separations

MDPI

Article

UHPLC–UV Fingerprinting for Wine Characterization by Geographical Origin and Grape Variety

Julia Calvo Salvador [1], Diego A. Ahumada Forigua [1], Sonia Sentellas [1,2,3] and Javier Saurina [1,2,*]

[1] Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona, Spain; jcalvosalvador@gmail.com (J.C.S.); daahumadaf@gmail.com (D.A.A.F.); sonia.sentellas@ub.edu (S.S.)
[2] Institut de Recerca en Nutrició i Seguretat Alimentària, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain
[3] Serra Húnter Fellow Programme, Generalitat de Catalunya, Via Laietana 2, 08003 Barcelona, Spain
* Correspondence: xavi.saurina@ub.edu

Abstract

The characterization and authentication of wines are essential to ensure quality and traceability, as well as to protect the added value linked to geographical origin, vinification processes, and grape variety. This study develops a non-targeted approach based on ultra-high-performance liquid chromatography with ultraviolet detection (UHPLC-UV) to obtain chromatographic fingerprints at 280 nm, mainly reflecting phenolic acids and flavanols. Data preprocessing, including correction of peak shifts, baseline drifts, and other artifacts, was carried out using open-source software further adapted in-house. Principal Component Analysis (PCA) revealed clear sample clustering according to Protected Designation of Origin (PDO) and grape variety, demonstrating the discriminative potential of the fingerprints. Supervised classification methods, including Partial Least Squares Discriminant Analysis (PLS-DA) and Soft Independent Modelling of Class Analogies (SIMCA), provided good predictive performance, with mean classification errors below 20.6%. Misclassifications were mainly associated with geographically close PDOs and multivarietal blends. Somontano PDO and Tempranillo wines were the most accurately classified categories.

Keywords: red wine; liquid chromatography; fingerprinting; protected designation of origin; grape variety; chemometrics; wine classification

1. Introduction

Wine is one of the oldest and most emblematic beverages in human history, with a documented presence across nearly all Mediterranean civilizations for thousands of years [1,2]. Over time, vineyards have become an essential component of cultural heritage, contributing to rural identity while supporting economic development. As a result, wine production represents a strategic sector in many economies, encompassing a complex value chain that spans from agricultural practices to commercialization and wine tourism [2,3].

According to the International Organisation of Vine and Wine (OIV), global wine consumption reached approximately 197 million hectoliters in 2025, with the European Union remaining the leading producer and consumer [4]. Countries such as France, Italy, and Spain play a dominant role due to their production capacity, diversity, and international reputation [4,5]. In this competitive context, wine quality is increasingly associated not only with sensory attributes but also with authenticity and geographical origin [6,7]. Consequently, certification systems such as Protected Designations of Origin (PDOs), Protected

Check for updates

Academic Editor: Fabio Gosetti

Received: 28 August 2026
Revised: 8 October 2026
Accepted: 13 October 2026
Published: 14 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Geographical Indications (PGIs), and other regulatory frameworks have been established to guarantee provenance, grape variety, and production methods [5].

Despite these regulatory efforts, the wine sector remains vulnerable to fraud and misleading practices, including mislabeling and adulteration, which can undermine consumer trust and cause economic damage. This situation needs reliable analytical methodologies for ensuring wine authenticity and detecting fraudulent practices [8,9]. From this analytical perspective, wine characterization is inherently complex, involving multiple factors such as grape variety, geographical origin, vintage, environmental conditions, and winemaking practices, all of which directly influence the chemical composition of the final product. This remains a highly topical issue in the field of food analysis, involving the development and application of effective, robust, rapid, and cost-efficient analytical methods, compatible with the analysis of large sample sets, together with appropriate data treatment, typically based on chemometric approaches. These strategies are reflected in recent review papers, which highlight the continuous growth and relevance of this research area [3,5,8,9]. A wide range of analytical techniques has been applied for wine characterization, including spectroscopy, chromatography, mass spectrometry, and elemental analysis. Regarding data preprocessing and chemometric approaches, principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and other machine learning algorithms have demonstrated high performance in applications related to wine authentication and quality assessment [10–15].

Focusing on the possibilities of liquid chromatography with different detection techniques, many publications are based on compositional profiles of various families of compounds commonly found in wines, among which polyphenols, organic acids, amino acids, biogenic amines, and volatile organic compounds stand out, among others [5,9]. On the other hand, approaches based on chromatographic fingerprints—which do not require quantification nor assume which compounds may serve as descriptors of specific attributes—are gaining increasing relevance due to their versatility and procedural simplicity [16,17]. Nevertheless, the most common fingerprinting strategies still rely on rapid spectroscopic techniques for fast screening, then using the resulting spectral data as sample fingerprints [18–22].

Regarding liquid chromatographic fingerprints, some representative recent studies include: LC-QTOF-MS-based untargeted metabolomic authentication of Chinese red wines according to their grape varieties [23] or geographical origin [24]; LC-IM-QTOF-MS metabolomics for distinguishing geographical origin and vintage of Chinese red wine [25], which provides an additional ion mobility dimension for enhanced discrimination power; LC-HRMS with a hybrid quadrupole-orbitrap analyzer for varietal authentication of wines based on data [26]. Interestingly, the resulting features may vary depending on the LC–MS platform employed, which adds complexity but also offers complementary information when needed [27]. As can be seen, these studies are based on high-resolution platforms, which are associated with considerable complexity and cost. Therefore, alternative approaches based on simpler and more economical chromatographic fingerprints would be highly desirable. In this regard, liquid chromatography with UV–Vis detection emerges as a very attractive and practical option, offering low cost, simplicity, robustness, and widespread availability in many analytical laboratories.

In this paper, we aim to apply a simple non-targeted approach combined with open-source preprocessing tools developed by the authors for the characterization of Spanish wines from different geographical origins and grape varieties. The available wine set comprises 198 samples analyzed with an ultrafast separation method optimized to maximize the number of resolved compounds—estimated from the chromatographic peak count—while minimizing the total analysis time. Exploratory studies based on fingerprint

show remarkable trends and distributions related to attributes such as geographical origin and grape variety. These features provide the basis for subsequent classification and authentication studies, in which both one-class and multiclass models, as well as different classification rules, are explored.

2. Materials and Methods

2.1. Reagents

For the chromatographic analysis, the following solvents and reagents were used: formic acid (purity $\geq$ 95%) from Sigma-Aldrich (St. Louis, MO, USA); purified water obtained from a Milli-Q system equipped with a Millipak Express 40 filter from Merck (Darmstadt, Germany); HPLC-grade acetonitrile (ACN) from PanReac AppliChem (Castellar del Vallès, Spain).

2.2. Wine Samples

A total of 198 red wine samples purchased in 2025 were analyzed. The samples were stored frozen and allowed to thaw at room temperature prior to processing. The study focused on the classification of wines from four PDOs (Montsant, Ribera del Duero, Rioja, and Somontano) and three monovarietal grape varieties: Garnacha (G), Syrah (S), and Tempranillo (T). The remaining PDOs and grape varieties, whether mono- or multivarietal, were grouped into a single category labeled "Others." As an example, Figure 1 shows pie charts illustrating the distribution of samples by geographical origin and grape variety. A complete description of the samples is provided in Table S1 of the Supplementary Material.

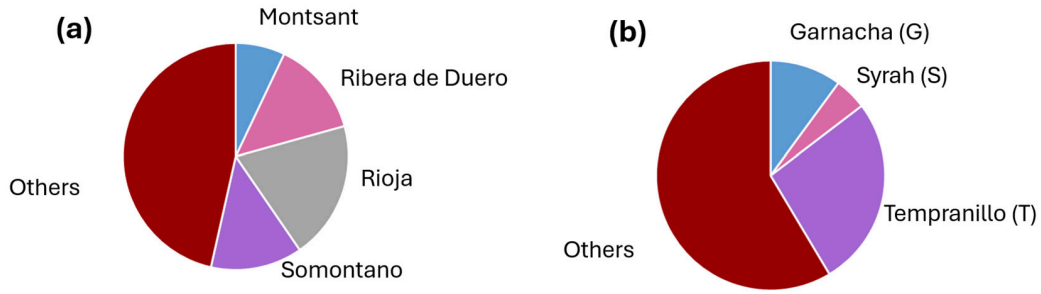

Figure 1. Pie charts with the distribution of samples by geographical origin (**a**) and grape variety (**b**).

2.3. Sample Treatment

The wine samples were filtered using a 6 mL syringe equipped with a 0.45 µm nylon membrane filter (Agilent Technologies, Santa Clara, CA, USA). The filtered samples were then transferred to 2 mL clear vials and stored at -4 °C until analysis. In parallel, a Quality Control (QC) sample was prepared by pooling 100 µL from each filtered wine. This QC sample was used to evaluate the repeatability of the developed methods, as well as the robustness of the chemometric results.

2.4. Chromatographic Method

The chromatographic fingerprints were obtained using an Acquity UPLC system (Waters, Milford, MA, USA) equipped with a photodiode array (PDA) detector, an automatic injection system (Sample Manager) and a Sample Organizer (refrigerated at 6 °C to ensure sample stability). The system was controlled using Empower software (version 3.6.1, Waters). The chromatographic separation was performed on an Acquity UPLC BEH C18 reversed-phase column (50 mm $\times$ 2.1 mm i.d., 1.7 µm, 130 Å; Waters) equipped with a 0.2 µm in-line filter. The mobile phase consisted of 0.1% (v/v) aqueous formic acid and ACN. The chromatographic gradient started at 2% ACN, which was maintained from 0.0

to 0.5 min. The ACN proportion was then linearly increased to 20% at 7.5 min, followed by a rapid increase to 90% ACN at 7.6 min. This composition was maintained until 8.1 min. The mobile phase was then returned to the initial composition (2% ACN) at 8.2 min and maintained until 10.0 min to allow column re-equilibration, resulting in a total run time of 10 min. The flow rate was 0.6 mL min^{-1} and the injection volume was 2 µL. Wine samples were analyzed in triplicate throughout the analytical sequence, with a QC sample injected every 15 samples. The injection order within each sequence was randomized to minimize potential confounding effects associated with instrumental drift or other uncontrolled variables that could influence subsequent chemometric analyses.

2.5. Data Analysis

A total of 636 chromatograms (594 sample injections and 42 QC injections) recorded at 280 nm were exported and assembled into a data matrix using the R programming language (version 4.4.2) [28], where rows corresponded to samples and columns to temporal measurement channels. Chromatograms were aligned using the most intense peaks as references, and baseline correction and smoothing were performed using the R packages alkahest [29] and pracma [30], respectively. A working time window between 0.76 and 7.5 min was then selected. This window encompassed the region containing relevant analytical information while excluding dead-volume, washing, and re-equilibration segments. After preprocessing, autoscaling was applied to ensure that all variables contributed equally to the description of sample variance. The resulting data matrices were initially explored by PCA to identify potential clustering patterns among sample classes.

Subsequently, multivariate classification models were developed to evaluate the predictive performance for PDO and grape-variety classification based on the instrumental responses obtained. For this purpose, the data matrices were randomly divided into two subsets: 70% of the samples were used for model calibration and the remaining 30% for external validation. Hence, all replicate injections corresponding to a given wine were always assigned to the same subset (calibration or prediction), thus preventing any data leakage between datasets. Classification models based on PLS-DA and SIMCA were constructed, with PLS-DA implemented in both binary (class membership vs. non-membership) and multiclass modes. In all cases, an "Others" category was included, comprising samples not belonging to the target classes.

For repeated external validation of the PLS-DA models, 100 independent calibration/prediction partitions were generated. Within each calibration set, 10-fold cross-validation was performed at the independent-wine level, with the folds constructed separately for each class. For binary PLS-DA models, the classification threshold was selected from the cross-validation predictions by maximizing balanced accuracy, defined as the arithmetic mean of sensitivity and specificity. Before calculating the external prediction metrics, replicate injections from each wine were aggregated at the sample level.

Chemometric analyses were carried out using SOLO 9.5 (Eigenvector Research, Manson, WA, USA). Additional statistical analyses were performed using R Studio (Version 2026.07.1+147). Support Vector Machine (SVM) algorithms were implemented in R using the e1071 package with a radial basis function (RBF) kernel. The model hyperparameters, namely the penalty parameter (C) and kernel bandwidth (γ), were optimized by grid search during internal cross-validation using only the calibration dataset. Within each repetition, five-fold grouped cross-validation was performed using the calibration dataset for hyperparameter optimization. The predictive performance of the optimized SVM models was subsequently evaluated using the external validation subset. Performance variability across the 100 iterations was assessed, and the reported C and γ values represent the arithmetic means of the values selected across the 100 internal cross-validation procedures.

Classification performance was evaluated using class-specific metrics, including sensitivity, specificity, classification error, and balanced accuracy. For multiclass PLS-DA models, these metrics were calculated for each target class using a one-versus-rest approach. For each external prediction partition, 1000 bootstrap resamples were generated separately from the positive and negative wine samples, preserving the original number of samples in each class. The 95% bootstrap confidence intervals were calculated from the 2.5th and 97.5th percentiles of the resulting bootstrap distributions.

3. Results

3.1. Optimization of the Chromatographic Separation

The selection of the mobile phase components (0.1% formic acid in water and ACN) and the separation column (reversed-phase C18) was based on previous studies focused on polyphenol determination [31]. In this study, the elution gradient was optimized to maximize the number of resolved compounds within the shortest possible analysis time, thereby achieving a compromise between chromatographic resolution and overall method throughput.

A full factorial experimental design involving two factors, namely the final ACN percentage and the gradient duration, each evaluated at five levels, was initially established, resulting in a total of 25 experimental runs. The final ACN percentages evaluated were 15%, 18%, 20%, 22%, and 25%, whereas the gradient durations tested were 4, 5, 6, 7, and 9 min. In all cases, the gradient started at an initial composition of 2% ACN. The number of visible peaks in each chromatogram and the retention time of the last detected peak were used to assess the method performance in terms of both separation capability and analysis time. These responses were subsequently used to generate the corresponding response surfaces, which are shown as 3D plots in Figure 2.

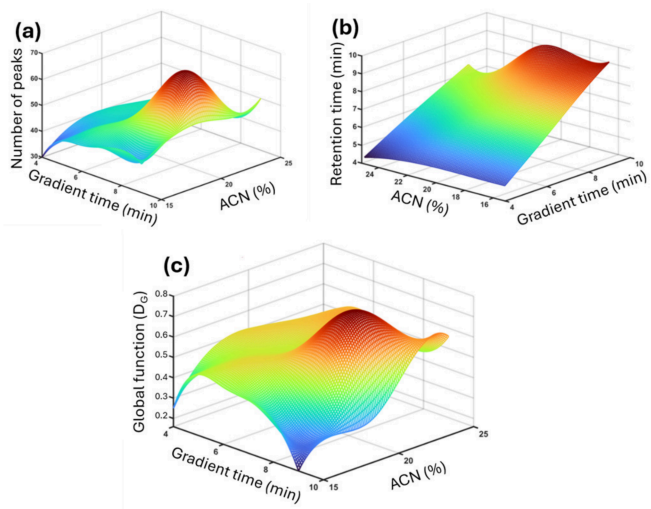

Figure 2. Response surfaces used for the optimization of the UHPLC-UV chromatographic method: (**a**) number of resolved peaks; (**b**) retention time of the last eluting compound; and (**c**) global objective function (D_G) calculated from the normalized responses. The red colors indicate the regions of maximum response, whereas the blue colors correspond to regions with lower response values.

Regarding the final ACN percentage, the separation capacity generally increases at lower ACN values, reaches a maximum at intermediate percentages, and then decreases at higher ACN percentages due to faster elution, which leads to peak overlapping (Figure 2a). The effect of gradient time shows a similar pattern: initially, the number of detected peaks increases because longer gradients reduce coelution and improve peak resolution. At longer gradient times, a moderate increase may still be observed, but in some cases the number

of peaks decreases because excessively broadened peaks (especially those from minor components) may be confounded with the baseline noise. In summary, as illustrated in the figure, optimal separations are obtained at intermediate ACN percentage and intermediate gradient durations.

The effect of these factors on the total analysis time is more straightforward: higher final ACN percentage and shorter gradient durations accelerate the elution of the strongest retained analytes, and vice versa. Therefore, from this perspective, the fastest separation is achieved using a high ACN percentage and short gradients (Figure 2b).

Since the optimal regions identified for separation capacity and analysis time do not coincide, a compromise solution was sought by balancing both objectives through a combined response calculated as the geometric mean of the individual responses after normalization. To this end, desirability functions were used to transform the experimental results into values ranging from 0 to 1, where 0 represents an unacceptable outcome and 1 corresponds to the optimal response. The definition of these normalization thresholds was as follows: 25 chromatographic peaks were chosen as the lower acceptable limit (desirability = 0) and 60 peaks as the maximum (desirability = 1). Similarly, an analysis time of 3 min was set as the most favorable condition (desirability = 1), whereas 10 min was considered excessively long (desirability = 0). The overall response was then calculated from these two normalized measures, assigning a greater weight to separation performance, as this factor was expected to have a stronger impact on the information content available for subsequent chemometric analysis:

$$D_G = (d_p{}^2 \times d_t)^{1/3}$$

with D_G representing the global response, and d_p and d_t the normalized responses for the number of peaks and the analysis time, respectively. The term d_p is squared to increase the weight of this response in the equation relative to d_t. The results of this combined criterion are shown in Figure 2c, from which the compromise solution, based on the maximization of D_G, corresponds to 20% of the final ACN% and 7 min of gradient duration. These conditions were therefore adopted as the final method, as described in Section 2.4.

As an example, Figure 3 presents the chromatograms of three wine samples analyzed under the selected separation conditions. Several peaks corresponding to compounds particularly abundant in these samples have been highlighted.

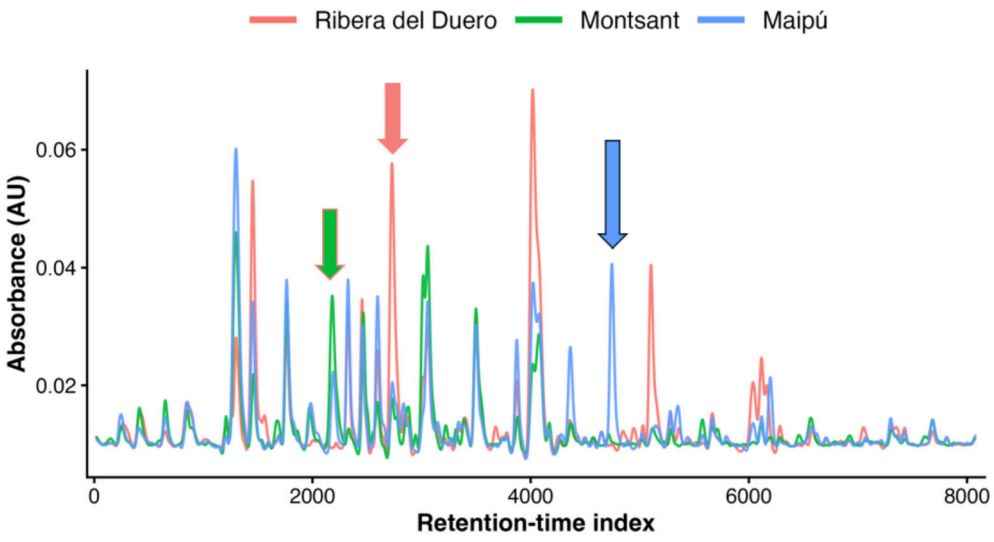

Figure 3. Comparison of chromatographic fingerprints of wine samples from different PDOs. Arrows indicate examples of potentially discriminant peaks contributing to the differentiation of samples from Ribera del Duero, Montsant, and the "Others" (with a representative wine from Maipú).

3.2. Wine Exploration by PCA

Among the different chemometric data preprocessing tested (including no preprocessing, mean-centering, autoscaling, and first derivative), autoscaling proved to be the most effective for maximizing sample discrimination according to PDO and grape variety. A preliminary study was also conducted to assess the discriminatory performance of different temporal ranges (0–1000, 1000–4200, and 4200–8100 channels) within the selected working window, between 0.76 and 7.5 min. The results indicated that these narrower regions offered no additional advantages over the broader interval and, in some cases, the models even resulted in poorer classification. Consequently, the complete chromatographic fingerprint comprising 8100 channels was selected for the subsequent multivariate classification models.

Under these circumstances, the PCA score plots, simplified by considering only the major classes for clearer visualization, show a reasonable distribution of wine samples according to PDO (see Figure 4). In the Figure, the QC clustering in a compact group at the center of the model confirms the reproducibility and stability of the analytical data throughout the measurement sequence. This behavior can also be clearly observed in the comparison of QC chromatograms recorded from the beginning to the end of the analytical series (see Figure S1 in the Supplementary Material, which depicts the chromatographic fingerprints of representative QC samples acquired at injections 1, 277, 373, and 602 across the entire analytical sequence). Excellent peak alignment and highly consistent signal intensities across the entire sequence are observed. The projection of PC2 versus PC3 reveals informative trends on wine distribution. Somontano samples are distinguished from the remaining classes mainly along PC2 and, despite some overlap, are generally concentrated on the right-hand side of the plot. The other classes cluster predominantly on the left-hand side, with Montsant samples showing a slight predominance in the upper central region and Ribera del Duero samples appearing mainly in the lower left region. Rioja wines are more widely dispersed across both areas. Overall, the chromatographic fingerprints contain information related to the geographical origin of the wines. However, the compositional complexity resulting from the combined influence of multiple factors suggests that subsequent classification and authentication studies may be challenging, potentially limiting model performance and reliability.

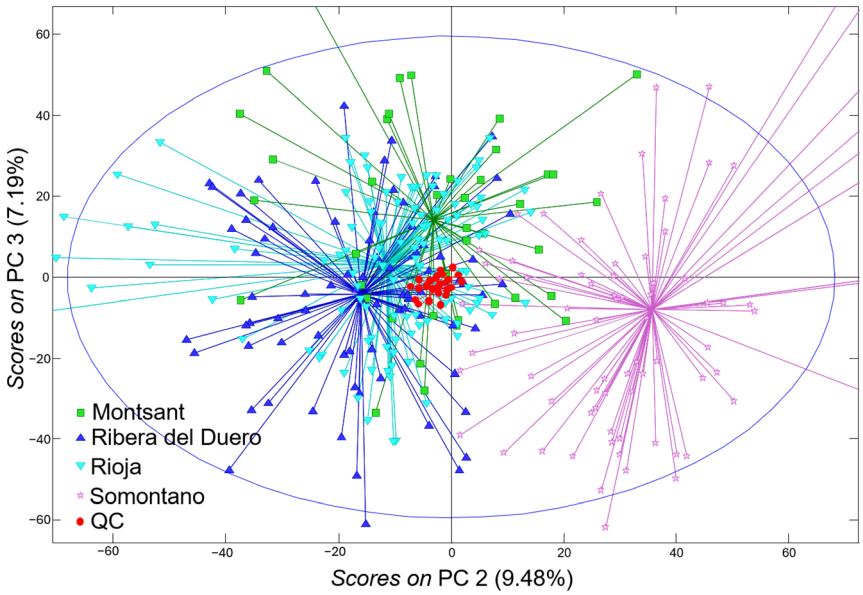

Figure 4. PCA score plot of PC2 vs. PC3 of a set of wine samples labelled according to PDO.

Figure S2 in the Supplementary Material displays two additional score plots, PC1 vs. PC2 and PC2 vs. PC4, providing further insight into the distribution and clustering of the different classes within these coordinate spaces. Although substantial overlap among classes is observed in all cases, these projections facilitate the visualization of differences between certain groups, which are less apparent in Figure 4, particularly for the differentiation between Ribera del Duero and Rioja wines.

An analogous study was performed to evaluate the potential of the chromatographic profiles for differentiating wine samples according to grape variety. While the predominant phenolic compounds are generally shared among varieties, differences in their relative abundances and distribution patterns may constitute a valuable source of discriminative information useful for characterization and classification purposes. To facilitate the interpretation of the main classes considered, namely Tempranillo, Garnacha, and Syrah, other minor monovarietal wines and blended wines were excluded from the dataset.

The PCA results obtained are shown in Figure S3 of the Supplementary Material. Interesting patterns were observed in the PC2 versus PC3 score plot. Along PC2, from left to right, samples belonging predominantly to the Garnacha, Syrah, and Tempranillo groups were distributed in this order. In addition, PC3 positioned most Syrah wines in the upper part of the plot. As in the previous case, these results indicate a relationship between the chromatographic fingerprints and the varietal characteristics of the wines.

It is well established that the composition of wine constituents, particularly phenolic compounds, evolves during aging as a result of a series of chemical and biochemical transformations. Consequently, exploratory PCA may be useful for investigating whether aging-related patterns can be detected from the analytical data. In our case, however, the global PCA model built using all wine samples did not reveal any clear clustering according to ageing. Hence, the score plots obtained from the different PC projections showed that young, semi-aged, Crianza, and Reserva wines occupied the same regions of the multivariate space. This finding can be attributed to the stronger influence of the factors previously examined, namely geographical origin and grape variety, on the compositional profiles of the samples. As a result, these factors account for most of the variance captured by the PCA model, potentially masking any discrimination related to wine aging. To highlight the potential effect of aging on the recorded fingerprints, a stratification approach was applied, focusing on a wine subset in which both PDO and grape variety were fixed. For this purpose, Ribera del Duero wines produced from the Tempranillo variety were selected, as this was the largest group available with these characteristics. The results shown in Figure S4 (subplot S4a) of the Supplementary Material provide a clearer visualization of this behavior. Despite the substantial overlap between classes, a discernible trend can be observed. Young wines are predominantly located on the right-hand side of the score plot, while Reserva wines subjected to longer aging periods are mainly concentrated on the left-hand side. Crianza wines, with intermediate aging times, are generally distributed in the central area of the plot.

Stratification according to a given aging category may also facilitate the visualization of other underlying trends. For instance, the score plot shown in Figure S4b of the Supplementary Material, corresponding to the subset of Tempranillo Crianza wines, provides a clearer view of the distribution of samples according to their PDO. In particular, a pronounced separation between wines from Ribera del Duero and Rioja can be observed, primarily along PC1. This distinction is considerably more evident than that observed in Figure 4.

3.3. Supervised Wine Classification and Authentication

In this section, wine classification according to PDOs and grape varieties is addressed using PLS-DA and SIMCA. As a preliminary step, the available wine set was divided into the calibration and prediction subsets, accounting for 70% and 30% of the samples, respectively. As indicated in Section 2.4, the data split was performed at the sample level; therefore, all replicates of each wine belong either to the calibration set or to the prediction set. The QCs were not considered here. Regarding the X-matrix of experimental data, as for PCA, chromatographic fingerprints within the working window between 0.76 and 7.5 min were considered.

Binary PLS-DA classification models (target class vs. non-target class) and multiclass models for the target classes were developed, both incorporating an additional "Others" class that groups all remaining wines. The objective of this category was not to define a coherent oenological group, but rather to represent wines outside the PDOs (or grape varieties) of interest. In practical terms, "Others" acts as a catch-all category to which wines not belonging to the target classes can be assigned.

It should be noted that in all cases the models were constructed using information from both the target class(es) and the samples outside those main classes. Hence, the increased within-class variability of the "Others" category certainly influences the model construction and classification performance of PLS-DA; at the same time, "Others" provide a certain degree of generality to the approach, as it acknowledges that a study may contain unknown wines that do not belong to any of the target classes; however, this generality is inherently limited by the variability represented within the available out-of-class samples. As an additional remark, the final class assignment of each sample was determined by the highest probability score, forcing every sample to be assigned to a given class and preventing multiple class assignments.

The selection of the number of latent variables (LVs) in the PLS-DA models was carried out by minimizing the mean classification error obtained by Venetian Blinds cross-validation, while avoiding data overfitting. Different splits and blind thicknesses were tested to verify that the optimal values were similar in all cases and that no remarkable discrepancies emerged that could indicate potential issues or overfitting.

Table 1 presents the cross-validation and test-set prediction results obtained for the PLS-DA classification models developed for the four representative PDOs. Overall, the cross-validation performance metrics were highly satisfactory, with the models requiring only a moderate number of latent variables (4–6). This finding is consistent with the underlying complexity of the dataset and the multifactorial relationships embedded within the chromatographic fingerprints.

Cross-validation PLS-DA results obtained for the calibration set are generally quite satisfactory; those derived from binary models consistently appear to outperform the multiclass approaches. However, when the comparison is based on the prediction stage, the resulting performance metrics are largely equivalent, with neither strategy clearly emerging as superior.

It is worth noting that the sensitivity values for PDO classification are consistently lower than the specificity counterparts. This may be attributed to the relatively limited number of samples belonging to the target classes within the calibration set. This is the case, for example, for the Montsant PDO (sensitivity = 0.50). From this perspective, classes that are better represented in the dataset, such as the Ribera del Duero PDO, achieve somewhat higher sensitivity values.

By comparison, the number of samples included in the "Others" class is considerably larger and, despite the inherent heterogeneity of this group, specificity values remain above 0.90 for PDO classification.

Table 1. Cross-validation and prediction classification performance of one-vs-rest binary and multiclass PLS-DA models applied to the four representative PDOs.

PDO	Model Parameters	Binary One-vs.-Rest	Multiclass	Binary One-vs.-Rest	Multiclass
		Cross-Validation		External Prediction	
Montsant	LVs	4	6	4	6
	Sensibility	1.0	0.528	0.500	0.5
	Specificity	0.929	0.955	0.902	0.914
	Error (%)	6.5	8.2	11.1	10.0
Ribera del Duero	LVs	5	6	5	6
	Sensibility	1.0	0.905	0.778	0.611
	Specificity	0.983	0.963	0.944	0.963
	Error (%)	1.4	4.6	7.2	7.2
Rioja	LVs	5	6	5	6
	Sensibility	0.973	0.853	0.595	0.595
	Specificity	0.959	0.944	0.942	0.971
	Error (%)	3.9	7.2	13.9	11.7
Somontano	LVs	5	6	5	6
	Sensibility	1.0	0.854	0.9	0.9
	Specificity	0.983	0.951	0.980	0.973
	Error (%)	1.4	6.0	3.3	3.9

Figure 5 shows a representative classification plot for the Somontano PDO obtained using a binary PLS-DA model. On the left, the cross-validation assignments of the wines in the calibration set are displayed, whereas the samples in the prediction set are shown on the right, separated by the vertical maroon dashed line. The horizontal dashed line at a score of approximately 0.31 represents the class-separation threshold, corresponding to the optimal balance between sensitivity and specificity. Accordingly, wines with scores above this threshold are assigned to the Somontano PDO, whereas those with lower scores are not recognized as belonging to this class. Overall, both cross-validation and prediction results demonstrate a high classification performance, with most Somontano wines being correctly identified and only a small number of wines from the "other" class, which includes wines from multiple origins, being incorrectly assigned to the Somontano class.

To evaluate whether the presence of highly heterogeneous wines within the "Others" category had a negative impact on prediction performance, additional PLS-DA studies were carried out considering only Spanish wines with a declared PDO. Consequently, 10 international wines and 11 Spanish wines without a PDO label were excluded from the dataset. The results obtained yielded sensitivity and specificity values very similar to those reported in Table 1, indicating that, in this case, the simplification of the sample diversity did not significantly affect the predictive performance of the models.

A parallel study was carried out, using the same modelling strategies, to classify wines according to the main grape varieties represented in the available sample set. In this case, the classes considered were Tempranillo, Garnacha, and Syrah, while all remaining monovarietal wines and blends involving these or other varieties were grouped into the "other" class. Table 2 summarizes the outcomes for both PLS-DA binary and multiclass models in the cross-validation and prediction phases.

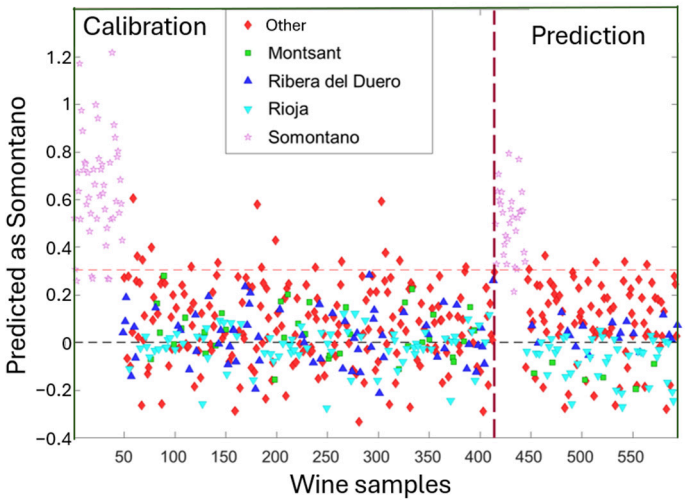

Figure 5. Classification plot for the Somontano PDO obtained from a binary PLS-DA model. Cross-validation assignments for the calibration set are shown on the left, and prediction samples on the right. The horizontal dashed line represents the class-separation threshold. The red dashed vertical line indicates the separation between the calibration sample set and the prediction sample set.

Table 2. Cross-validation and prediction classification performance of one-vs-rest binary and multiclass PLS-DA models applied to the three representative grape varieties.

Variety	Model Parameters	Binary One-vs.-Rest	Multiclass	Binary One-vs.-Rest	Multiclass
		Cross-Validation		External Prediction	
Garnacha	LVs	5	6	5	6
	Sensibility	0.909	0.667	0.481	0.556
	Specificity	0.945	0.929	0.850	0.869
	Error (%)	5.8	9.2	20.6	17.8
Syrah	LVs	6	6	6	6
	Sensibility	0.952	0.762	1.000	0.500
	Specificity	0.980	0.969	0.920	0.966
	Error (%)	2.2	4.1	7.8	5.0
Tempranillo	LVs	5	6	5	6
	Sensibility	0.889	0.889	0.833	0.929
	Specificity	0.865	0.845	0.841	0.928
	Error (%)	12.8	14.3	16.1	7.2

For varietal classification, the performance metrics should be interpreted in the context of the unequal representation of the different classes. Lower sensitivity values are mainly associated with varieties represented by a limited number of samples in the prediction set. For example, Garnacha exhibited a sensitivity of 0.481 in the binary classification model. In contrast, the Tempranillo class, which was represented by a substantially larger number of samples, achieved sensitivity values above 0.8. The apparently optimal result obtained for the Syrah class should be interpreted with caution, as it is based on only two wines available for prediction. Thus, it most likely reflects the correct classification of these few samples rather than a genuinely superior model performance. Nevertheless, specificity values remained consistently high across all classes. Overall, increasing the number of samples per variety would likely improve the robustness, reliability, and generalizability of the classification models.

As an illustrative example, Figure 6 presents the classification plot for the Tempranillo class obtained using a binary PLS-DA model. Nearly all Garnacha and Syrah wines are correctly recognized as non-Tempranillo samples. However, several misclassifications are observed, both during cross-validation and prediction, particularly for blended wines in which Tempranillo is the predominant grape variety.

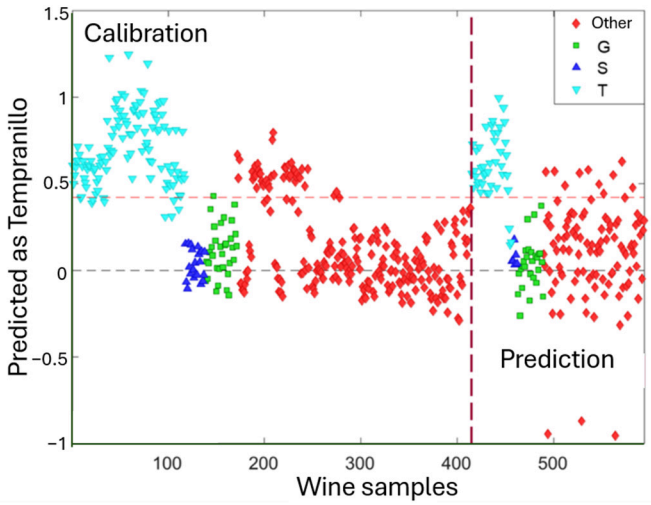

Figure 6. Classification plot for the Tempranillo monovarietal class obtained from a binary PLS-DA model. Cross-validation assignments for the calibration set are shown on the left, and prediction samples on the right. The horizontal dashed red line is the class-separation threshold. The red dashed vertical line indicates the separation between the calibration sample set and the prediction sample set.

SIMCA was employed as the preferred strategy for class-specific authentication (one-class modelling), using only information from the target class to construct the calibration model. The resulting model was subsequently applied to all remaining samples to assess whether they belonged to the target class. In this context, one-class methods such as SIMCA are particularly well suited for authentication purposes, as they evaluate whether an unknown sample conforms to the characteristic patterns of the reference class [3].

Individual one-class models were developed for each target PDO and grape variety by constructing PCA models exclusively from wines belonging to the target class within the calibration subset. The optimal number of PCs was selected by cross-validation, carried out on the full calibration set, including both target-class and out-of-class samples. As the statistical threshold for sample assignment, a combined criterion between Hotelling's T^2 and Q-residuals was used, according to the expression $(Q^2 + (T^2)^2)^{1/2}$, set equal to 0.95.

As performance metrics for sample authentication, the prediction error was calculated as the proportion of misclassified samples relative to the total number of samples. In addition, the ability of the models to correctly identify samples falling within the rejection region (i.e., with combined T^2 and Q scores > 0.95) was evaluated.

The results summarized in Table 3 reveal a reasonable authentication performance for the prediction set, with misclassification errors ranging from 5.0% to 16.7% for PDOs and from 3.3% to 18.3% for grape varieties. However, beyond these overall performance figures, a more detailed data examination reveals some weaker class-specific results. These appear to be mainly attributable to the unbalanced representation of the groups, rather than to limitations in the nature or discriminative power of the analytical fingerprints themselves.

For instance, the prediction sensitivity for the PDO Montsant class is 0.000. This result arises because only two wines from this PDO were included in the prediction set, substantially increasing the likelihood of obtaining poor performance metrics. In contrast, the specificity is very high (0.983), since the number of wines in the prediction set that

do not belong to PDO Montsant is relatively large (57 wines); consequently, the impact of an occasional misclassification on the overall percentage metric is much smaller. For PDO Ribera del Duero, a more reasonable sensitivity value (0.833) is obtained, reflecting the larger number of wines available in the prediction set (6 wines). The corresponding specificity is also satisfactory (0.833).

Table 3. Cross-validation and prediction results for single-class SIMCA models to classify representative PDOs and grape varieties.

Class	Model Parameters	Cross-Validation	External Prediction
Montsant	PCs	8	8
	Sensibility	1.000	0.000
	Specificity	0.974	0.983
	Error (%)	2.4	5.0
Ribera del Duero	PCs	6	6
	Sensibility	1.000	0.833
	Specificity	0.792	0.833
	Error (%)	17.6	16.7
Rioja	PCs	9	9
	Sensibility	0.947	0.476
	Specificity	0.959	0.978
	Error (%)	4.3	13.9
Somontano	PCs	9	9
	Sensibility	1.000	0.133
	Specificity	0.951	0.973
	Error (%)	4.3	16.7
Garnacha	PCs	8	8
	Sensibility	1.000	0.074
	Specificity	0.958	0.967
	Error (%)	3.9	16.7
Syrah	PCs	8	8
	Sensibility	1.000	0.000
	Specificity	1.000	1.000
	Error (%)	0.0	3.3
Tempranillo	PCs	8	8
	Sensibility	0.949	0.691
	Specificity	0.951	0.855
	Error (%)	4.3	18.3

Overall, the sensitivity values in the prediction stage are not optimal, most likely owing to the limited number of samples available for some of the PDOs being predicted. By contrast, specificity values are consistently much higher because a considerably larger number of samples are available for recognition as non-members of each class.

A similar situation is observed in the varietal classification studies, where the Garnacha and Syrah classes are represented by only a small number of wines in the prediction set. In contrast, the sensitivity towards Tempranillo is substantially higher (0.691), possibly due to the greater number of available samples. In all cases, specificity values remain good for the same reason as above, namely the larger number of samples that do not belong to the class under consideration.

To further investigate the discriminatory information contained in the chromatographic data and assess the potential contribution of nonlinear relationships, SVM models with a radial basis function (RBF) kernel were also evaluated as a complementary super-

vised classification approach. The SVM models were optimized by internal cross-validation using the calibration subset and subsequently evaluated on the prediction subset. The reported values of C and γ correspond to the arithmetic means of the values selected during the 100 repeated internal grouped cross-validation procedures. The corresponding cross-validation and external prediction results obtained from the SVM models are summarized in Table 4.

Table 4. Cross-validation and prediction classification performance of SVM models using a radial basis function (RBF) kernel for the classification of representative PDOs and grape varieties.

Class	Model Parameters	Cross-Validation	External Prediction
Montsant	Cost (C)	26.07	26.07
	Gamma (γ)	1.80×10^{-4}	1.80×10^{-4}
	Sensitivity	0.697	0.542
	Specificity	0.418	0.331
	Error (%)	56.4	65.2
Ribera del Duero	Cost (C)	14.09	14.09
	Gamma (γ)	6.80×10^{-5}	6.80×10^{-5}
	Sensitivity	0.991	0.908
	Specificity	0.015	0.015
	Error (%)	85.6	85.6
Rioja	Cost (C)	98.00	98.00
	Gamma (γ)	3.46×10^{-4}	3.46×10^{-4}
	Sensitivity	0.809	0.789
	Specificity	0.844	0.826
	Error (%)	16.3	18.1
Somontano	Cost (C)	10.09	10.09
	Gamma (γ)	1.63×10^{-4}	1.63×10^{-4}
	Sensitivity	0.576	0.471
	Specificity	0.467	0.451
	Error (%)	51.9	54.6
Garnacha	Cost (C)	11.87	11.87
	Gamma (γ)	3.24×10^{-4}	3.24×10^{-4}
	Sensitivity	0.621	0.547
	Specificity	0.414	0.406
	Error (%)	56.5	58.0
Syrah	Cost (C)	18.69	18.69
	Gamma (γ)	3.80×10^{-4}	3.80×10^{-4}
	Sensitivity	0.678	0.373
	Specificity	0.506	0.548
	Error (%)	48.6	46.1
Tempranillo	Cost (C)	50.91	50.91
	Gamma (γ)	5.67×10^{-4}	5.67×10^{-4}
	Sensitivity	0.905	0.901
	Specificity	0.794	0.766
	Error (%)	17.6	19.8

4. Discussion

Although this strategy relies on untargeted analysis and is not intended for compound profiling, the chemical information captured by the chromatographic fingerprints is primarily associated with phenolic compounds. Detection at 280 nm records the contributions of the most relevant phenolic compounds present in red wines, including hydroxybenzoic

and hydroxycinnamic acids, flavanols, and several minor phenolic families. Accordingly, chromatograms were extracted at 280 nm, providing a suitable compromise between methodological simplicity, compositional coverage and analytical performance. In this regard, although the software developed for data extraction and processing (see Section 2.5) allows chromatographic information to be extracted at multiple wavelengths (e.g., 320 and 370 nm are particularly useful for hydroxycinnamic acids and flavonols, respectively), we consider 280 nm as the most balanced option in terms of instrumental sensitivity and overall informational content.

The selected working conditions yielded highly satisfactory chromatograms, characterized by rich peak patterns and a total run time of 10 min. Such analytical throughput is particularly advantageous for the analysis of large sample sets within a reasonable time frame, as demonstrated in the present study, where the complete chromatographic sequence, comprising more than 600 injections, was completed within 108 h.

Representative chromatograms shown in Figure 3 illustrate the high quality of the separation, with several well-resolved peak clusters reflecting the occurrence of compounds detectable at 280 nm. These chromatographic fingerprints provide the basis for the subsequent characterization and authentication studies. From the perspective of characterization potential, the overall distribution of compounds in selected wines from Ribera del Duero, Montsant, and "Others" origins is broadly similar. Nevertheless, certain compounds display higher relative abundances and, in some cases, appear to be almost specific to a particular class. Such features may constitute valuable markers for the subsequent characterization and classification stages.

Additionally, the preprocessing workflow ensures excellent peak alignment while effectively reducing noise and correcting baseline drift through a fully integrated and automated procedure. Consequently, the high data quality, together with the excellent chromatographic reproducibility achieved, enables the reliable analysis of multiple wine subsets across different analytical batches. This aspect is particularly relevant in real-world applications, where wine samples submitted for authentication often arrive at control laboratories at different times. Furthermore, this capability facilitates the expansion and updating of calibration models through the incorporation of new samples over time.

PCA enables the unsupervised exploration of samples with the aim of identifying natural groupings and trends without considering class labels such as geographical origin or grape variety. The distribution of samples in the PCA space can provide insights into the most relevant factors or attributes underlying such patterns. This information is highly valuable for assessing whether the data, in this case the chromatographic fingerprints, contain sufficient discriminatory information to support the intended authentication task or, alternatively, whether they fail to provide meaningful class differentiation.

PCA results reveal a certain degree of sample structuring according to geographical origin, grape variety, and other relevant attributes. However, considerable overlap is observed among samples with different characteristics owing to the multifactorial nature of the underlying relationships. The composition of the phenolic fraction, which largely determines the obtained fingerprints, depends on both geographical origin and grape variety. In addition, aging processes can substantially modify the compositional profiles.

It should be noted that many samples originate from geographically close regions and, although they do not belong to the same PDO, they are expected to exhibit similar compositional profiles. Likewise, many wines are blends with relatively complex compositions, which adds an additional layer of complexity to the study. Moreover, wines produced from coupages in which a particular grape variety predominates tend to behave similarly to pure monovarietal wines and therefore occupy similar regions of the PCA space.

Despite the complexity of the dataset, the results indicate that supervised classification and authentication of wines based on chromatographic fingerprints are feasible. However, the strong interdependence among the factors governing wine composition may limit the achievable classification performance.

In PLS-DA, both binary and multiclass models were built including an additional "Others" category to accommodate wines that do not belong to the target classes. By its nature, this category is highly heterogeneous, as it comprises wines from diverse geographical origins and production backgrounds. For instance, in PDO classification, the "Others" class includes wines from regions such as Navarra, Campo de Borja, Penedès, Priorat, and other denominations. Similarly, in varietal classification, it encompasses all monovarietal wines not produced from Tempranillo, Garnacha, or Syrah, as well as a wide range of blended wines.

Such heterogeneity could potentially have a negative impact on prediction performance, since substantially different groups are forced into a single class and may not share common underlying patterns. Nevertheless, from an operational perspective, "Others" is highly useful because it helps define the markers of the target classes and effectively acts as a catch-all group for samples that deviate from the targe patters.

An alternative strategy would be to reject samples based on class assignment confidence, excluding from the target classes those wines that cannot be classified with sufficient certainty. Although this approach was also evaluated, the models including the "Others" category provided superior overall performance. In contrast, confidence-based rejection resulted in an excessively high proportion of unassigned samples, reducing the practical applicability of the classification models. As for the classification models, we first discuss the results for PDOs. Cross-validation figures, including sensitivity, specificity, and prediction error, are slightly better for the binary models, possibly because these models can be more specifically adjusted to deal with each individual target class, which particularly enhances sensitivity. Regarding external prediction, the quality metrics are lower, consistent with predictions on independent samples not used in the calibration step. Sensitivities are affected by the relatively small number of class samples available for this independent prediction step (so a single misclassified sample can markedly worsen this metric), whereas specificity remains very high due to the larger number of out-of-class samples.

The classification error rates, calculated as the proportion of misclassified samples relative to the total number of samples, range from 3.3% to 14%. No substantial differences are observed between the binary and multiclass strategies. Although the binary approach shows slightly better performance, the multiclass strategy is more straightforward to implement. Nevertheless, binary models can be recommended because of the greater ease with which their results can be interpreted.

The results in Table 1 indicate that the Somontano PDO achieves the best prediction performance, likely because the climatic conditions and winemaking practices in this region are particularly distinctive. The Montsant PDO also exhibits characteristic features, although its figures are somewhat lower, probably due to the more limited representativeness of the corresponding sample subset. In contrast, the Ribera del Duero and Rioja PDOs are represented by sufficiently large datasets, but some confusions between these regions may occur because both PDOs use predominantly Tempranillo as their principal grape variety, leading to greater similarity in their compositional fingerprints.

Shifting the discussion to grape variety modeling (see Table 2), cross-validation results are slightly better for the binary models than for the multiclass ones, mainly due to their higher sensitivity in identifying in-class samples. At this stage, the Tempranillo class shows the least satisfactory performance metrics, likely due to confusion with certain multivarietal wines in which Tempranillo represents a large proportion of the blend.

The results for the prediction subsets are more limited, with the most notable issue being the low sensitivity in identifying the Syrah class; this can be attributed to the small number of available wines for this variety. On a more positive note, the estimates for the Garnacha class are quite good, likely due to the more distinctive fingerprints associated with this variety.

To further assess the robustness of the PLS-DA classification results, repeated external validation was performed across 100 data partitions. Overall, the classification performance remained relatively consistent across the repeated validations, although greater variability was observed for some classes, particularly those with limited sample representation. This effect was most evident for the grape varieties, especially Syrah and Garnacha, for which sensitivity showed greater variability across repeated partitions. Bootstrap analysis further illustrated the uncertainty associated with the different classes, with wider intervals observed for several classes with limited sample representation. The complete results of the repeated validation and bootstrap analyses are provided in Supplementary Material (Table S2 and Figures S5 and S6).

SIMCA results should be interpreted with caution, particularly for classes with limited sample representation (see Table 3). Although specificity values were generally high, indicating a good ability to reject wines outside the target class, the low prediction sensitivity observed for several classes indicates that reliable authentication capability cannot be demonstrated under the current sample representation. In particular, the prediction sensitivities obtained for Montsant and Syrah were 0.000, while those for Somontano and Garnacha were 0.133 and 0.074, respectively. Therefore, these models cannot be considered effective authentication tools for these classes at this stage. Because these sensitivity estimates are based on a single external prediction split, they should be regarded as exploratory and potentially highly unstable, particularly for classes represented by very few target samples. The low sensitivity values are strongly influenced by the limited number of representative samples available in the prediction set, whereas the high specificity values are partly a consequence of the larger number of out-of-class samples available for evaluation. Accordingly, the present results should be regarded as an indication of the potential of the chromatographic fingerprints for class-specific authentication rather than as evidence of reliable authentication capability. Larger and more representative datasets would be required to properly assess the robustness and generalizability of the SIMCA models.

SVM models were subsequently evaluated to investigate whether nonlinear relationships within the chromatographic fingerprints could improve classification performance. Initially, linear SVM models were explored; however, these models provided unsatisfactory results and were therefore not considered further. Nonlinear SVM models using an RBF kernel were subsequently evaluated (Table 4). The results showed variable performance among the different classes. For PDOs, prediction errors ranged from 18.1% for Rioja to 85.6% for Ribera del Duero, whereas for grape varieties they ranged from 19.8% for Tempranillo to 58.0% for Garnacha. Tempranillo showed the best performance among the grape varieties, with a prediction sensitivity of 0.901, specificity of 0.766, and error of 19.8%. In contrast, Garnacha and Syrah showed substantially higher prediction errors, of 58.0% and 46.1%, respectively. Overall, the nonlinear SVM models did not provide a consistent improvement over the PLS-DA models, indicating that, in these particular cases, the introduction of nonlinear relationships did not substantially increase the discriminatory capability of the chromatographic fingerprints.

5. Conclusions

The UHPLC-UV methodology developed generates highly information-rich fingerprints suitable for the effective study of wine attributes such as geographical origin and

varietal composition. Although it is a purely untargeted approach, the resulting chromatograms correspond predominantly to the polyphenolic fraction, mainly represented by hydroxybenzoic and hydroxycinnamic acids, as well as flavonoids. We believe that this methodology is efficient in terms of speed, robustness, and the informative value of the chromatographic fingerprints generated. In our case, it should be noted that the measurements are relatively fast and, above all, extremely stable and reproducible over time and across different analytical sessions.

The chromatographic fingerprints are characterized by the presence of numerous peaks which, although not specific to a given sample type, do exhibit cross-selectivity, with some signals being overexpressed in certain classes. This diversity in feature intensities forms the basis for the characterization and authentication studies presented here. In this regard, the methodology is well positioned to deal with real-world scenarios, particularly because it is fully compatible with the analysis of different series or batches, as well as with the gradual incorporation of new calibration samples to improve the representativeness and performance of the models.

PCA reveals some distributional patterns associated with variety or geographical origin. Nevertheless, class overlap is substantial, as the typologies do not form compact groups but instead occupy broad regions in the score plots due to concomitant multifactorial dependencies. The PLS-DA classification models showed variable performance among the different PDO and varietal classes under the heterogeneous vintage and aging conditions represented in the dataset. Specificity was generally high, whereas sensitivity varied substantially among classes. In some cases, lower sensitivity values were obtained, which can be attributed to the limited number of available samples for those classes; this is therefore an aspect that could be improved by increasing class representativeness. On the other hand, the SIMCA models showed their ability to reject wines outside the target class, although their ability to recognize authentic samples was highly variable among classes and was insufficient for some classes under the current sample representation. Nonlinear SVM models were also evaluated to assess whether nonlinear relationships could provide additional discriminatory information. However, their performance was variable among classes and did not show a consistent improvement over PLS-DA, suggesting that the introduction of nonlinear relationships did not substantially enhance the discriminatory capability of the chromatographic fingerprints.

The chemometric results could be further improved if broader and more representative calibration sets were available for each class of interest. However, the current wine dataset is unbalanced for several classes, and vintage and aging categories are also unevenly distributed among PDOs. Therefore, part of the observed classification performance may be influenced by these factors and cannot be unequivocally attributed to geographical origin. Validation using more balanced datasets, with comparable vintage and aging distributions across classes, would be valuable to further assess the robustness and generalizability of the classification models. The scope of the study The PLS could also be extended to other wine classes or attributes through the proper design of representative calibration sets.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/separations13100296/s1. Table S1. List of red wines analyzed. Table S2. Summary of PLS-DA performance across repeated external validation procedures for the evaluation of the uncertainty in the predictions of PLS-DA models. Figure S1. Chromatographic fingerprints of representative quality control (QC) samples acquired at injections 1, 277, 373, and 602 throughout the analytical sequence, illustrating the stability and reproducibility of the analytical response over the course of the study. Figure S2. PCA score plot of PC1 vs. PC2 (a) and PC2 vs. PC4 (b) for a set of wine samples labelled according to PDO. Figure S3. PCA score plot of PC2 vs. PC3 for a set of wine samples labelled according to grape variety. Sample assignation: Blue,

Tempranillo; Fuchsia, Garnacha; Grey, Syrah. Figure S4. Supplementary score plots. (a) Evaluation of the effect of aging on the distribution of wines from the Tempranillo Ribera del Duero subset. (b) Evaluation of the effect of PDO on the distribution of wines from the Tempranillo Crianza subset. Figure S5. Repeated external validation performance of PLS-DA models. Distribution of balanced accuracy obtained across 100 repeated external validation procedures for binary and multiclass PLS-DA models. Figure S6. Bootstrap uncertainty assessment of PLS-DA classification performance. Mean balanced accuracy and corresponding bootstrap 95% confidence intervals obtained for each target class from the repeated external validation procedures. G, Garnacha; S, Syrah; T, Tempranillo.

Author Contributions: Conceptualization, J.C.S., S.S. and J.S.; methodology, J.C.S., D.A.A.F., S.S. and J.S.; formal analysis, J.C.S., S.S. and J.S.; investigation, J.C.S.; resources, D.A.A.F., S.S. and J.S.; data curation, J.C.S. and D.A.A.F.; writing—original draft preparation, J.C.S., S.S. and J.S.; writing—review and editing, S.S. and J.S.; supervision, S.S. and J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Agencia Estatal de Investigación del Ministerio de Ciencia, Innovación y Universidades MICIU/AEI/10.13039/501100011033 (ref. PID2023-147160OB-C22) and the INSA-UB María de Maeztu Unit of Excellence (Grant CEX2021-001234-M) funded by MICIU/AEI and ERDF EU.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We gratefully acknowledge our colleagues from the Department of Chemical Engineering and Analytical Chemistry at the University of Barcelona (M. Granados, A. de Juan, O. Nuñez, J.M. Díaz-Cruz, among others), as well as our friends, for their assistance in collecting wine samples from diverse origins.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACN	Acetonitrile
HPLC	High-Performance Liquid Chromatography
HRMS	High-Resolution Mass Spectrometry
IM	Ion Mobility
LC	Liquid Chromatography
LV	Latent Variable
MS	Mass Spectrometry
PC	Principal Component
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares-Discriminant Analysis
PDO	Protected Designation of Origin
QC	Quality Control
QTOF	Quadrupole-Time of Flight
SIMCA	Soft Independent Modelling of Class Analogies
UHPLC	Ultra-High-Performance Liquid Chromatography
UV	Ultraviolet detection

References

1. Jackson, R.S. Introduction. In *Wine Science*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2014; pp. 1–19. [CrossRef]
2. Moehring, M.J.; de B. Harrington, P. Analysis of Wine and Its Use in Tracing the Origin of Grape Cultivation. *Crit. Rev. Anal. Chem.* **2022**, *52*, 1901–1912. [CrossRef] [PubMed]

3. Koljančić, N.; Furdíková, K.; de Araújo Gomes, A.; Špánik, I. Wine Authentication: Current Progress and State of the Art. *Trends Food Sci. Technol.* **2024**, *150*, 104598. [CrossRef]
4. OIV. State of the World Wine Sector in 2025. 2026. Available online: https://www.oiv.int/sites/default/files/2026-05/OIV-State_of_the_World_Wine_Sector_in_2025_0.pdf (accessed on 26 June 2026).
5. Morales, L.V.; Ahumada, D.A.F.; Deben, A.; Sentellas, S.; Saurina, J. Revisiting Wine Authentication in a Changing Landscape: Trends, Challenges, and Future Directions. *Trends Anal. Chem.* **2026**, *202*, 118917. [CrossRef]
6. Elumalai, A.; Natarajan, V. Advancements in Analytical Technologies for Ensuring Food Quality and Authentication: A Comprehensive Review. *J. Food Compos. Anal.* **2025**, *139*, 107075. [CrossRef]
7. Navarro-Abril, A.; Saurina, J.; Sentellas, S. Simultaneous Determination of Amino Acids and Biogenic Amines by Liquid Chromatography Coupled to Mass Spectrometry for Assessing Wine Quality. *Beverages* **2022**, *8*, 69. [CrossRef]
8. Leleu, G.; Butelle, R.; de Moura, P.H.B.; Valls-Fonayet, J.; Da Costa, G.; Richard, T. Wine Authenticity Controls: Analytical Methods, Data Integration and Emerging Perspectives. *Food Anal. Methods* **2026**, *19*, 98. [CrossRef]
9. Qiang, H.Q.; Gu, H.W.; Yin, X.L.; Meng, Z.; Long, W.; Fu, H.; She, Y. A Comprehensive Review on Wine Authentication and Traceability: Current Issues, Solutions and Future Perspectives. *Food Chem.* **2026**, *508*, 148456. [CrossRef] [PubMed]
10. Tsagkaris, A.S.; Kalogiouri, N.; Tokarova, V.; Hajslova, J. The Impact of Spectral Data Pre-Processing on the Assessment of Red Wine Vintage through Spectroscopic Methods. *J. Sci. Food Agric.* **2025**, *105*, 5986–5998. [CrossRef] [PubMed]
11. González-Domínguez, R.; Sayago, A.; Fernández-Recamales, Á. An Overview on the Application of Chemometrics Tools in Food Authenticity and Traceability. *Foods* **2022**, *11*, 3940. [CrossRef] [PubMed]
12. Sentellas, S.; Saurina, J. The Role of Chemometrics in Food Integrity and Authenticity. In *Chromatographic and Related Separation Techniques in Food Integrity and Authenticity*; Núñez, O., Campmajó, G., Eds.; World Scientific: Singapore, 2021; pp. 167–200. [CrossRef]
13. Rodionova, O.Y.; Oliveri, P.; Malegori, C.; Pomerantsev, A.L. Chemometrics as an Efficient Tool for Food Authentication: Golden Pillars for Building Reliable Models. *Trends Food Sci. Technol.* **2024**, *147*, 104429. [CrossRef]
14. Aleixandre-Tudo, J.L.; du Toit, W. A Chemometric Approach to the Evaluation of the Ageing Ability of Red Wines. *Chemom. Intell. Lab. Syst.* **2020**, *203*, 104067. [CrossRef]
15. Mafata, M.; Brand, J.; Medvedovici, A.; Buica, A. Chemometric and Sensometric Techniques in Enological Data Analysis. *Crit. Rev. Food Sci. Nutr.* **2022**, *63*, 10995–11009. [CrossRef] [PubMed]
16. Castañeda, F.N.; Pellegrino Vidal, R.B.; Aspromonte, J. Untargeted Chromatographic Methods Coupled with Chemometric Strategies for the Analysis of Food and Related Samples. *Trends Anal. Chem.* **2024**, *173*, 117650. [CrossRef]
17. Zhong, P.; Wei, X.; Li, X.; Wei, X.; Wu, S.; Huang, W.; Koidis, A.; Xu, Z.; Lei, H. Untargeted Metabolomics by Liquid Chromatography–Mass Spectrometry for Food Authentication: A Review. *Compr. Rev. Food Sci. Food Saf.* **2022**, *21*, 2455–2488. [CrossRef] [PubMed]
18. Ríos-Reina, R.; Segura-Borrego, M.P.; Camiña, J.M.; Callejón, R.M.; Azcarate, S.M. Multiplatform Spectralprint Strategies for the Authentication of Spanish PDO Fortified Wines Using AHIMBU. *Chemom. Intell. Lab. Syst.* **2025**, *257*, 105311. [CrossRef]
19. Ranaweera, R.K.R.; Capone, D.L.; Bastian, S.E.P.; Cozzolino, D.; Jeffery, D.W. A Review of Wine Authentication Using Spectroscopic Approaches in Combination with Chemometrics. *Molecules* **2021**, *26*, 4334. [CrossRef] [PubMed]
20. Pinto, T.; Cosme, F.; Correia, E.; Vilela, A. Spectroscopic Techniques Application for Wine and Wine Byproduct Authentication. *Appl. Sci.* **2025**, *15*, 4457. [CrossRef]
21. Ranaweera, R.K.R.; Bastian, S.E.P.; Gilmore, A.M.; Capone, D.L.; Jeffery, D.W. A-TEEM with Multiblock Data Analysis and Machine Learning for Intraregional Classification of Barossa Shiraz Wine. *Food Control* **2023**, *144*, 109335. [CrossRef]
22. Le Mao, I.; Da Costa, G.; Bautista, C.; de Revel, G.; Richard, T. Application of ^{1}H NMR Metabolomics to French Sparkling Wines. *Food Control* **2023**, *145*, 109423. [CrossRef]
23. Yin, X.L.; Peng, Z.X.; Pan, Y.; Lv, Y.; Long, W.; Gu, H.W.; Fu, H.; She, Y. UHPLC-QTOF-MS-Based Untargeted Metabolomic Authentication of Chinese Red Wines. *Food Res. Int.* **2024**, *178*, 113923. [CrossRef] [PubMed]
24. Pan, Y.; Gu, H.-W.; Lv, Y.; Yin, X.-L.; Long, Y.; Chen, W.; Fu, H.; She, Y. Untargeted Metabolomic Analysis of Chinese Red Wines for Geographical Origin Traceability. *Food Chem.* **2022**, *377*, 133473. [CrossRef] [PubMed]
25. Wang, Z.; Chen, X.; Liu, Q.; Zhang, L.; Liu, S.; Su, Y.; Ren, Y.; Yuan, C. Untargeted Metabolomics Analysis for Discriminating Geographical Origin and Vintage of Chinese Red Wine. *Food Res. Int.* **2023**, *165*, 112547. [CrossRef] [PubMed]
26. Uttl, L.; Bechynska, K.; Ehlers, M.; Kadlec, V.; Navratilova, K.; Dzuman, Z.; Fauhl-Hassek, C.; Hajslova, J. Critical Assessment of Chemometric Models for Varietal Authentication of Wine. *Food Control* **2023**, *147*, 109336. [CrossRef]
27. Díaz, R.; Gallart-Ayala, H.; Sancho, J.V.; Nuñez, O.; Zamora, T.; Martins, C.P.B.; Hernández, F.; Hernández-Cassou, S.; Saurina, J.; Checa, A. LC-MS Interplatform Comparison in Non-Targeted Wine Metabolomics. *J. Chromatogr. A* **2016**, *1433*, 90–97. [CrossRef] [PubMed]
28. R: The R Project for Statistical Computing. 2026. Available online: https://www.r-project.org/ (accessed on 26 June 2026).

29. CRAN: Package Alkahest. 2025. Available online: https://cran.r-project.org/web/packages/alkahest/index.html (accessed on 26 June 2026).
30. CRAN: Package Pracma. 2025. Available online: https://cran.r-project.org/web/packages/pracma/index.html (accessed on 26 June 2026).
31. Mir-Cerdà, A.; Granados, M.; Saurina, J.; Sentellas, S. Green Extraction of Antioxidant Compounds from Olive Tree Leaves Based on Natural Deep Eutectic Solvents. *Antioxidants* **2023**, *12*, 995. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.