# 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.

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

Leticia Sarlo[1,2], Coraline Duroux[2], Yohann Clément[3], Pierre Lanteri[3], Fabien Rossetti[1], Olivier David[4], Augustin Tillement[2,5], Philippe Gillet[2], Agnès Hagège[3], Laurent David[5], Michel Dumoulin[6], Richard Marchal[7,8], Théodore Tillement[2], François Lux[1,9,*], Olivier Tillement[1]

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

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

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

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

[5] Universite Claude Bernard Lyon 1, Institut National des Sciences Appliquées, Université Jean Monnet, CNRS, UMR 5223, Ingénierie des Matériaux Polymères, 15 bd Latarjet, 69622 Villeurbanne, France

[6] Agro Œno Conseil, Mâcon, France

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

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

[9] Institut Universitaire de France (IUF), Paris

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

## Abstract

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

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

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

## Introduction

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

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

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

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

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

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

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

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

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

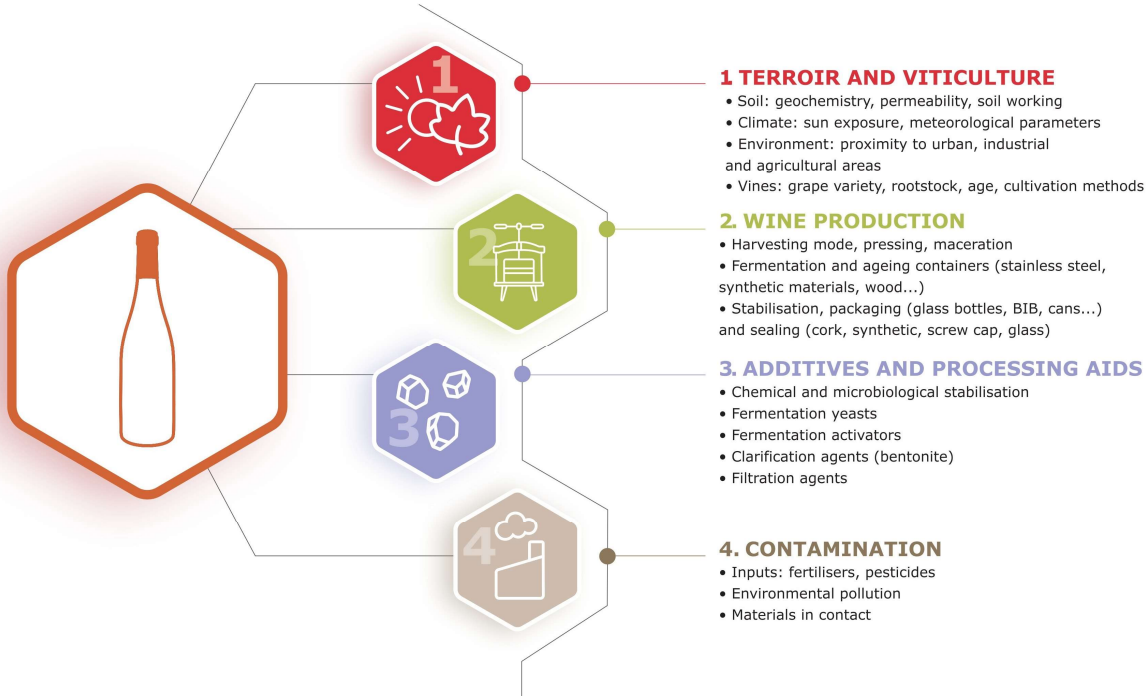



**Figure 1. Description of different factors that can influence the concentration of major, minor and trace mineral elements in bottled wine, namely terroir and viticulture, wine production, additives and contamination.**

## Materials and methods

### 1. Reagents and materials

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

metal-free tubes (VWR®) were employed for collecting and preparing both samples and standards. A semi-quantitative calibration standard was prepared by diluting the multi-element standard (Reference 85006.186), purchased from VWR, with 100 mg/L of Al, Ag, As, B, Ba, Be, Bi, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, Pb, Sb, Se, Sr, Ti, Tl, V and Zn, in 5% $HNO_3$ (v/v). ICP-MS tuning solution, containing 1 µg/L of Ce, Co, Li, Tl and Y in 2% $HNO_3$ (v/v) (Agilent Technologies), was used to optimise the ICP-MS signal intensities. The use of these solutions is described below. Moreover, a commercial red wine was used as a final control to ensure reproducible results (i.e., not exceeding 15% variations) over time. An indium solution, used as an internal standard to ensure good sample conservation, was prepared by diluting the 1000 mg/L indium standard in 4% $HNO_3$ (v/v) (purchased from SCP Science) and then added to each sample.

## 2. Sample preparation

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

Samples of approximately 30 mL of wine were put in metal-free tubes. Direct wine dilution was found to provide optimal balance in terms of user-friendliness, result accuracy and precision (Godshaw, 2017). Samples were diluted 1:3 using 1% $HNO_3$ (v/v) and 10 µg/L of indium standard solution. This initial dilution provides sample storage in acidic conditions, ensuring the preservation of elemental composition over time and limiting mineral precipitation and adsorption onto the metal-free tube walls. A second dilution of 1:5 using 1% $HNO_3$ (v/v) was performed just before the analysis. The total dilution factor (1:15) was fine-tuned to minimise matrix effects, which can occur due to the presence of alcohol or other organic matter when performing trace element quantification (Catarino, 2006).

## 3. ICP-MS analysis

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

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

Elemental concentrations were obtained through SQ analysis using the 28-element standard at a concentration of 20 µg/L and 1% $HNO_3$ as the blank. SQ approach was performed for 42 elements, with 100 sweeps and one replicate: $^{11}B$, $^{23}Na$, $^{24}Mg$, $^{27}Al$, $^{28}Si$, $^{31}P$, $^{34}S$, $^{35}Cl$, $^{39}K$, $^{43}Ca$, $^{45}Sc$, $^{47}Ti$, $^{51}V$, $^{52}Cr$, $^{55}Mn$, $^{56}Fe$, $^{59}Co$, $^{60}Ni$, $^{63}Cu$, $^{66}Zn$, $^{75}As$, $^{79}Br$, $^{85}Rb$, $^{88}Sr$, $^{89}Y$, $^{90}Zr$, $^{93}Nb$, $^{111}Cd$, $^{115}In$, $^{118}Sn$, $^{127}I$, $^{133}Cs$, $^{137}Ba$, $^{139}La$, $^{140}Ce$, $^{141}Pr$, $^{146}Nd$, $^{147}Sm$, $^{182}W$, $^{205}Tl$, $^{208}Pb$ and $^{238}U$. These 41 elements, except for In used as internal standard, constitute the mineral wine profile (MWP). The following elements were absent from the calibration standard: $^{28}Si$, $^{31}P$, $^{34}S$, $^{35}Cl$, $^{45}Sc$, $^{79}Br$, $^{85}Rb$, $^{89}Y$, $^{90}Zr$, $^{93}Nb$, $^{115}In$, $^{118}Sn$, $^{127}I$, $^{133}Cs$, $^{139}La$, $^{140}Ce$, $^{141}Pr$, $^{146}Nd$, $^{147}Sm$, $^{182}W$ and $^{238}U$. Their concentrations are interpolated between elements present in the calibration standard, applying response factors, which depend on their isotopic mass, isotopic abundance and ionisation energy.

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

## 4. Statistical analysis

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

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

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

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

## 5. Sample classification

### 5.1.Selection of machine learning model

For model selection, the dataset underwent 80:20 stratified random split, to compose the train and the test set, respectively. The test set was used to verify the trained model performance, as it had not been previously seen by the model. All samples with unknown label values were taken out of the dataset before the stratified split was done. Both sets were composed of all 39 elements (B, Na, Mg, 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, La, Ce, Pr, Nd, Sm, W, Tl, Pb and U).

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

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

### 5.2.Application of XGB in Sample Classification

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

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

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

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

## Results and discussion

### 1. Determining the elemental composition of the wine samples

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

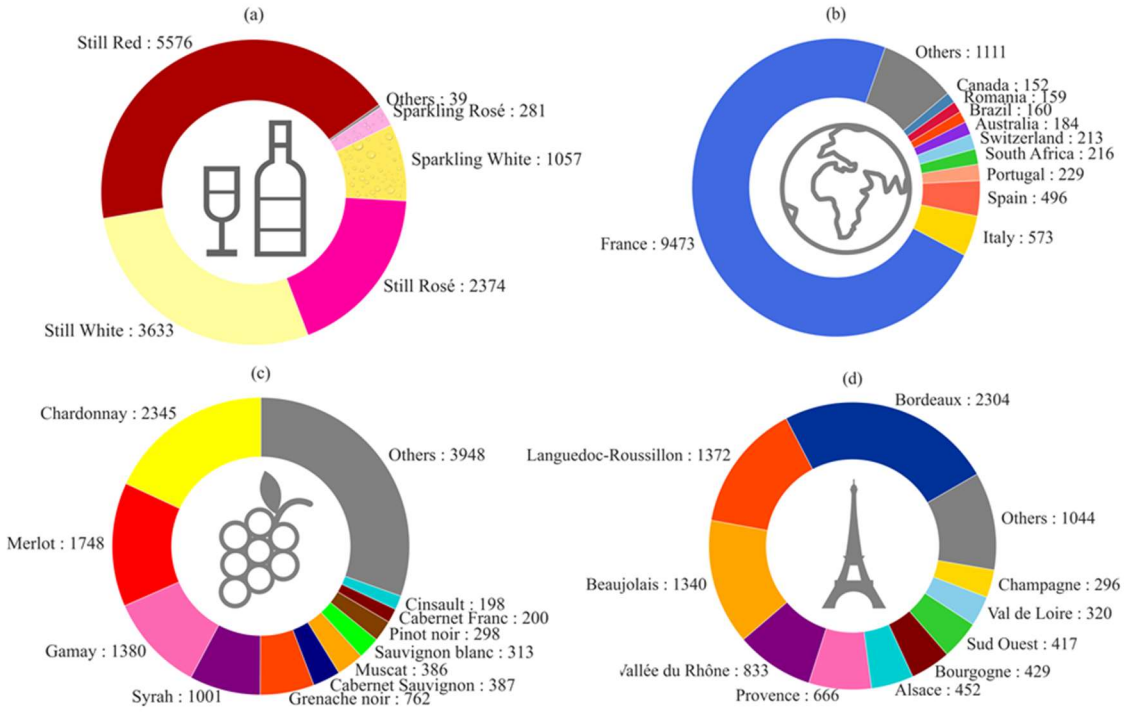


**Figure 2. Distribution of the 12966 analysed wines based on the wine type (a), country (b), grape variety (c), and French wine region (d). The category "Others" contains labels with fewer than 150 samples. A more detailed description of each category is given in Tables S1, S2 and S3.**

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

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

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

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

**Table 1. Mean, median and interquartile range (IQR) for the 39 concentrations of mineral 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

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

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

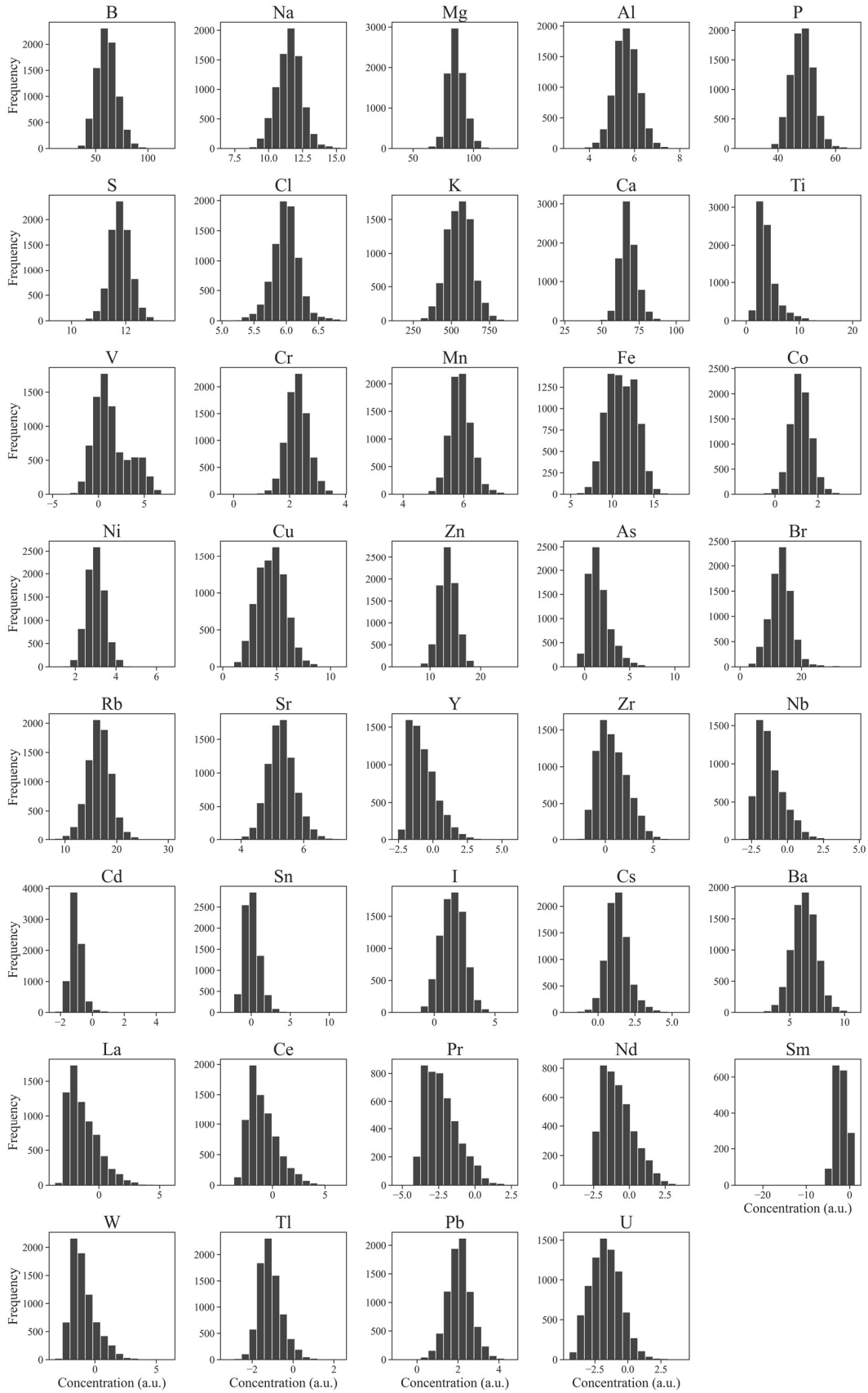

B
Na
Mg
Al
P
S
Cl
K
Ca
Ti
V
Cr
Mn
Fe
Co
Ni
Cu
Zn
As
Br
Rb
Sr
Y
Zr
Nb
Cd
Sn
I
Cs
Ba
La
Ce
Pr
Nd
Sm
W
Tl
Pb
U
Frequency
Concentration (a.u.)

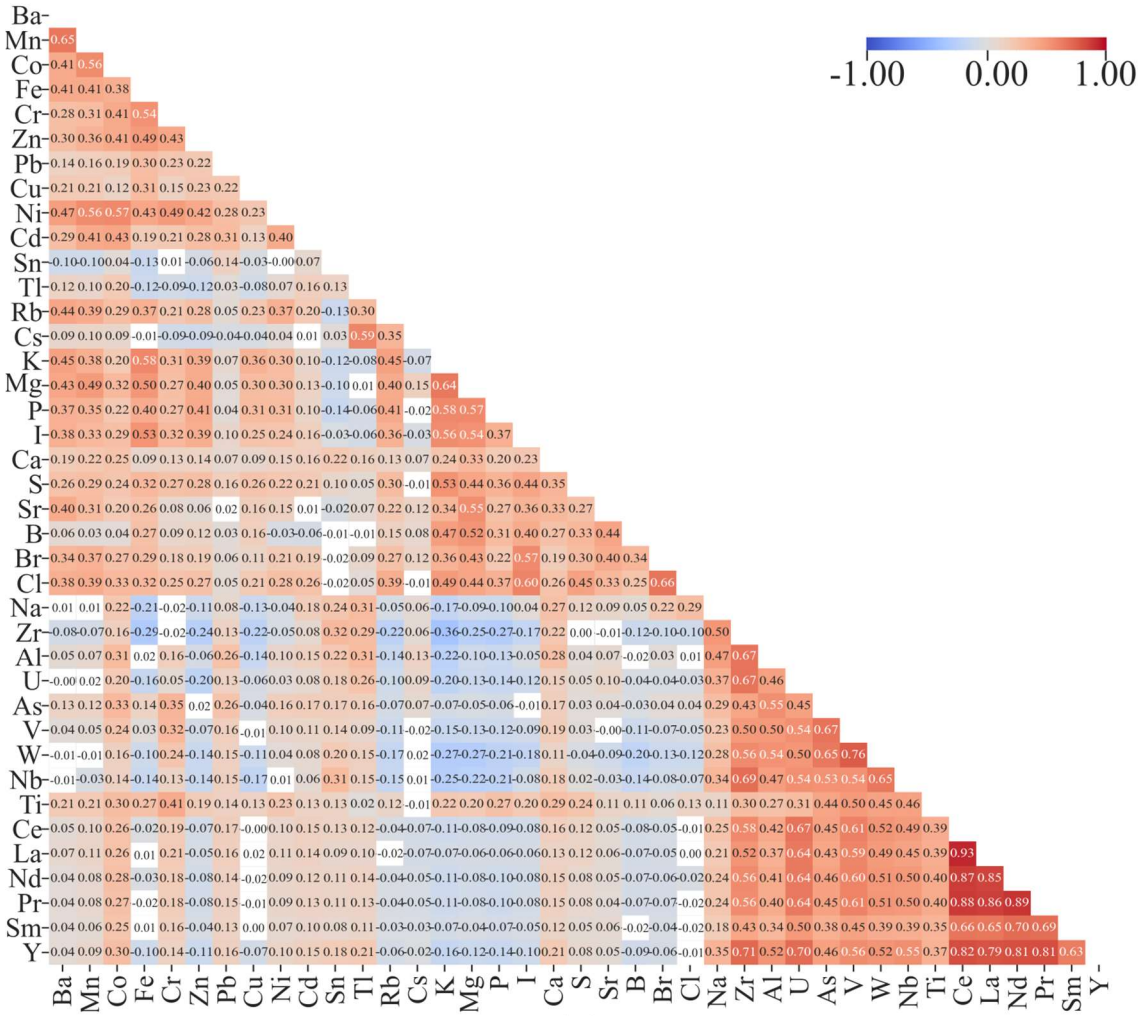
**Figure 3. Element concentration distribution in the database. Concentration (x-axis) is transformed using the Box-Cox power transformation. If the concentration value was under the limit of quantification, the data was not plotted. The concentration and frequency scales have been adapted for each element.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

## 2. Statistical analysis

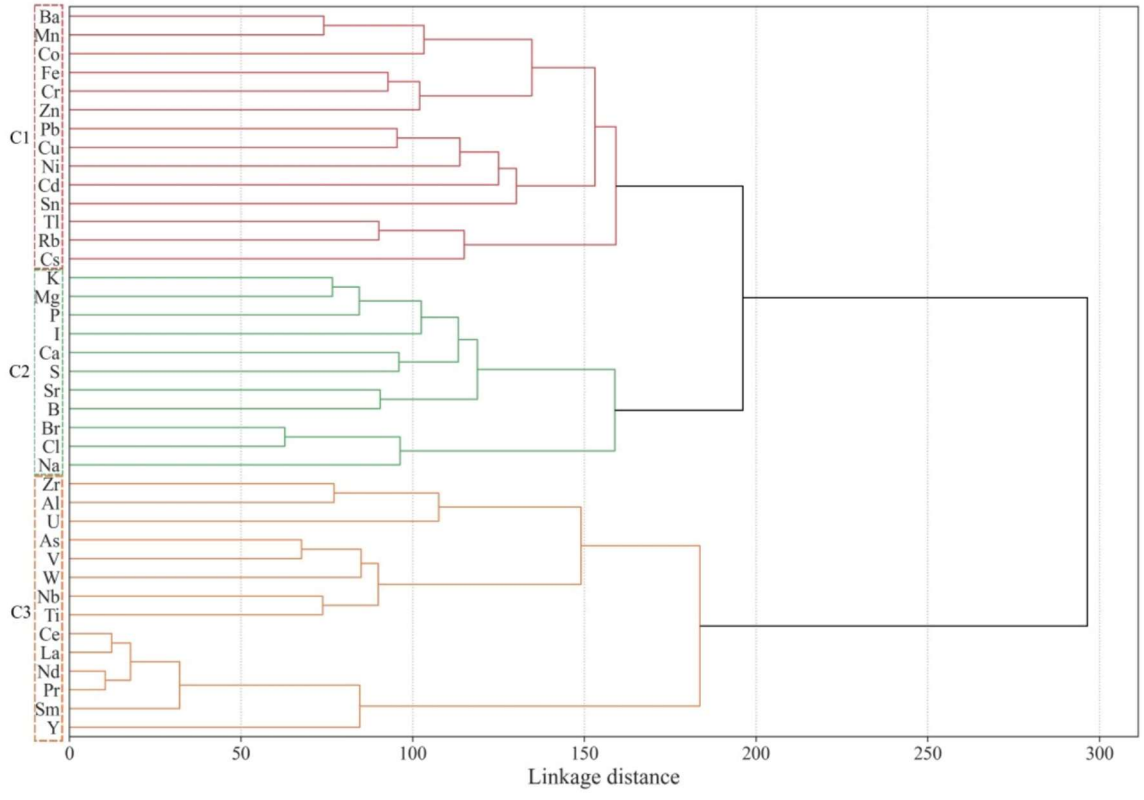
### 2.1. Correlation analysis

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

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


-1.00
0.00
1.00
Ba Mn Co Fe Cr Zn Pb Cu Ni Cd Sn Tl Rb Cs K Mg P I Ca S Sr B Br Cl Na Zr Al U As V W Nb Ti Ce La Nd Pr Sm Y

(a)


C1
C2
C3
Ba Mn Co Fe Cr Zn Pb Cu Ni Cd Sn Tl Rb Cs K Mg P I Ca S Sr B Br Cl Na Zr Al U As V W Nb Ti Ce La Nd Pr Sm Y
0 50 100 150 200 250 300
Linkage distance

(b)

**Figure 4. (a) Correlation matrix plot illustrating the relationships between 39 pairs of element concentrations across all samples of the database. The colour gradient reflects the Spearman correlation coefficient (ρ), with statistically significant correlations (*p*-value < 0.05) indicated in red (positive correlation) or blue (negative correlation). Non-significant correlations are in white cells. (b) Hierarchical cluster dendrogram generated using Ward's method and Euclidean distance for the 39 elements. Distances indicate the degree of correlation between different elements. Three clusters are identifiable: C1 (Ba, Mn, Co, Fe, Cr, Zn, Pb, Cu, Ni, Cd, 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, Nb, Ti, Ce, La, Nd, Pr, Sm and Y).**

## 2.2. Cluster analysis

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

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

## 2.3. Principal Component Analysis

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

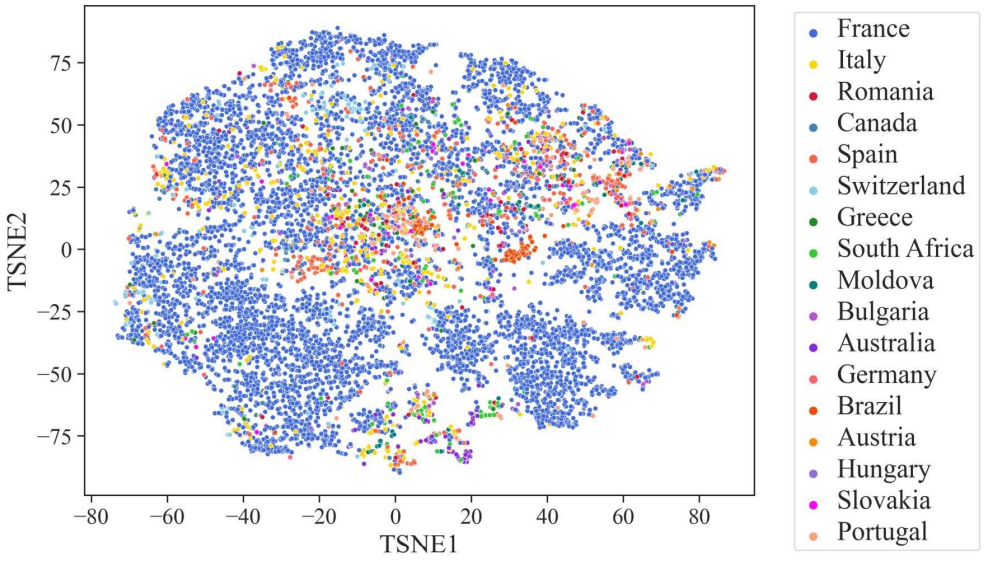
Initially, the grouping of all the samples was studied, as shown in Figure 5 and Figure S2. In terms of wine type, red wines are well separated from rosé and white wines, which is expected due to the

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

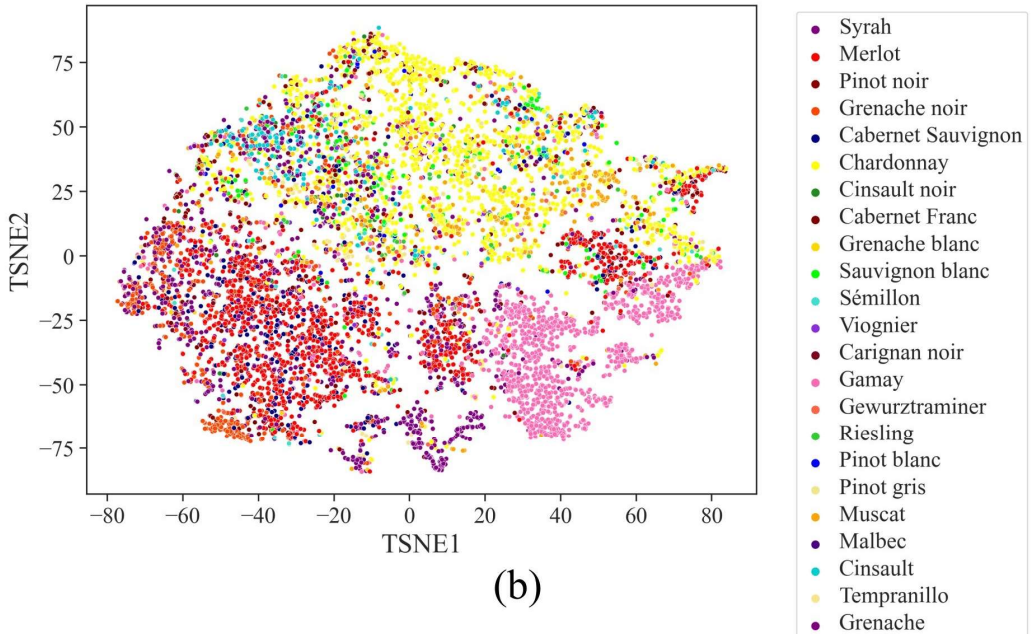
For French wine type, the separation of red wines from white and rosés is clear. For the French wine regions and principal grape variety, all the categories with more than 50 samples have been represented and only French samples with known categories have been plotted. Clusters are evident for the Beaujolais,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

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

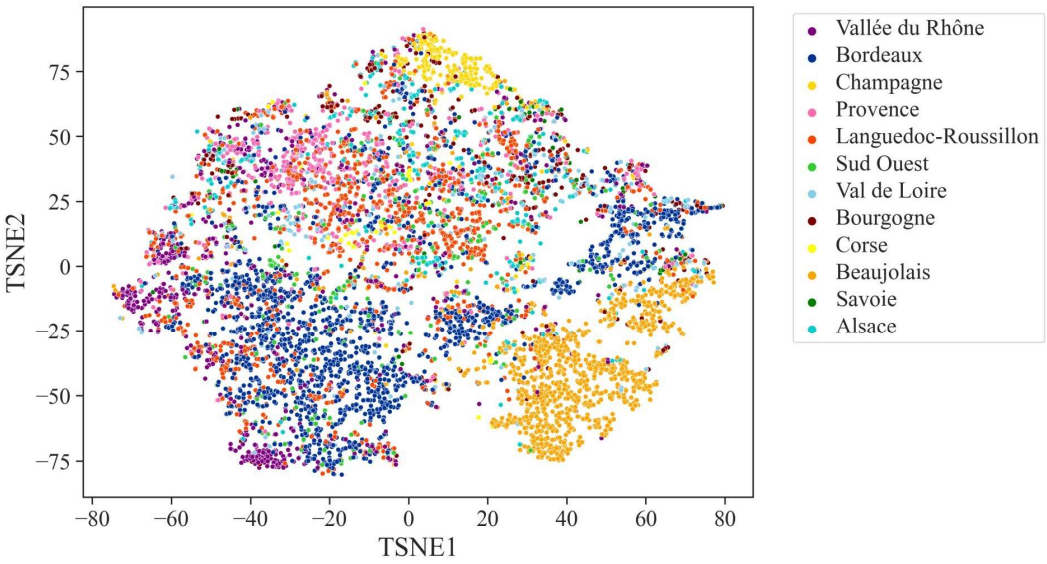
The promising patterns observed in both the data grouping and a rich database containing more than 12000 samples have motivated the development of machine learning techniques tailored to sample classification.


TSNE2
TSNE1
75
50
25
0
−25
−50
−75
−80
−60
−40
−20
0
20
40
60
80
France
Italy
Romania
Canada
Spain
Switzerland
Greece
South Africa
Moldova
Bulgaria
Australia
Germany
Brazil
Austria
Hungary
Slovakia
Portugal

(a)


TSNE2
TSNE1
75
50
25
0
−25
−50
−75
−80
−60
−40
−20
0
20
40
60
80
Syrah
Merlot
Pinot noir
Grenache noir
Cabernet Sauvignon
Chardonnay
Cinsault noir
Cabernet Franc
Grenache blanc
Sauvignon blanc
Sémillon
Viognier
Carignan noir
Gamay
Gewurztraminer
Riesling
Pinot blanc
Pinot gris
Muscat
Malbec
Cinsault
Tempranillo
Grenache

(b)


TSNE2
TSNE1
75
50
25
0
−25
−50
−75
−80
−60
−40
−20
0
20
40
60
80
Vallée du Rhône
Bordeaux
Champagne
Provence
Languedoc-Roussillon
Sud Ouest
Val de Loire
Bourgogne
Corse
Beaujolais
Savoie
Alsace

(c)

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

## 3. Sample classification

### 3.1.Selecting the machine learning model

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

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

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

### 3.2.Application of XGB in Sample Classification

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

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

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

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

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

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

**Table 3. Mean ($N_{iteration}$ = 10) performance metrics for predicting a wine's country, French region and principal grape variety. Only labels with 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 | Gewurztraminer | 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 | | | | | | |

414

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

**Table 4. Mean ($N_{iteration}$ = 10) performance metrics when classifying the countries of French, 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 |

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

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

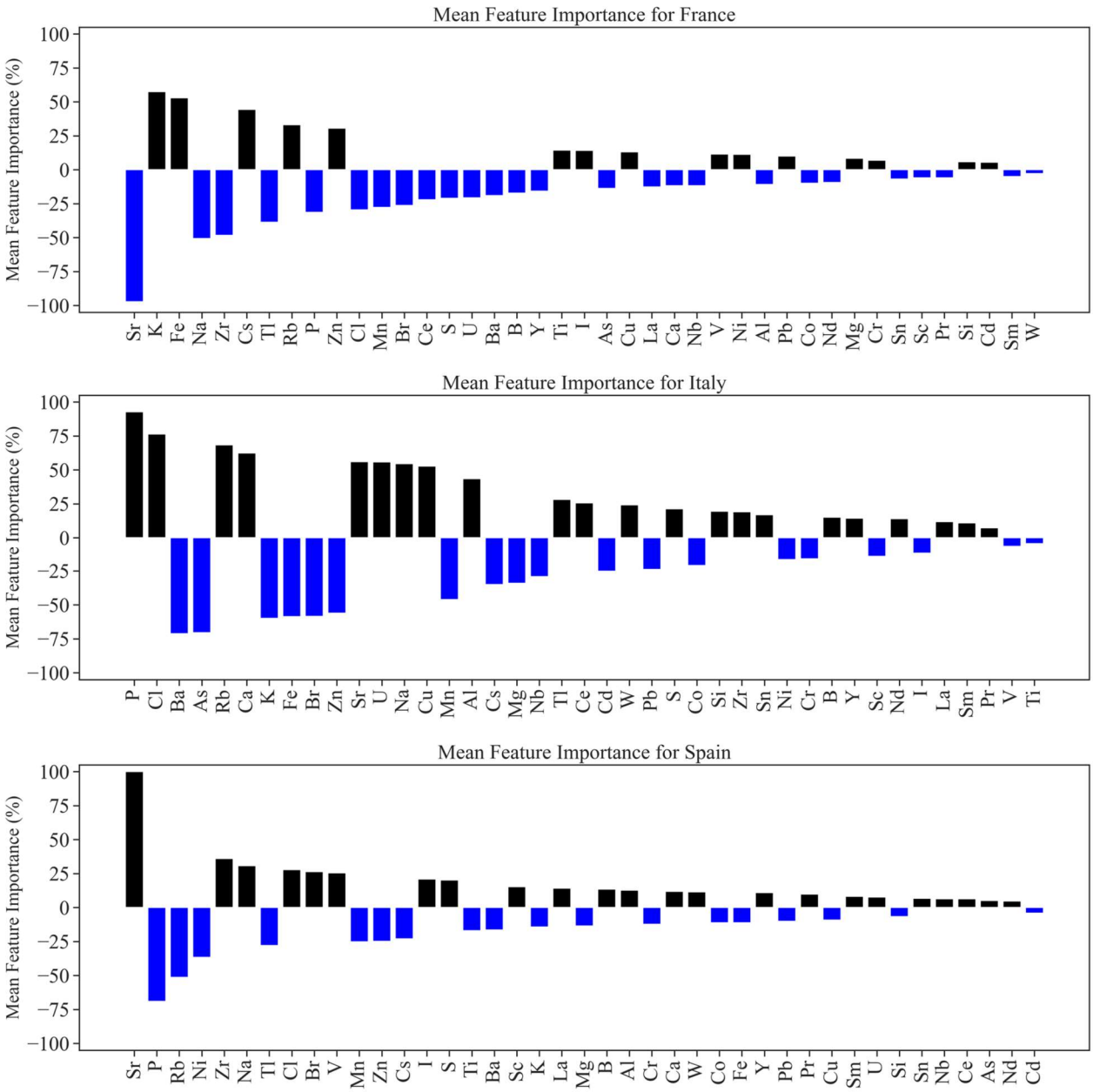



**Figure 6. Element Mean Feature Importance for the three major countries. Values are presented in decreasing order of importance. Values in black indicate positive correlations and values in blue, negative correlations between the element's concentration and the wine category. These values for the three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

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

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

**Table 5. Mean ($N_{iteration}$ = 10) performance metrics for classifying the three major categories of 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 |

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

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

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

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

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

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

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

## Conclusion

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

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

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

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

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

## Acknowledgements

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

## References

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

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

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

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

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

Castiñeira, M. del M., Brandt, R., Jakubowski, N., & Andersson, J. T. (2004). Changes of the Metal 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., & Szynkowska-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

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., Ettler, 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). $^{1}$H-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/oeno-
630 one.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*.

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

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

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

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

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

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

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

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

winemaking process and recognise the blending percentage of different varietal wines. *OENO One*, *56*(1), 189–196. https://doi.org/10.20870/oeno-one.2022.56.1.4904

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

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

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

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

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

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

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

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

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

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

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

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

Wu, H., Lin, G., Tian, L., Yan, Z., Yi, B., Bian, X., Jin, B., Xie, L., Zhou, H., & Rogers, K. M. (2021). Origin verification of French red wines using isotope and elemental analyses coupled with chemometrics. *Food Chemistry*, *339*, 127760. 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