737
Cabernet Franc	199	0.861	0.823	0.763	0.764	Carignan noir	62	0.899	0.767	0.882	0.882
Cinsault	198	0.959	0.905	0.891	0.891	Pinot blanc	51	0.932	0.830	0.861	0.861
Grenache	139	0.952	0.911	0.866	0.867						

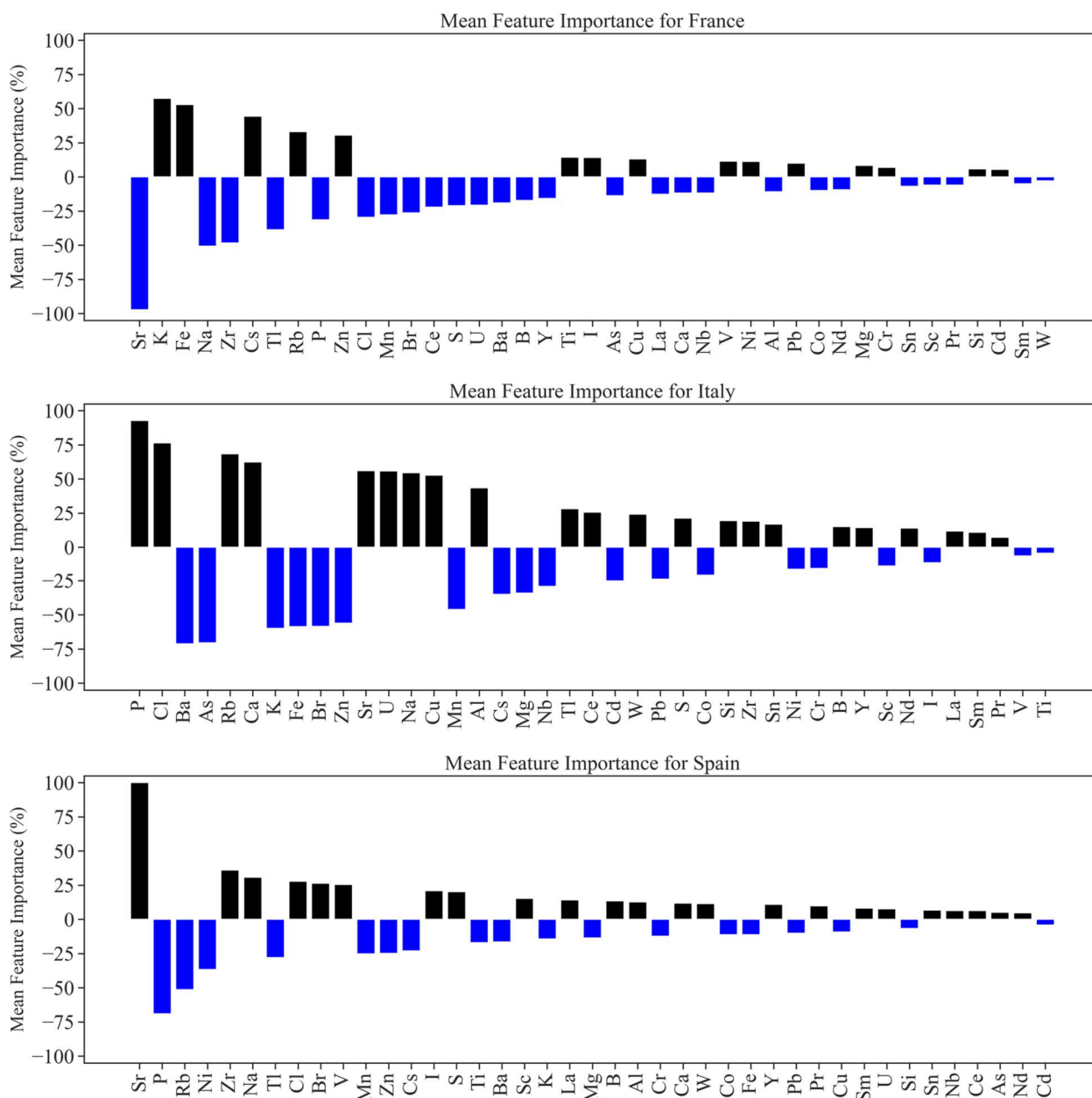
415 In order to further adapt the model to wine authentication, the type of wine can be taken into account
 416 when segmenting the dataset. To this end, red wines were used to carry out a binary classification of
 417 French, Italian and Spanish wines and thus evaluate how the metrics would change. The results are
 418 presented in Table 4. It can be seen that model performance for all three countries displays an AUC
 419 higher than 0.9. The same behaviour was found in terms of the model's accuracy, which reached at
 420 least 94% of correct classifications for France and Spain. This performance warrants further
 421 exploitation of the data, and it reinforces the need for a polyvalent database that can be refined as
 422 required.

423 **Table 4. Mean ($N_{\text{iteration}} = 10$) performance metrics when classifying the countries of French,**
 424 **Italian and Spanish red wines.**

Country	Number of samples	AUC	Sensitivity	Specificity	Accuracy
France	4799	0.989	0.947	0.960	0.949
Italy	131	0.965	0.908	0.903	0.903
Spain	109	0.980	0.927	0.948	0.947

425 The differentiation between the previously classified categories relies exclusively on the features
 426 available within the database, consisting of 39 elements in the MWP. When differentiating
 427 distinguishing between these categories, their significance is evaluated through the mean importance
 428 feature, as depicted in Figure 6 for the three countries. This metric has positive and negative values,
 429 depending on how the presence (positive, black shading) or absence (negative, blue shading) of a
 430 feature helps in distinguishing each category. In the differentiation of French wines, the absence of
 431 strontium is the most important feature that differentiates them from all the other wines, while for
 432 Spanish, the presence of this element helps in their classification.

433 Fifteen elements contribute to the differentiation of the wines originating from Italy. The variation
 434 in the features influencing the model highlights the necessity of conducting a comprehensive MWP
 435 assessment of each wine, as different categories can be differentiated by various features; this is
 436 further illustrated for the three main regions and grape varieties in Figures S3 and S4.



437

438 **Figure 6. Element Mean Feature Importance for the three major countries. Values are**
 439 **presented in decreasing order of importance. Values in black indicate positive correlations and**
 440 **values in blue, negative correlations between the element's concentration and the wine**
 441 **category. These values for the three main regions and three main grape varieties are shown in**
 442 **Figures S2 and S3**

443 The segmentation of the dataset previously carried out for the country classifications is not the only
 444 strategy that can be used to adapt the model to sample classification. The decision threshold of the
 445 model can be tuned in order to achieve higher values of specificity or sensitivity, which could be
 446 used to better determine whether the sample is from the target category or not. A specificity of over
 447 99% can be thereby be obtained, increasing the reliability of the classification as non-belonging to a

448 category. This optimisation was conducted for the three major countries, French regions and grape
 449 variety, producing the metrics presented in Table 5.

450 **Table 5. Mean ($N_{\text{iteration}} = 10$) performance metrics for classifying the three major categories of**
 451 **country, French region and grape variety. Specificity is set to 0.99.**

Country	Number of samples	Specificity	Sensitivity	Accuracy
France	9454	0.991	0.747	0.813
Italy	568	0.991	0.617	0.974
Spain	495	0.991	0.607	0.976
French region	Number of samples	Specificity	Sensitivity	Accuracy
Bordeaux	2303	0.991	0.734	0.922
Languedoc-Roussillon	1370	0.990	0.464	0.906
Beaujolais	1340	0.991	0.935	0.982
Principal grape variety	Number of samples	Specificity	Sensitivity	Accuracy
Chardonnay	2344	0.991	0.578	0.905
Merlot	1747	0.990	0.386	0.897
Gamay	1379	0.991	0.938	0.985

452 When comparing the models described in this study with others in the literature, similar
 453 performances were found. Tanabe *et al.* (2020) analysed 62 elements to differentiate neighbouring
 454 American viticultural regions, with an accuracy of over 94%. In terms of grape variety, their study
 455 was limited to Pinot noir with a limited number of samples ($n=53$) (Tanabe, 2020). As a comparison,
 456 the model developed in this study obtained a region classification accuracy of up to 98% using more
 457 samples and a more diverse dataset.

458 Griboff *et al.* analysed 18 elements by ICP-MS and 2 isotopes by isotope ratio mass spectrometry of
 459 62 wine samples from Argentina and Australia (Griboff, 2021). As already explained in the
 460 Introduction, sample preparation for isotope ratio analysis is time-consuming and hinders the
 461 acquisition of data for a large and comprehensive dataset. The method developed in this study

462 provides a more time-efficient analysis, as well as a more comprehensive database to be exploited
463 via machine learning methods.

464 Forina *et al.* analysed a dataset extracted from the European Wine Databank. It was composed of 58
465 selected organic and inorganic analytical parameters of 1188 wine samples that were available in the
466 databank (Forina, 2009). This was the most comprehensive study found in the literature, but it was
467 still limited to four countries and the methods of data acquisition were costly and time-consuming.

468 When distinguishing varieties, other studies have explored the elemental content of wine as a
469 fingerprint (Feher, 2019; Temerdashev, 2019), using chemometric or machine learning approaches.
470 This study achieved comparable results with a larger and more origin-diverse dataset. As this profile
471 is usually only associated with a wine's origin, being able to differentiate varieties in a multiple-
472 origin set is promising for the future of wine authentication. Recently Temerdashev *et al.* (2024)
473 have shown that, using chemometric analysis and 153 samples, it is possible to distinguish between
474 three grape varieties (Chardonnay, Riesling and Muscat) and four regions of the Krasnodar territory:
475 this therefore also validates our ICP/MS mineral analysis methodology for classifying wines.

476 While previous studies have obtained good results for wine classification, no other existing research
477 has used the same number of samples and representation of countries, wine regions and varieties as
478 in the present study. Such a large database is essential for creating a polyvalent model that can verify
479 the origin of an unknown wine by exploiting exclusively its Mineral Wine Profile.

480 To the best of our knowledge, this study is the first that involves the analysis of over twelve thousand
481 wine samples and their corresponding MWP. The extensive dataset opens up numerous avenues for
482 further research. For example, MWP could be used as a tool for studying various ecological
483 phenomena over time and to support necessary adaptations to climate change and modifications in
484 viticultural practices. For instance, elements like potassium are already closely monitored by
485 winegrowers, as potassium nutrition is directly correlated with grapevine growth and ultimately with
486 wine quality (Villette, 2020). Interestingly, potassium levels in berries have been steadily increasing
487 over the past few decades and serve as a reliable indicator of climate change, which is linked to a
488 decline in wine quality (Nistor, 2022). Similarly, an increase in calcium levels in wine has been
489 observed, attributed to global warming-induced water stress in plants, which is also linked to changes
490 in wine quality (Fioschi, 2024). In addition to these well-known minerals associated with global
491 warming, MWP, when integrated with large datasets, may be used in the future to identify new
492 indicators related to subtle climate changes in specific regions.

493 **Conclusion**

494 The findings of this study demonstrate the remarkable capabilities of MWP in determining the
495 country and region of wine production. It is noteworthy that contemporary consumers increasingly
496 seek detailed information regarding authenticity that goes beyond just the region of origin. The
497 concept of terroir, ranging from MACRO-terroir to MICRO-terroir via MESO-terroir (Marre *et al.*,
498 2012), underscores the intricate interplay of factors shaping wine characteristics. Regions like
499 Bourgogne, Bordeaux, and Champagne boast diverse soils, microclimates, grape varieties and
500 cultivation methods.

501 Analysing the mineral composition of wines and leveraging AI to process this data unlocks the
502 potential of authenticating wines at a granular geographical level. This necessitates working within
503 specific regions with hundreds of wines sourced from geologically homogeneous plots to ensure
504 precise metadata. In the medium-term, correlating this metadata with sensory profiles of wines
505 promises a deeper understanding of their origins and thus quality. The combination of Mineral Wine
506 Profile and Artificial Intelligence could thus be an indispensable tool for such investigations.

507 This study pioneers the development of a semi-quantitative method that enables rapid and robust
508 screening of 41 elements present in wines (about 200 samples can be analysed in just one day),
509 leading to the creation of a database of over 12000 Mineral Wine Profiles in just over a year. Here,
510 correlations between metal traces, rare earth elements, macro and micronutrients were initially
511 analysed, and their further exploration could be an intriguing avenue for future research endeavours.

512 By leveraging a large and diverse dataset, the present study developed an Extreme Gradient Boosting
513 model, which achieved mean accuracies of 92% for country classification, 91% for French wine
514 region and 85% for grape variety. Additionally, the initial specialisation of the dataset to assess the
515 performance of the model separating countries for red wines produced promising results, with an
516 increase in AUC scores (>0.9) and accuracy (>90%) for the classification of the three countries
517 tested. These findings have practical implications for the wine industry in that this comprehensive
518 dataset serves as a robust foundation for a versatile AI model capable of identifying a wine's origin
519 with over 99% specificity solely based on its Mineral Wine Profile.

520

521 Future research should focus on correlating the MWP and geological data to explore terroir signature,
522 as well as correlating the MWP and sensory profiles to delve more deeply into association of MWP
523 with the quality of wine. In conclusion, combining MWP and AI is indispensable for the wine
524 industry, which needs to cater to the ever-evolving demands of consumers for detailed origin
525 authentication beyond mere geographical regions.

526 **Acknowledgements**

527 We would like to express our gratitude to Victor Gomez and Henri-Laurent Arnould, wine
528 competition organisers, and to Gilles Masson, president of Centre du Rosé, for providing the majority
529 of the samples.

530 **References**

531 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean,
532 J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y.,
533 Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). *TensorFlow: Large-Scale*
534 *Machine Learning on Heterogeneous Distributed Systems*.

535 Alonso Gonzalez, P., Parga-Dans, E., Arribas Blázquez, P., Pérez Luzardo, O., Zumbado Peña, M.
536 L., Hernández González, M. M., Rodríguez-Hernández, Á., & Andújar, C. (2021). Elemental
537 composition, rare earths and minority elements in organic and conventional wines from
538 volcanic areas: The Canary Islands (Spain). *PLOS ONE*, *16*(11), e0258739.
539 <https://doi.org/10.1371/journal.pone.0258739>

540 Astray, G., Martinez-Castillo, C., Mejuto, J.-C., & Simal-Gandara, J. (2021). Metal and metalloid
541 profile as a fingerprint for traceability of wines under any Galician protected designation of
542 origin. *Journal of Food Composition and Analysis*, *102*, 104043.
543 <https://doi.org/10.1016/j.jfca.2021.104043>

544 Baleiras-Couto, M. M., & Eiras-Dias, J. E. (2006). Detection and identification of grape varieties in
545 must and wine using nuclear and chloroplast microsatellite markers. *Analytica Chimica Acta*,
546 *563*(1–2), 283–291. <https://doi.org/10.1016/j.aca.2005.09.076>

547 Bentlin, F. R. S., Pulgati, F. H., Dressler, V. L., & Pozebon, D. (2011). Elemental analysis of wines
548 from South America and their classification according to country. *Journal of the Brazilian*
549 *Chemical Society*, *22*(2), 327–336. <https://doi.org/10.1590/S0103-50532011000200019>

550 Castiñeira, M. del M., Brandt, R., Jakubowski, N., & Andersson, J. T. (2004). Changes of the Metal
551 Composition in German White Wines through the Winemaking Process. A Study of 63

552 Elements by Inductively Coupled Plasma–Mass Spectrometry. *Journal of Agricultural and*
553 *Food Chemistry*, 52(10), 2953–2961. <https://doi.org/10.1021/jf035119g>

554 Catarino S., Curvelo-Garcia A. S., & Bruno de Sousa, R. (2006). Measurements of contaminant
555 elements of wines by inductively coupled plasma-mass spectrometry: A comparison of two
556 calibration approaches. *Talanta*, 70, 1073-1080.
557 <https://doi.org/10.1016/j.talanta.2006.02.022>

558 Catarino, S., Madeira, M., Monteiro, F., Rocha, F., Curvelo-Garcia, A. S., & De Sousa, R. B. (2008).
559 Effect of Bentonite Characteristics on the Elemental Composition of Wine. *Journal of*
560 *Agricultural and Food Chemistry*, 56(1), 158–165. <https://doi.org/10.1021/jf0720180>

561 Cellier, R., Berail, S., Barre, J., Epova, E., Claverie, F., Ronzani, A.-L., Milcent, S., Ors, P., &
562 Donard, O. F. X. (2021). Analytical strategies for Sr and Pb isotopic signatures by MC-ICP-
563 MS applied to the authentication of Champagne and other sparkling wines. *Talanta*, 234,
564 122433. <https://doi.org/10.1016/j.talanta.2021.122433>

565 Chen, C., Dabek-Zlotorzynska, E., Rasmussen, P. E., Hassan, H., Lanouette, M. Evaluation of
566 semiquantitative analysis in ICP-MS. *Talanta*, 74, 1547-1555.
567 <https://doi.org/10.1016/j.talanta.2007.09.037>

568 Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the*
569 *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*,
570 785–794. <https://doi.org/10.1145/2939672.2939785>

571 Da Costa, N. L., Ximenez, J. P. B., Rodrigues, J. L., Barbosa, F., & Barbosa, R. (2020).
572 Characterization of Cabernet Sauvignon wines from California: Determination of origin
573 based on ICP-MS analysis and machine learning techniques. *European Food Research and*
574 *Technology*, 246(6), 1193–1205. <https://doi.org/10.1007/s00217-020-03480-5>

- 575 Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to
576 determine the geographical origin of foods in the European Union. *TrAC Trends in Analytical*
577 *Chemistry*, 40, 38–51. <https://doi.org/10.1016/j.trac.2012.08.003>
- 578 Ellis, D. I., Brewster, V. L., Dunn, W. B., Allwood, J. W., Golovanov, A. P., & Goodacre, R. (2012).
579 Fingerprinting food: Current technologies for the detection of food adulteration and
580 contamination. *Chemical Society Reviews*, 41(17), 5706. <https://doi.org/10.1039/c2cs35138b>
- 581 Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
582 <https://doi.org/10.1016/j.patrec.2005.10.010>
- 583 Feher, I., Magdas, D. A., Dehelean, A., & Sârbu, C. (2019). Characterization and classification of
584 wines according to geographical origin, vintage and specific variety based on elemental
585 content: A new chemometric approach. *Journal of Food Science and Technology*, 56(12),
586 5225–5233. <https://doi.org/10.1007/s13197-019-03991-4>
- 587 Fioschi, G., Prezioso, I., Sanarica, L., Pagano, R., Bettini, S., & Paradiso, V. M. (2024). Carrageenan
588 as possible stabilizer of calcium tartrate in wine. *Food Hydrocolloids*, 157, 110403.
589 <https://doi.org/10.1016/j.foodhyd.2024.110403>
- 590 Forina, M., Oliveri, P., Jäger, H., Römisch, U., & Smeyers-Verbeke, J. (2009). Class modeling
591 techniques in the control of the geographical origin of wines. *Chemometrics and Intelligent*
592 *Laboratory Systems*, 99(2), 127–137. <https://doi.org/10.1016/j.chemolab.2009.08.002>
- 593 Gajek, M., Pawlaczyk A., & Szykowska-Jozwik M. I. (2021). Multi-elemental analysis of wine
594 samples in relation to their type, origin, and grape variety. *Molecules*, 26, 214.
595 <https://doi.org/10.3390/molecules26010214>
- 596 Giaccio, M., & Vicentini, A. (2008). Determination of the geographical origin of wines by means of
597 the mineral content and the stable isotope ratios: A review. *Journal of Commodity Science,*
598 *Technology and Quality*, 47, 267–284.

- 599 Godshaw, J., Hopfer, H., Nelson, J., & Ebeler, S. (2017). Comparison of Dilution, Filtration, and
600 Microwave Digestion Sample Pretreatments in Elemental Profiling of Wine by ICP-MS.
601 *Molecules*, 22(10), 1609. <https://doi.org/10.3390/molecules22101609>
- 602 Griboff, J., Horacek, M., Wunderlin, D. A., & Monferrán, M. V. (2021). Differentiation Between
603 Argentine and Austrian Red and White Wines Based on Isotopic and Multi-Elemental
604 Composition. *Frontiers in Sustainable Food Systems*, 5, 657412.
605 <https://doi.org/10.3389/fsufs.2021.657412>
- 606 Hatzakis, E. (2019). Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A
607 Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, 18(1),
608 189–220. <https://doi.org/10.1111/1541-4337.12408>
- 609 Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa,
610 S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific
611 Reports*, 12(1), 5979. <https://doi.org/10.1038/s41598-022-09954-8>
- 612 International Organisation of Vine and Wine. (2024). *International Standard For The Labelling Of
613 Wines*. [https://www.oiv.int/sites/default/files/publication/2024-03/OIV-
614 %20Wine%20labelling%20Standard%20EN_2024%20final%20.pdf](https://www.oiv.int/sites/default/files/publication/2024-03/OIV-%20Wine%20labelling%20Standard%20EN_2024%20final%20.pdf)
- 615 Kang, X., Zhao, Y., & Tan, Z. (2023). An explainable machine learning for geographical origin
616 traceability of mussels *Mytilus edulis* based on stable isotope ratio and compositions of C, N,
617 O and H. *Journal of Food Composition and Analysis*, 123, 105508.
618 <https://doi.org/10.1016/j.jfca.2023.105508>
- 619 Kment, P., Mihaljevič, M., Ettlér, V., Šebek, O., Strnad, L., & Rohlová, L. (2005). Differentiation
620 of Czech wines using multielement composition – A comparison with vineyard soil. *Food
621 Chemistry*, 91(1), 157–165. <https://doi.org/10.1016/j.foodchem.2004.06.010>
- 622 Le Mao, I., Da Costa, G., & Richard, T. (2023). ¹H-NMR metabolomics for wine screening and
623 analysis. *OENO One*, 57(1), 15–31. <https://doi.org/10.20870/oeno-one.2023.57.1.7134>

- 624 Leeuwen, C. van, Barbe, J.-C., Darriet, P., Geffroy, O., Gomès, E., Guillaumie, S., Helwi, P.,
625 Laboyrie, J., Lytra, G., Menn, N. L., Marchand, S., Picard, M., Pons, A., Schüttler, A., &
626 Thibon, C. (2020). Recent advancements in understanding the terroir effect on aromas in
627 grapes and wines: This article is published in cooperation with the XIIIth International Terroir
628 Congress November 17-18 2020, Adelaide, Australia. Guest editors: Cassandra Collins and
629 Roberta De Bei. *OENO One*, 54(4), Article 4. <https://doi.org/10.20870/oenone.2020.54.4.3983>
- 631 Li, C., Kang, X., Nie, J., Li, A., Farag, M. A., Liu, C., Rogers, K. M., Xiao, J., & Yuan, Y. (2023).
632 Recent advances in Chinese food authentication and origin verification using isotope ratio
633 mass spectrometry. *Food Chemistry*, 398, 133896.
634 <https://doi.org/10.1016/j.foodchem.2022.133896>
- 635 Lima, M. M. M., Hernandez, D., & Runnebaum, R. C. (2023). Reproducibility of the Elemental
636 Profile of Pinot Noir Wines: A Comparison across Three Vintages. *ACS Food Science &*
637 *Technology*, 3(10), 1646–1653. <https://doi.org/10.1021/acsfoodscitech.3c00183>
- 638 Lima, M. M. M., Hernandez, D., Yeh, A., Reiter T., & Runnebaum, R. C. (2021). Reproducibility of
639 elemental profile across two vintages in Pinot noir wines from fourteen different vineyard
640 sites (2021). *Food Research International*. 141, 110045.
641 <https://doi.org/10.1016/j.foodres.2020.110045>
- 642 Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., & Wang, Z. (2021). XGBoost-based method
643 for flash flood risk assessment. *Journal of Hydrology*, 598, 126382.
644 <https://doi.org/10.1016/j.jhydrol.2021.126382>
- 645 Marre, A., Combaud, A., Chalumeau, L., & Philbiche, C. (2012, August). *Le concept de terroir en*
646 *Champagne: Un outil adaptable à toutes les échelles.*

647 Nistor, E., Dobrei, A. G., Mattii, G. B., Dobrei, A. Calcium and potassium accumulation
648 during the growth season in cabernet sauvignon and merlot grape variety (2022). *Plants*, *11*, 1536.
649 <https://doi.org/10.3390/plants11121536>

650 Pasvanka, K., Kostakis, M., Tarapoulouzi, M., Nisianakis, P., Thomaidis, N. S., & Proestos, C.
651 (2021). ICP–MS Analysis of Multi-Elemental Profile of Greek Wines and Their
652 Classification According to Variety, Area and Year of Production. *Separations*, *8*(8), 119.
653 <https://doi.org/10.3390/separations8080119>

654 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller,
655 A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,
656 A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). *Scikit-learn: Machine*
657 *Learning in Python*. <https://doi.org/10.48550/ARXIV.1201.0490>

658 Pérez-Magariño, S. (2004). Comparative study of artificial neural network and multivariate methods
659 to classify Spanish DO rose wines. *Talanta*, *62*(5), 983–990.
660 <https://doi.org/10.1016/j.talanta.2003.10.019>

661 Perini, M., & Bontempo, L. (2022). Liquid Chromatography coupled to Isotope Ratio Mass
662 Spectrometry (LC-IRMS): A review. *TrAC Trends in Analytical Chemistry*, *147*, 116515.
663 <https://doi.org/10.1016/j.trac.2021.116515>

664 Pohl, P. (2007). What do metals tell us about wine? *TrAC Trends in Analytical Chemistry*, *26*(9),
665 941–949. <https://doi.org/10.1016/j.trac.2007.07.005>

666 Popîrdă, A., Luchian, C. E., Cotea, V. V., Colibaba, L. C., Scutarașu, E. C., & Toader, A. M. (2021).
667 A Review of Representative Methods Used in Wine Authentication. *Agriculture*, *11*(3),
668 Article 3. <https://doi.org/10.3390/agriculture11030225>

669 Ranaweera, R. K. R., Gilmore, A. M., Bastian, S. E. P., Capone, D. L., & Jeffery, D. W. (2022).
670 Spectrofluorometric analysis to trace the molecular fingerprint of wine during the

671 winemaking process and recognise the blending percentage of different varietal wines.
672 *OENO One*, 56(1), 189–196. <https://doi.org/10.20870/oenone.2022.56.1.4904>

673 Salazar, F. N., & Achaerandio, I. (2006). Comparative Study of Protein Stabilization in White Wine
674 Using Zirconia and Bentonite: Physicochemical and Wine Sensory Analysis. *Journal of*
675 *Agricultural and Food Chemistry*, 54(26), 9955–9958. <https://doi.org/10.1021/jf062632w>

676 Schartner, M., Beck, J. M., Laboyrie, J., Riquier, L., Marchand, S., & Pouget, A. (2023). Predicting
677 Bordeaux red wine origins and vintages from raw gas chromatograms. *Communications*
678 *Chemistry*, 6(1), 247. <https://doi.org/10.1038/s42004-023-01051-9>

679 Silva-Barbieri, D., Salazar, F. N., López, F., Brossard, N., Escalona, N., & Pérez-Correa, J. R. (2022).
680 Advances in White Wine Protein Stabilization Technologies. *Molecules*, 27(4), 1251.
681 <https://doi.org/10.3390/molecules27041251>

682 Su, Y., Li, Y., Zhang, J., Wang, L., Rengasamy, K. R., Ma, W., & Zhang, A. (2023). Analysis of
683 soils, grapes, and wines for Sr isotope characterisation in Diqing Tibetan Autonomous
684 Prefecture (China) and combining multiple elements for wine geographical traceability
685 purposes. *Journal of Food Composition and Analysis*, 122, 105470.
686 <https://doi.org/10.1016/j.jfca.2023.105470>

687 Tanabe, C. K., Nelson, J., Boulton, R. B., Ebeler, S. E., & Hopfer, H. (2020). The Use of Macro,
688 Micro, and Trace Elemental Profiles to Differentiate Commercial Single Vineyard Pinot noir
689 Wines at a Sub-Regional Level. *Molecules*, 25(11), 2552.
690 <https://doi.org/10.3390/molecules25112552>

691 Temerdashev, Z., Khalafyan, A., Kaunova, A., Abakumov, A., Titarenko, V., & Akin'shina, V.
692 (2019). Using neural networks to identify the regional and varietal origin of Cabernet and
693 Merlot dry red wines produced in Krasnodar region. *Foods and Raw Materials*, 124–130.
694 <https://doi.org/10.21603/2308-4057-2019-1-124-130>

695 Temerdashev, Z., Khalafyan, A., Abakumov A., Bolshov M., Akin'shina V., & Kaunova A. (2024)
696 Authentication of selected white wines by geographical origin using ICP spectrometric and
697 chemometric analysis. *Heliyon*, *10*, e29607. <https://doi.org/10.1016/j.heliyon.2024.e29607>

698 van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning*
699 *Research*, *9*, 2579–2605.

700 Villano, C., Lisanti, M. T., Gambuti, A., Vecchio, R., Moio, L., Frusciante, L., Aversano, R., &
701 Carputo, D. (2017). Wine varietal authentication based on phenolics, volatiles and DNA
702 markers: State of the art, perspectives and drawbacks. *Food Control*, *80*, 1–10.
703 <https://doi.org/10.1016/j.foodcont.2017.04.020>

704 Villette, J., Cuéllar T., Verdeil J. L., Delrot S., & Gaillard I. (2020) Grapevine potassium nutrition
705 and fruit quality in the context of climate change. *Frontiers in Plant Science*, *11*, 123.
706 <https://doi.org/10.3389/fpls.2020.00123>

707 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski,
708 E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J.,
709 Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... Vázquez-
710 Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python.
711 *Nature Methods*, *17*(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>

712 Wen, J., Li, J., Wang, D., Li, C., Robbat, A., & Xia, L. (2023). Identification of geographical origin
713 of winter jujube based on GC–MS coupled with machine-learning algorithms. *Journal of*
714 *Food Composition and Analysis*, *124*, 105710. <https://doi.org/10.1016/j.jfca.2023.105710>

715 Wu, H., Lin, G., Tian, L., Yan, Z., Yi, B., Bian, X., Jin, B., Xie, L., Zhou, H., & Rogers, K. M.
716 (2021). Origin verification of French red wines using isotope and elemental analyses coupled
717 with chemometrics. *Food Chemistry*, *339*, 127760.
718 <https://doi.org/10.1016/j.foodchem.2020.127760>

719 Zambianchi, S., Soffritti, G., Stagnati, L., Patrone, V., Morelli, L., & Busconi, M. (2022). Effect of
720 storage time on wine DNA assessed by SSR analysis. *Food Control*, *142*, 109249.
721 <https://doi.org/10.1016/j.foodcont.2022.109249>

722 Zhang, D., Wei, Z., Han, Y., Duan, Y., Shi, B., & Ma, W. (2023). A Review on Wine Flavour Profiles
723 Altered by Bottle Aging. *Molecules*, *28*(18), Article 18.
724 <https://doi.org/10.3390/molecules28186522>

725