

Article

# Machine Learning Model Stability for Sub-Regional Classification of Barossa Valley Shiraz Wine Using A-TEEM Spectroscopy

Han Wang and David W. Jeffery \* 

School of Agriculture, Food and Wine, and Waite Research Institute, The University of Adelaide, PMB 1, Glen Osmond, SA 5064, Australia

\* Correspondence: david.jeffery@adelaide.edu.au

**Abstract:** With a view to maintaining the reputation of wine-producing regions among consumers, minimising economic losses caused by wine fraud, and achieving the purpose of data-driven terroir classification, the use of an absorbance–transmission and fluorescence excitation–emission matrix (A-TEEM) technique has shown great potential based on the molecular fingerprinting of a sample. The effects of changes in wine composition due to ageing and the stability of A-TEEM models over time had not been addressed, however, and the classification of wine blends required investigation. Thus, A-TEEM data were combined with an extreme gradient boosting discriminant analysis (XGBDA) algorithm to build classification models based on a range of Shiraz research wines ( $n = 217$ ) from five Barossa Valley sub-regions over four vintages that had aged in bottle for several years. This spectral fingerprinting and machine learning approach revealed a 100% class prediction accuracy based on cross-validation (CV) model results for vintage year and 98.8% for unknown sample prediction accuracy when splitting the wine samples into training and test sets to obtain the classification models. The modelling and prediction of sub-regional production area showed a class CV prediction accuracy of 99.5% and an unknown sample prediction accuracy of 93.8% when modelling with the split dataset. Inputting a sub-set of the current A-TEEM data into the models generated previously for these Barossa sub-region wines yielded a 100% accurate prediction of vintage year for 2018–2020 wines, 92% accuracy for sub-region for 2018 wines, and 91% accuracy for sub-region using 2021 wine spectral data that were not included in the original modelling. Satisfactory results were also obtained from the modelling and prediction of blended samples for the vintages and sub-regions, which is of significance when considering the practice of wine blending.

**Keywords:** terroir; wine authentication; fluorescence spectroscopy; sub-regionalisation; artificial intelligence; *Vitis vinifera*



**Citation:** Wang, H.; Jeffery, D.W. Machine Learning Model Stability for Sub-Regional Classification of Barossa Valley Shiraz Wine Using A-TEEM Spectroscopy. *Foods* **2024**, *13*, 1376. <https://doi.org/10.3390/foods13091376>

Academic Editors: Chiara Baldacchini and Silvia Portarena

Received: 28 March 2024

Revised: 22 April 2024

Accepted: 24 April 2024

Published: 29 April 2024



**Copyright:** © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Understanding the value of wine requires an appreciation of the influence of terroir—the interaction of physical, biological, and cultural aspects related to provenance and distinctive traits that influence product image, style, and quality. From its creation in the 1960s to today, the term terroir has endured and has even become the focus of investigations that aim to relate terroir to the properties of grapes and wine [1–7]. Considering its underpinnings, terroir necessarily encompasses research disciplines ranging from microbiology, plant and soil science, and oenology to marketing, consumers, humanities, and philosophy. Aside from the influence on grape and wine composition, the complex interactions contributing to terroir lead to a degree of recognition of wine produced in a certain (especially renowned) region, known as its ‘sense of place’ [8]. Despite the complexity and remaining scientific need to elaborate on the influence of terroir, wine producers endow terroir with commercial value and stimulate the potential institutional nature of terroir, making it a valuable marketing tool [1,9,10].

The institutionalisation of terroir has arisen because production regions are controlled according to Protected Designation of Origin (PDO) or Geographical Indication (GI) regulations, which aim to guarantee the authenticity and quality of a wine from a delimited region [1,8]. As reflected in the sub-regionalisation of wine regions, the subdivision of production zones has become a necessary means for the development of a wine-producing area because it is related to the reputation of the region and the interests of local enterprises [11–13]. This can be exemplified by the Barossa Valley, a typical Shiraz-producing region with a long history in Australia, which has stood at a historic turning point in the development of wine sub-regions. Five different potential sub-regions—Northern Grounds (NG), Central Grounds (CGs), Eastern Ridge (ER), Southern Grounds (SG), and Western Ridge (WR)—have been divided and are beginning to be recognised within the industry, becoming a tool to assist in the marketing of Barossa Valley wines [14,15].

In addition to the pursuit of quality, the sub-regionalisation of wine production regions is intended to be used as a means of wine marketing, thereby further increasing a wine's value from a certain production area (with its associated terroir). This can be seen in the division of land into village appellations, which led to the rise in wine prices in Champagne and Burgundy [2,9,12,16]. Wine is thus viewed as a value-added luxury product and an important contributor to the global beverage market, with the European wine sector alone generating billions of dollars in revenue each year [17,18]. In this significant global market with substantial economic benefits, cases of profiteering through wine counterfeiting are common and have long plagued wine producers and local governments [19]. According to an EUIPO report in 2016, the existence of counterfeit wine in the European Union leads to an estimated annual revenue loss of about USD 1.3 billion, equivalent to 3.3% of total sales and an employment loss of about 4800 jobs [20]. As a mainstay of the global wine industry, Australia has also been affected by wine fraud, with counterfeit wine under the famous Penfolds brand, for example, flowing freely in overseas markets [21]. Within Australia, wine fraud more likely relates to honesty around wine label information, the underlying details of which could be altered during the winemaking stage in terms of vintages, varieties, and regional blends, especially in relation to the 85% blending principle [22].

As a wine authentication method, the absorbance–transmission and fluorescence excitation–emission matrix (A-TEEM) approach is based on absorbance and fluorescence spectroscopy [23,24] using an Aqualog instrument with right-angle optical geometry for fluorescence detection [25,26]. This method can generate multidimensional spectral information in the UV–Vis range for all chromophores and fluorophores and simultaneously combines absorbance–transmittance data with an excitation–emission matrix (EEM) to provide unique molecular fingerprints of wine [27]. The total EEM data obtained with this technology involve a set of emission spectra across different wavelengths ( $\lambda_{em}$ ), recorded within a range of excitation wavelengths ( $\lambda_{ex}$ ). This provides information on fluorescent substances in each wine sample and is effective for comparing samples with small compositional differences [27–29]. Ranaweera et al. [30] explored the use of A-TEEM for encapsulating the influence of biophysical and cultural factors associated with wine terroir by tracking wines through the winemaking process, and A-TEEM yielded impressive results for discriminating regions within a single GI [31]. In a study on the origin traceability and authenticity verification of Chinese wine, the combination of EEM and chemometrics was once again proven effective as a wine traceability technology [32]. In addition, compared with traditional wine analysis techniques like high-performance liquid chromatography (HPLC), specific natural isotope fractionation–nuclear magnetic resonance (SNIF-NMR), and isotope ratio mass spectrometry (IRMS) [33–35], acquiring A-TEEM data is relatively simple and accessible, especially for those working in winery laboratories who are already familiar with UV–Vis spectrophotometry.

In terms of spectral data analysis, principal component analysis (PCA) and parallel factor analysis (PARAFAC) are commonly used chemometric techniques. A-TEEM data can also be combined with a machine learning algorithm, such as extreme gradient boosting discrimination analysis (XGBDA), to classify wines from different vintages, varieties,

and regions and even different sub-regions in Barossa Valley. Such chemometric methods demonstrate the ability to analyse nuanced EEM data. PCA and XGBDA can find patterns in datasets and classify samples based on similarities and differences in unsupervised and supervised modes, respectively. As an auxiliary technique, PARAFAC can effectively identify the type and concentration of fluorophores underpinning a dataset classification [24,26,27,31].

Previous work has undoubtedly provided encouraging results for terroir classification at the sub-regional level, with the discrimination of close geographical regions being difficult to accomplish with other analytical methods [31]. The present study aimed to further explore the ultimate depth that this method can reach and consider scenarios encountered in the practical application of the method in the industry. The Barossa Valley GI remained the target region, with an additional vintage of 2021 added along with an assessment of stored wines used in the previous study [31], to explore the influence of bottle ageing. XGBDA was used for machine learning classification modelling, including for the prediction of unknown samples using models developed with a split dataset. The newly collected sample data for the stored wines were assessed with the prediction model established by Ranaweera, Bastian, Gilmore, Capone, and Jeffery [31] two years prior to determine how relevant that previous model was for classifying the current dataset. In addition, samples from the four vintages and five sub-regions were mixed in certain proportions to investigate the practice of wine blending, thereby providing relevance to a typical winemaking scenario.

## 2. Materials and Methods

### 2.1. Chemicals

High-purity water was obtained with the Milli-Q purification system (Elga Labwater, Woodridge, USA). Absolute ethanol for chromatography and analytical-grade hydrochloric acid (HCl, 37% *w/v*) were purchased from Rowe Scientific (Lonsdale, SA, Australia).

### 2.2. Wine Samples

Shiraz research wines ( $n = 217$ ) produced in 2018, 2019, 2020, and 2021 from fruit collected in 20 vineyards were available from a previous project that investigated Barossa Shiraz terroir [36]. As reported before, the wines were made with 100% single-site Shiraz grapes, with fruit parcels obtained from four sites within each of the five sub-regions of Barossa Valley, South Australia, across the four vintages, as follows: Sites 1–4, Northern Grounds (NG) = 42 wines; Sites 5–8, Central Grounds (CG) = 36; Sites 9–12, Eastern Edge (ER) = 48; Sites 13–16, Southern Grounds (SG) = 45; Sites 17–20, Western Ridge (WR) = 46. Replicate wines were available for each site (A, B, C), as shown in Table S1 of the Supporting Information. Winemaking was undertaken by WIC Winemaking Services and wines were bottled in 750 mL glass bottles with screw caps. Bottled wines were stored in a wine cellar under controlled temperature and humidity. Wines from 2018–2020 analysed in the present work (excluding Eden Valley given the focus on Barossa Valley) had aged for an additional 2 years since their previous A-TEEM analysis reported by Ranaweera, Bastian, Gilmore, Capone, and Jeffery [31] (total ageing time of 3–5 years), whereas 2021 wines had been cellared for only 2 years and were analysed for the first time.

### 2.3. Sample Preparation and A-TEEM Procedure

Wine samples (1 mL) obtained from freshly opened bottles were centrifuged (Eppendorf 5415D, Adela Scientific, Thebarton, SA, Australia) at  $9300 \times g$  for 10 min. The supernatant (40  $\mu$ L) was obtained and diluted by 1:150 with degassed and filtered (0.45  $\mu$ m PTFE membrane) 50% aqueous ethanol adjusted to pH 2 with HCl, according to Ranaweera, Gilmore, Capone, Bastian, and Jeffery [23]. After dilution, samples were mixed with a benchtop vortexer (MS1 Minishaker IKA) for 60 s and sonicated for 15 min (SONICLEAN 250HD, Rowe Scientific, Lonsdale, SA, Australia) to remove air bubbles. Samples were analysed in duplicate using Hellma type 1FL (1 cm path length) macroscopic fluorescence

cuvettes (Sigma-Aldrich, Castle Hill, NSW, Australia) using an Aqualog spectrophotometer (UV-800-C, HORIBA Scientific, Quark Photonics, Adelaide, SA, Australia). The settings consisted of 0.2 s integration time, excitation wavelength range 240–800 nm in 5 nm increments, emission range 242–824 nm in 4.66 nm increments, saturation mask width 10 nm, medium detector gain, and automatic spectral pre-processing including the correction of inner filter effects and Rayleigh masking. The EEMs were normalised by the measurement of a standard, sealed, high-purity water cuvette each time the instrument was used, as previously reported [26,31]. The diluted wine sample was stirred in the 1FL cuvette within the sample holder for 120 s with a stir bar before the start of the analysis. Dilution solvent blanks were recorded in the same way prior to sample analysis for auto-subtraction from each sample in the batch [37]. Absorption spectra (240–700 nm) and EEMs were recorded using Aqualog software (version 4.3, HORIBA Scientific, Quark Photonics).

#### 2.4. Preparation of Wine Blends

Wines were prepared in a 12 mL glass vial with silicone/PTFE screw cap (Agilent Technologies, Santa Clara, CA, USA), following the mixture proportions shown in Tables S2–S4 to obtain a final volume of 10 mL of mixed wine sample (in duplicate). The approach was similar to that reported by Ranaweera et al. (2022). After thoroughly mixing vials for 60 s using a benchtop vortex, samples were prepared as above (i.e., centrifuged at  $9300\times g$  for 10 min in a 10 mL centrifuge tube; 40  $\mu$ L of supernatant diluted 150-fold with dilution solvent; samples vortexed to mix and finally sonicated) and analysed to obtain A-TEEM data.

#### 2.5. Statistical Analysis

SOLO + MIA (version 9.2.1.0, Eigenvector Research, Inc., Manson, WA, USA) was used for data processing and analysis.

##### 2.5.1. Data Fusion (Multi-Block Modelling)

The 3D EEMs and corresponding 2D absorbance datasets from A-TEEM were combined to enhance the classification and prediction accuracy [24]. Then, 3D EEM data were reshaped into a two-way data array (unfold multiway mode 1) and joined with the absorbance data. Fused data were used in statistical analyses requiring a 2D dataset, namely, PCA and extreme gradient boosting (XGBoost) modelling.

##### 2.5.2. Unsupervised Data Analysis

PCA and PARAFAC were variously applied to analyse the different datasets collected using the A-TEEM method. For PCA, fused data were auto-scale pre-processed with five principal components selected to classify the five different sub-regions for each of the four vintages. PARAFAC was used to decompose the 3D EEM data of the wine samples into the most dominant fluorophores. For pre-processing, normalisation of spectra to 1 (default) and EEM filtering were applied, with  $\pm 16$  nm and  $\pm 32$  nm for the first-order and second-order Rayleigh filters, respectively [38]. Non-negativity constraints were imposed in all three modes (intensity, emission, and excitation wavelengths) of EEM data, and components were selected based split-half analysis results [38].

##### 2.5.3. Data Analysis with Machine Learning

XGBDA was applied as a classification machine learning algorithm to build the wine authentication models with the fused dataset according to vintage, sub-region, and the specified blends. XGBDA was applied with partial least squares (PLS) compression, using a maximum for latent variables (LVs) of 10 for vintage and vintage blending classification and 20 LVs for sub-region and sub-region blending classification (blends as specified in Table S2 of the Supporting Information). The models developed for vintage and sub-region were applied to the blends specified in Tables S3 and S4, respectively. The number of LVs was selected according to the cross-validation (CV) result accuracy when comparing 10–45 LVs.

Pre-processing was undertaken with mean centring, autoscaling, and generalised least squares weighting (GLSW) with the declutter threshold at 0.02 to calibrate and cross-validate (Venetian blinds procedure,  $k = 10$ ). The xgboost algorithm and gbtrees booster of XGBDA had an  $\eta = 0.3$ ,  $\max\_depth = 1$ , and  $\text{num\_round} = 200$ . Model testing for both vintage and sub-region was taken further by splitting the data into about 80% used for calibration ( $n = 354$ ) and about 20% used for validation ( $n = 80$ ) (keeping the replicates together), using the same XGBDA approach as just described. Further validation was obtained by loading a random subset of the multi-block sample data obtained in the present work (for vintage based on 2018–2020 and for sub-region using 2018 and 2021 as examples) into the previously established model (based on the combination of vintage and sub-region) [31] to test prediction accuracy with newly recorded data for the wines that had aged for a further 2 years and for 2021 wines that were not used before in the classification modelling.

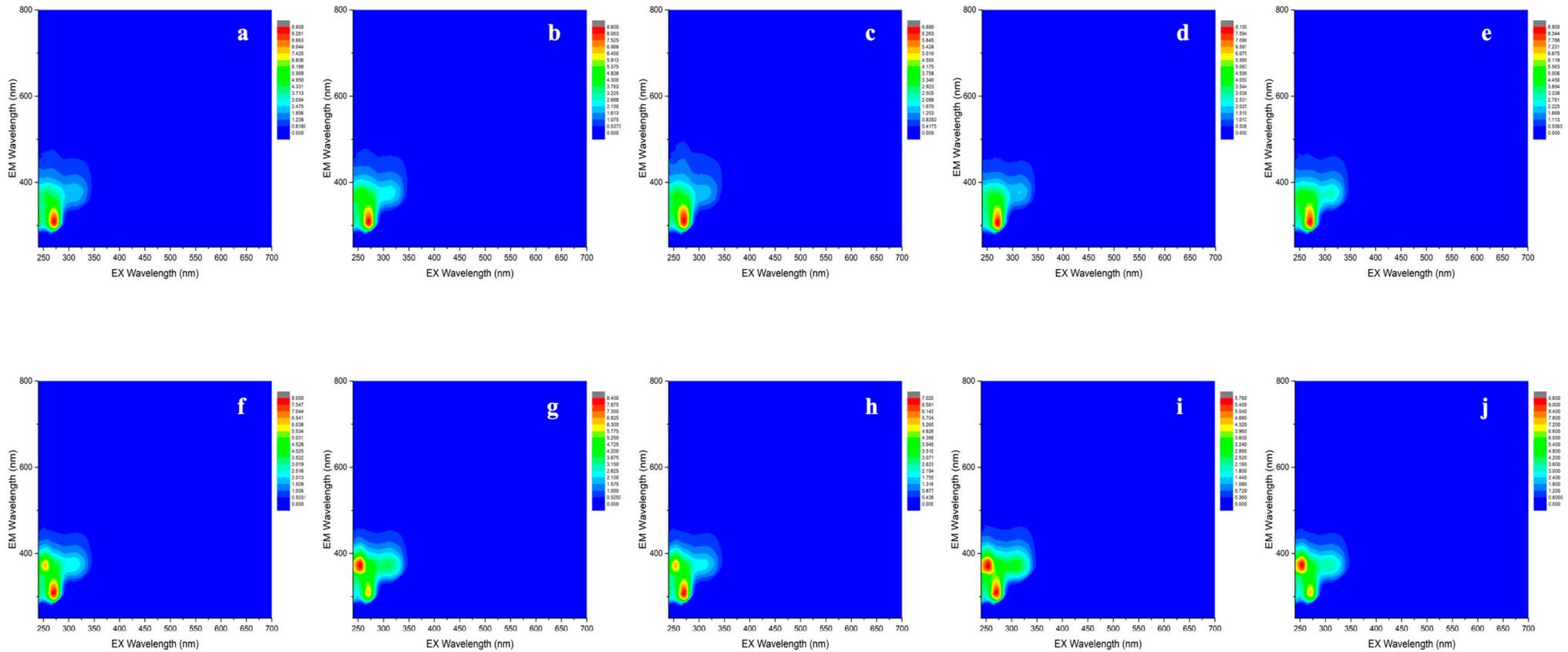
According to the most probable prediction rule, which assigns samples to the class with the highest probability overall, the validity of the model's prediction results was evaluated by the confusion matrix score probability. The scoring probabilities included true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The magnitude of the probability was expressed as a number from 0 to 1 and a percentage.

### 3. Results and Discussion

#### 3.1. Molecular Fingerprints (EEMs)

Figure 1 shows examples of molecular fingerprints for experimental Shiraz wines from Barossa Valley GI, indicating the variance between the 2018 and 2021 vintages for the five sub-regions (NG, CG, ER, SG, WR). The vintage difference can primarily be seen through the gross differences in the EEM fingerprints. Each panel in the first row from vintage 2018 (Figure 1a–e) had only one intense peak at around  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  270/310 nm, with panels in the second row from vintage 2021 (Figure 1f–j) having two intense peaks with  $\lambda_{\text{ex}}/\lambda_{\text{em}}$  at around 270/310 nm and 250/370 nm. Comparing the fingerprints of vintage 2021 wines, Northern Grounds (Figure 1f) and Eastern Ridge (Figure 1h) tended to have a similar fingerprint, as did Central Grounds (Figure 1g) and Western Ridge (Figure 1j), whereas Southern Grounds (Figure 1i) had a more unique fingerprint. In contrast, spectra for the 2018 wines (5 years old) were more similar across the sub-regions. The differences between vintage 2018 and 2021 and sub-region differences within vintage 2021 could be explained on the basis of climatic data such as growing season rainfall, mean January temperature, and growing degree days, as well as other terroir influences [31,39].

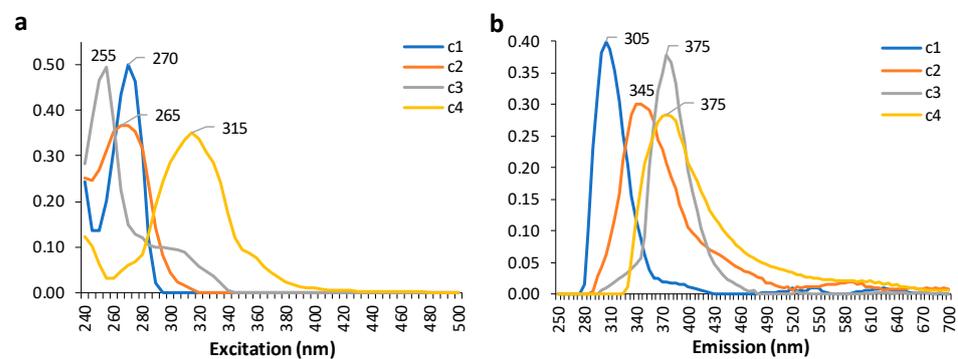
The EEMs were generally representative of the spectral fingerprints obtained with the A-TEEM approach, and, as seen with the sub-regions for vintage 2018, differences may not have been easily discernible by simple visual inspection. This demonstrates the importance of using chemometrics with these datasets to identify subtle patterns in the EEM fingerprints, as elaborated in subsequent sections.



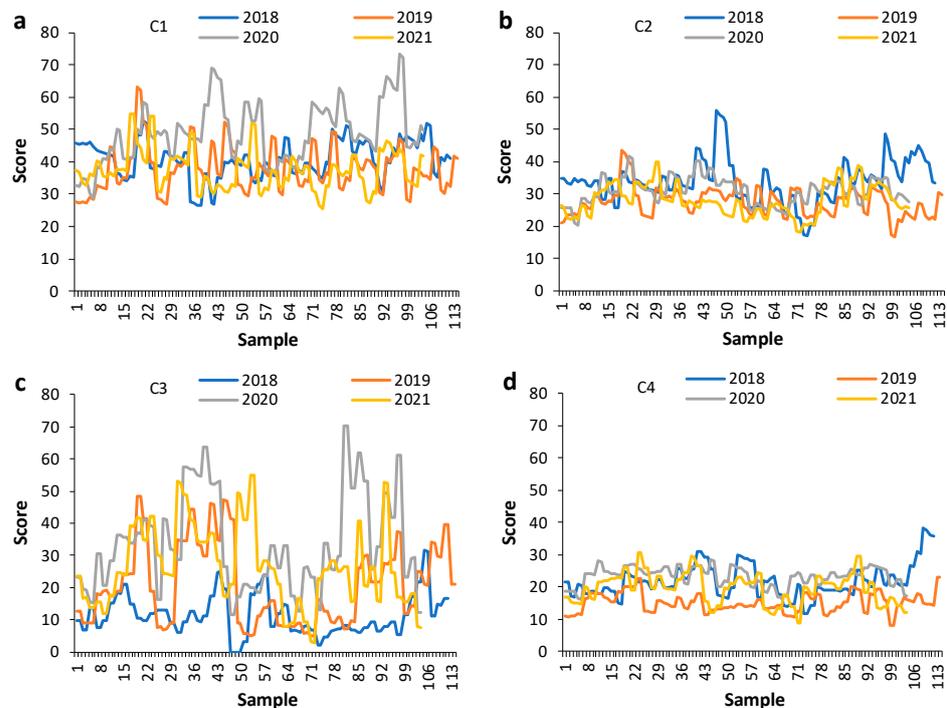
**Figure 1.** Fluorescence excitation and emission matrix (EEM) contour plots (molecular fingerprints) of Shiraz wines, comparing vintage 2018 for subregions (a) Northern Grounds, (b) Central Grounds, (c) Eastern Ridge, (d) Southern Grounds, and (e) Western Ridge and vintage 2021 for subregions (f) Northern Grounds, (g) Central Grounds, (h) Eastern Ridge, (i) Southern Grounds, and (j) Western Ridge.

### 3.2. PARAFAC Decomposition of EEMs

PARAFAC was undertaken to tentatively identify the main fluorophores that characterised the samples. These results could then be used to provide some understanding of the possible compositional drivers underpinning sub-regional classification. A four-component model comprising all sub-regions and vintage years was selected based on a split-half analysis of 97%. PARAFAC modelling (Figure 2) yielded a component 1 peak at 270/305 nm ( $\lambda_{ex}/\lambda_{em}$ ), component 2 peak at 265/345 nm, component 3 peak at 255/375 nm, and component 4 peak at 315/375 nm. Components were tentatively assigned to respective compound classes: 1. flavan-3-ols [28,40]; 2. anthocyanins, aromatic amino acids, and hydroxybenzoic acids [28,40,41]; 3. phenolic acids/aldehydes and flavonols [28]; 4. caffeic and *p*-coumaric acids [40] and stilbenes like resveratrol and *trans*-piceid [41,42] or perhaps grape seed oils from maceration during fermentation (e.g., tocopherols and tetraenes) [43]. The PARAFAC score plots (Figure 3a–d) revealed how the vintages differed based on the tentatively assigned fluorophores.



**Figure 2.** Loadings from parallel factor analysis (PARAFAC) decomposition modelling of 3D EEM data for Shiraz wine samples from all sub-regions and vintages showing (a) excitation wavelengths (nm) and (b) emission wavelengths (nm) of components 1–4.



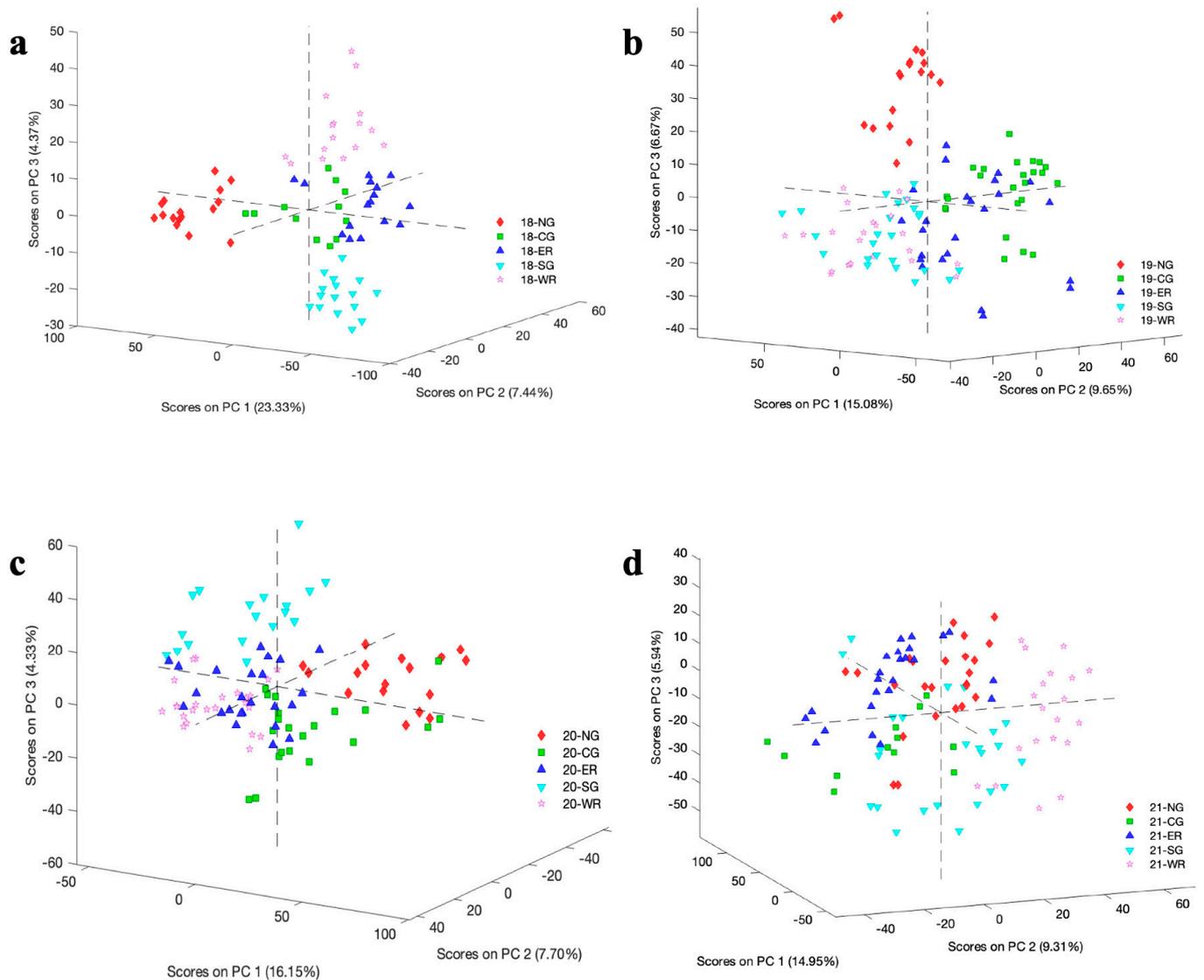
**Figure 3.** PARAFAC score plots for samples from different sub-regions according to vintages 2018–2021 for (a) component 1, (b) component 2, (c) component 3, and (d) component 4.

The PARAFAC components related to ordinary red wine constituents that are influenced by grape growing conditions and terroir more broadly, which could be variable among the sub-regions used in this study. Components 2 (anthocyanins, amino and hydroxybenzoic acids) and 4 (hydroxycinnamates, stilbenes) showed less fluctuation according to vintage than components 1 (flavan-3-ols) and 3 (phenolic acids, flavonols). The seasonal climate could be a particular factor contributing to the variability in some vintage years more than others, with differences in growing season rainfall and temperature for 2018–2021 according to the vintage reports from Barossa Australia [44]. As noted earlier, the gross differences observed in the EEMs presented in Figure 1 that underpin the PARAFAC results could also be related to this observation. Relative wine age could also exert some influence based on the evolution of phenolic profiles over time. Depending on vintage year, greater differences among the sub-regions were also evident, especially for components 1 and 3 (Figure 3a,c).

### 3.3. PCA Decomposition of A-TEEM Data

Dimensionality reduction with PCA was applied to multi-block A-TEEM data (i.e., combined absorbance and EEM datasets) to explore the separation of Barossa sub-regions (Figure 4). The first three principal components accounted for a total variance explained of 35.1%, 31.4%, 28.2%, and 30.2% for vintages from 2018 to 2021, respectively. Vintage 2018 in particular showed an impressive result, with each sub-region tightly grouped and almost completely separated from each other (Figure 4a). This was reminiscent of the results obtained for this vintage in the previous work [31]. However, apart from NG (red diamonds) in 2019 (Figure 4b), and 2020 to a lesser extent (Figure 4c), the other vintages did not exhibit significant differentiation of sub-regions according to PCA. SG (light blue inverted triangles) and WR (lilac stars) were similar to each other in vintages 2019–2021, whereas NG/CG/SG showed an obvious degree of separation in the four vintages, although less so in 2021 (Figure 4d), especially for NG and CG (green squares). The separation of WR and ER (dark blue triangles) from the other three sub-regions largely depended on the vintage, and WR and ER were themselves separated to a degree in vintages 2018 and 2021 (Figure 4a,d). These results were consistent with the study published by Ranaweera, Bastian, Gilmore, Capone, and Jeffery [31]: besides climate factors across vintages (and regions) as mentioned in previous sections, which could lead to more or less differentiation, localised factors of terroir such as soil properties and topography across sub-regions could play a role [15,39] via their influences on grape (and, thus, wine) composition.

Ageing could be another factor that correlated with the separation of sub-regions in the PCA plot (greater separation for older wines), although vintage differences might have a more pronounced influence, considering that the separation of sub-regions for 2018–2020 wines was more or less maintained upon re-analysis of the wines after several years of bottle ageing. This is an important result from an implementation perspective—despite the compositional changes associated with red wines as they age, which can impart changes in wine EEM fingerprints and absorbance values, the original differentiation among the sub-regions according to PCA was still evident several years later. Even so, PCA with multi-block spectral data for these aged wines from different vintages was not sufficient to consistently separate the sub-regions, although k-means clustering was able to resolve vintage year in the previous work [31]. Improvement in sub-regional classification across multiple vintages was necessary, with supervised methods and particularly machine learning algorithms providing a possible solution, as evidenced previously [24,31].



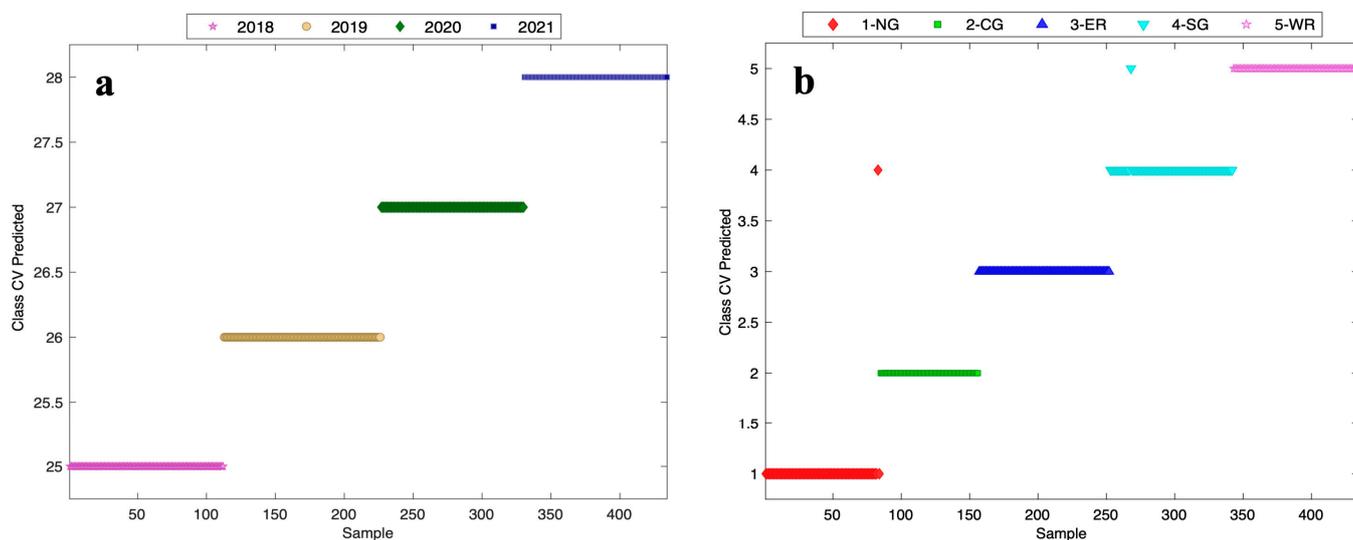
**Figure 4.** Principal component analysis score plots showing the first three principal components (PC 1, PC 2, PC 3) from multi-block wine sample data for five sub-regions across vintages (a) 2018, (b) 2019, (c) 2020, and (d) 2021. NG, Northern Grounds (red diamonds); CG, Central Grounds (green squares); ER, Eastern Ridge (dark blue triangles); SG, Southern Grounds (light blue inverted triangles); WR, Western Ridge (lilac stars).

### 3.4. XGBDA Predictive Modelling

#### 3.4.1. Vintage and Sub-Region Validation

As an effective machine learning classification algorithm [23,30], XGBDA was carried out in an attempt to improve classification across the vintages and sub-regions. The XGBDA approach with CV afforded an excellent classification result (Figure 5 and Table S5 of the Supporting Information), with 100% accuracy for the vintage model (Figure 5a), and 99.5% accuracy (2 misclassified out of 434 sample spectra, Figure 5b) for the sub-region model. These exemplary results were consistent with the work of Ranaweera, Bastian, Gilmore, Capone, and Jeffery [31] and remarkable, considering the proximity of the sub-regions. A further step of splitting the datasets into about 80% for calibration ( $n = 354$ ) and about 20% for validation ( $n = 80$ ) led to slightly lower accuracy than the CV model results shown in Figure 5: in this case, 1 out of 80 samples was misclassified for vintage, giving a 98.8% classification accuracy (Figure S1a–d of the Supporting Information), and 5 out of 80 samples were misclassified at a sub-region level, giving a 93.8% classification accuracy

(Figure S2a–e). Ideally, greater sample numbers would be used for splitting the datasets, but this outcome still highlights the robustness of the A-TEEM approach for classifying wine samples, with the accuracy easily being equal to other authentication techniques. Again, though, it is worth remarking that this study considered wines from sub-regions within a GI (as little as several km apart), thus highlighting the ability to authenticate at a fine scale and the potential of the approach for helping to objectively define unique terroirs within regions.



**Figure 5.** Class predicted from cross-validation (CV) using extreme gradient boosting discriminant analysis (XGBDA) modelling of multi-block A-TEEM data according to (a) vintage and (b) sub-region. Wine samples ( $n = 217$  in duplicate) from four vintages in (a) 2018 (lilac stars), 2019 (yellow circles), 2020 (green diamonds), and 2021 (blue squares) and five sub-regions in (b) NG (red diamonds), CG (green squares), ER (dark blue triangles), SG (light blue inverted triangles), and WR (lilac stars).

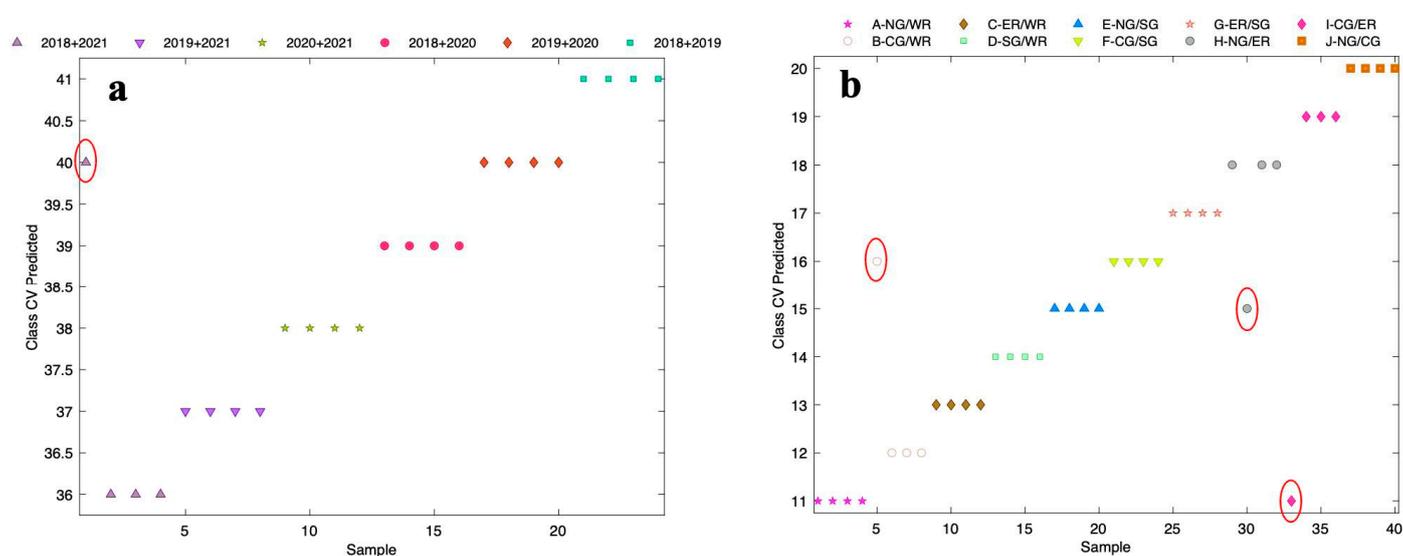
Furthermore, in view of the possible effects of bottle ageing on the A-TEEM data mentioned earlier, and to obtain a deeper understanding of the impact of bottle ageing on this authentication method, the multi-block A-TEEM data obtained in the present study were tested against the previous model developed by Ranaweera, Bastian, Gilmore, Capone, and Jeffery [31] using a subset of the same wine samples from 2018–2020, as well as wines from vintage 2021, which were not analysed previously. According to Figure S3a of the Supporting Information, the prediction of vintage for 2018, 2019, and 2020 wines with the previous model still showed 100% classification accuracy. For the prediction of sub-region, highlighted for 2018 and 2021 wines (excluding Eden Valley, which was not analysed in the present work), the model not only achieved prediction of the 2018 vintage samples with an accuracy of 92% (Figure S3b, four misclassified samples) but it also achieved a prediction accuracy of 91.2% for the samples from vintage 2021 (Figure S3c, three misclassified samples), which had not been involved in the development of the previous model. This was a highly encouraging result as it not only meant that the method was largely unaffected by bottle ageing (i.e., wines measured several years later could still be accurately predicted with the originally developed models) but also demonstrated the sub-region predictive ability using data from an ‘unknown’ vintage (i.e., wines from 2021).

### 3.4.2. Blended Wine Validation

Considering that commercial wines typically consist of blends, it was worthwhile considering the performance of the A-TEEM and XGBDA classification method upon wine blending. Previously, wines were tracked through the winemaking process and XGBoost regression (XGBR) modelling highlighted the sensitivity of the approach to blending one varietal wine with as little as 1% of another [30]. As an extension, selected samples in

the present study were blended across different vintages and separately for different sub-regions. The blending ratio for sub-regions was set as 50:50 and also 15:85, according to the Australian wine industry 85% principle [22], as well as 50:50, 10:90, and 5:95 for vintage (Tables S2–S4 of the Supporting Information).

As the first step, multi-block A-TEEM data of 50:50 mixed samples (Table S2) were selected to establish the model and explore the predictive power of XGBDA with CV. Figure 6a shows the class CV predicted results of 12 blended samples (analysed in duplicate) from combinations of the four vintages, with only one wine misclassified and achieving 95.8% overall classification accuracy (Table S5); the combination of 2018 + 2021 was classified as 2019 + 2020. Figure 6b shows the class CV predicted results of 20 samples (analysed in duplicate) from blending combinations of the five sub-regions, showing that three samples were misclassified (92.5% accuracy, Table S5): one from CG + WR was classified as CG + SG, one from NG + ER was classified as NG + SG, and one from CG + ER was classified as NG + WR. Despite the limited selection of data, the results were consistently quite outstanding for both vintage and sub-region blending.



**Figure 6.** Class CV predicted from XGBDA modelling of multi-block A-TEEM data for different blends (50:50, prepared as outlined in Table S2) according to (a) vintage and (b) sub-region. Samples were blended in duplicate, and each was analysed in duplicate. NG, Northern Grounds; CG, Central Grounds; ER, Eastern Ridge; SG, Southern Grounds; WR, Western Ridge. Samples outlined in red were misclassified.

A further step was carried out by using the XGBDA models established in Section 3.4.1 of the Results and Discussion with multi-block A-TEEM data from selected samples (prepared in duplicate for a single analysis of each) using a stricter blending ratio of 10:90 and 5:95 for vintages (Table S3 of the Supporting Information), along with 15:85 and 50:50 for sub-regions (Table S4), to predict the probable class of each. Table 1 shows the prediction probability based on the vintage and sub-region blending. For wine samples comprising 95% vintage 2018 and 5% vintage 2021 wines (S1 and S2 for vintage), applying the model developed for vintage for all wines gave an average probability of 97.5% that the wine was from the 2018 and a 1.2% probability it was from 2021. For blends containing 90% 2018 wine and 10% 2021 wine (S3 and S4 for vintage), the model predicted that the samples were from 2018 and 2021 with averages of 89.95% and 8.75% probability, respectively. One sample containing 85% SG and 15% WR (S1 for sub-region) was predicted to consist of SG and WR wine with 89% and 6% probability, respectively, but the prediction of S2 for sub-region was extremely low, with only 6.9% and 1.1% probability of the sample coming from SG and WR, respectively. This was an anomalous result without an apparent explanation. Blends of 50% SG and 50% WR (S3 and S4 for sub-region, Table 1) were predicted to come from SG

and WR with averages of 49.8% and 33.3% probability, respectively. Although these results did not automatically imply that class prediction should equal the percentage in the blend, the modelling did reasonably well (errant result aside) in reflecting the main vintage or sub-region component where one predominates, and at least indicates that any blend was not mistaken as a single vintage or individual sub-region.

**Table 1.** Class-predicted probability from XGBDA modelling of multi-block A-TEEM data for blended samples based on vintage (2018 and 2021, prepared as outlined in Table S3) and sub-region (Southern Grounds and Western Ridge, prepared as outlined in Table S4).

Sample	Vintage					
	2018			2021		
	Actual	Predicted	Relative Difference	Actual	Predicted	Relative Difference
S1	95%	97.30%	2.4%	5%	1.30%	−74%
S2	95%	97.80%	2.9%	5%	1.10%	−78%
S3	90%	89.10%	−1.0%	10%	6.30%	−37%
S4	90%	90.80%	0.9%	10%	4.90%	−51%
Sample	Sub-Region					
	Southern Grounds (SG)			Western Ridge (WR)		
	Actual	Predicted	Relative Difference	Actual	Predicted	Relative Difference
S1	85%	88.50%	4.1%	15%	6.20%	−59%
S2	85%	6.90%	−91.9%	15%	1.10%	−93%
S3	50%	41.00%	−18.0%	50%	32.20%	−36%
S4	50%	58.60%	17.2%	50%	34.50%	−31%

The results presented in Table 1 nicely supplement the work of Ranaweera, Gilmore, Bastian, Capone, and Jeffery [30], who reported the use of XGBR modelling of the percentage of grape variety in a blend, with the present study addressing some gaps related to vintage and sub-region blending using A-TEEM data and XGBDA. Importantly, the results in Table 1 were obtained by loading blended wine sample data into the models established using the entire wine datasets for vintage and sub-region that did not involve any blends, thus providing further insight into the potential for the application of this wine authentication methodology in an industry-relevant context.

#### 4. Conclusions

Overall, reliable results have been obtained for the classification of Shiraz research wines arising from adjacent areas within the Barossa Valley GI. The capability of A-TEEM and XGBDA has again shown its worth, with the important contribution of identifying subtle differences among wine samples after a period of bottle ageing but still being able to accurately classify such wines. This adds further weight to the utility of this approach regarding the stability of models over time, which is critical from a classification perspective as wines age. Notably, there was discernment of closely located vineyards even after wine ageing, thus highlighting the conservation of terroir influences on the spectral fingerprints of the wines. The ability of the technique to differentiate wine blends was also an important development in this work, considering that the blending of wine is a widespread and often necessary practice, but one that can be open to manipulation (through the falsification of region or variety, for example). Future improvements on the present outcomes can be envisaged by increasing the size of the dataset or, indeed, creating models for specific blends (potentially allowing for the detection of an unauthorised variety of a PDO wine). In addition, research could be extended to the analysis of commercial wines and the development of an authentication database over numerous vintages based on A-TEEM with machine learning classification.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/foods13091376/s1>: Table S1: Summary of the different vintages, sub-regions, and sites of the Shiraz wines from the Barossa Valley GI analysed in this study; Table S2: Combinations of wine samples in blending (50:50) for A-TEEM analysis with 6 groups for vintage blending and 10 groups for sub-region blending; Table S3: Percentage of wine in the blend of 2018 and 2021 vintages; Table S4: Percentage of wine in the blend of Southern Grounds (SG) and Western Ridge (WR) sub-regions; Table S5: Confusion matrices showing the performance parameters of different cross-validated XGBDA models from multi-block data; Figure S1: Class-predicted member for test set samples (n = 80) from XGBDA modelling of multi-block A-TEEM data for vintage; Figure S2: Class-predicted member for test set samples (n = 80) from XGBDA modelling of multi-block A-TEEM data for subregion; Figure S3: Class-predicted member using the XGBDA model established by Ranaweera et al. (2023) with multi-block A-TEEM test data from the present study [31].

**Author Contributions:** Conceptualisation, H.W. and D.W.J.; data curation, H.W.; formal analysis, H.W.; investigation, H.W.; methodology, H.W. and D.W.J.; project administration, D.W.J.; resources, D.W.J.; supervision, D.W.J.; visualisation, H.W. and D.W.J.; writing—original draft preparation, H.W. and D.W.J.; writing—review and editing, H.W. and D.W.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** H.W. acknowledges the financial support provided by the Mortlock Honours Scholarship and project funding from the School of Agriculture, Food and Wine, the University of Adelaide.

**Data Availability Statement:** The original contributions presented in the study are included in the article/Supplementary Materials; further inquiries can be directed to the corresponding author.

**Acknowledgments:** We are very grateful for being able to access research wines from a Wine Australia funded project (UA1602). John Gledhill from WIC Winemaking Services is acknowledged for his assistance with wine sample collection. We appreciate the technical support with the Aqualog instrument provided by Ruchira Ranaweera (formerly the University of Adelaide), Adam Gilmore from HORIBA Instruments Inc., and Andrew Jane from Quark Photonics.

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

## References

- Castelló, E. The will for terroir: A communicative approach. *J. Rural Stud.* **2021**, *86*, 386–397. [CrossRef]
- Hill, R.A.D. ‘Le terroir, c’est la vie’: Re-animating a concept among Burgundy’s wine producers. *Environ. Plan. E Nat. Space* **2022**, *5*, 447–472. [CrossRef]
- Parker, T. *Tasting French Terroir: The History of an Idea*, 1st ed.; University of California Press: Berkeley, CA, USA, 2015.
- Gladstones, J. *Wine, Terroir and Climate Change*; Wakefield Press: Mile End, SA, Australia, 2011.
- Kustos, M.; Gambetta, J.M.; Jeffery, D.W.; Heymann, H.; Goodman, S.; Bastian, S.E.P. A matter of place: Sensory and chemical characterisation of fine Australian Chardonnay and Shiraz wines of provenance. *Food Res. Int.* **2020**, *130*, 108903. [CrossRef] [PubMed]
- Souza Gonzaga, L.; Capone, D.L.; Bastian, S.E.P.; Jeffery, D.W. Defining wine typicity: Sensory characterisation and consumer perspectives. *Aust. J. Grape Wine Res.* **2021**, *27*, 246–256. [CrossRef]
- Cramer, G.R.; Cochetel, N.; Ghan, R.; Destrac-Irvine, A.; Delrot, S. A sense of place: Transcriptomics identifies environmental signatures in Cabernet Sauvignon berry skins in the late stages of ripening. *BMC Plant Biol.* **2020**, *20*, 41. [CrossRef] [PubMed]
- Unwin, T. *Terroir: At the heart of geography*. In *The Geography of Wine*; Dougherty, P., Ed.; Springer: Dordrecht, The Netherlands, 2011; pp. 37–48.
- Charters, S.; Spielmann, N. Characteristics of strong territorial brands: The case of champagne. *J. Bus. Res.* **2014**, *67*, 1461–1467. [CrossRef]
- Barham, E. Translating terroir: The global challenge of French AOC labeling. *J. Rural Stud.* **2003**, *19*, 127–138. [CrossRef]
- Rosen, S. Institutional transformation: Supply or demand? Concluding comment. *J. Inst. Theor. Econ.* **1996**, *152*, 275–286.
- Meloni, G.; Swinnen, J. Trade and terroir. The political economy of the world’s first geographical indications. *Food Policy* **2018**, *81*, 1–20. [CrossRef]
- Schamel, G. Geography versus brands in a global wine market. *Agribusiness* **2006**, *22*, 363–374. [CrossRef]
- Robins, N. The Barossa Grounds—The journey so far. *Aust. N. Z. Grapegrow. Winemak.* **2014**, *601*, 42.
- Bramley, R.G.V.; Ouzman, J. Underpinning terroir with data: On what grounds might subregionalisation of the Barossa Zone geographical indication be justified? *Aust. J. Grape Wine Res.* **2022**, *28*, 196–207. [CrossRef]
- Thach, L.; Charters, S.; Cogan-Marie, L. Core tensions in luxury wine marketing: The case of Burgundian wineries. *Int. J. Wine Bus. Res.* **2018**, *30*, 343–365. [CrossRef]

17. Anderson, K.; Nelgen, S.; Pinilla, V. *Global Wine Markets: 1860 to 2016: A Statistical Compendium*; University of Adelaide Press: Adelaide, SA, Australia, 2017.
18. Guy, K.M. "Oiling the wheels of social life": Myths and marketing in Champagne during the Belle Epoque. *Fr. Hist. Stud.* **1999**, *22*, 211–239. [[CrossRef](#)] [[PubMed](#)]
19. Holmberg, L. Wine fraud. *Int. J. Wine Res.* **2010**, *2*, 105–113. [[CrossRef](#)]
20. Gabel, B. Wine origin authentication linked to terroir—Wine fingerprint. *BIO Web Conf.* **2019**, *15*, 02033. [[CrossRef](#)]
21. Smith, M. Chinese police seize fake Penfolds haul. *Aust. Financ. Rev.* **2018**. Available online: <https://www.afr.com/world/asia/chinese-police-seize-50000-bottles-of-fake-penfolds-20180327-h0y19c> (accessed on 17 March 2024).
22. Wine Australia. The Blending Rules. Available online: <https://www.wineaustralia.com/labelling/further-information/the-blending-rules> (accessed on 17 March 2024).
23. Ranaweera, R.K.R.; Gilmore, A.M.; Capone, D.L.; Bastian, S.E.P.; Jeffery, D.W. Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling. *Food Chem.* **2021**, *335*, 127592. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. Gilmore, A.; Akaji, S.; Csatorday, K. Spectroscopic analysis of red wines with A-TEEM molecular fingerprinting. *Readout* **2017**, *E49*, 41–48.
26. Gilmore, A.M.; Sui, Q.; Blair, B.; Pan, B.S. Accurate varietal classification and quantification of key quality compounds of grape extracts using the absorbance-transmittance fluorescence excitation emission matrix (A-TEEM) method and machine learning. *OENO One* **2022**, *56*, 107–115. [[CrossRef](#)]
27. Ranaweera, R.K.R.; Gilmore, A.M.; Capone, D.L.; Bastian, S.E.P.; Jeffery, D.W. Spectrofluorometric analysis combined with machine learning for geographical and varietal authentication, and prediction of phenolic compound concentrations in red wine. *Food Chem.* **2021**, *361*, 130149. [[CrossRef](#)] [[PubMed](#)]
28. Airado-Rodríguez, D.; Durán-Merás, I.; Galeano-Díaz, T.; Wold, J.P. Front-face fluorescence spectroscopy: A new tool for control in the wine industry. *J. Food Compos. Anal.* **2011**, *24*, 257–264. [[CrossRef](#)]
29. Cabrera-Bañegil, M.; Hurtado-Sánchez, M.d.C.; Galeano-Díaz, T.; Durán-Merás, I. Front-face fluorescence spectroscopy combined with second-order multivariate algorithms for the quantification of polyphenols in red wine samples. *Food Chem.* **2017**, *220*, 168–176. [[CrossRef](#)] [[PubMed](#)]
30. Ranaweera, R.K.R.; Gilmore, A.M.; Bastian, S.E.P.; Capone, D.L.; Jeffery, D.W. Spectrofluorometric analysis to trace the molecular fingerprint of wine during the winemaking process and recognise the blending percentage of different varietal wines. *OENO One* **2022**, *56*, 189–196. [[CrossRef](#)]
31. Ranaweera, R.K.R.; Bastian, S.E.P.; Gilmore, A.M.; Capone, D.L.; Jeffery, D.W. Absorbance-transmission and fluorescence excitation-emission matrix (A-TEEM) with multi-block data analysis and machine learning for accurate intraregional classification of Barossa Shiraz wine. *Food Control* **2023**, *144*, 109335. [[CrossRef](#)]
32. Wu, Q.; Geng, T.; Yan, M.-L.; Peng, Z.-X.; Chen, Y.; Lv, Y.; Yin, X.-L.; Gu, H.-W. Geographical origin traceability and authenticity detection of Chinese red wines based on excitation-emission matrix fluorescence spectroscopy and chemometric methods. *J. Food Compos. Anal.* **2024**, *125*, 105763. [[CrossRef](#)]
33. Gomes, S.; Castro, C.; Barrias, S.; Pereira, L.; Jorge, P.; Fernandes, J.R.; Martins-Lopes, P. Alternative SNP detection platforms, HRM and biosensors, for varietal identification in *Vitis vinifera* L. using *F3H* and *LDOX* genes. *Sci. Rep.* **2018**, *8*, 5850. [[CrossRef](#)] [[PubMed](#)]
34. Martin, G.J.; Guillou, C.; Martin, M.L.; Cabanis, M.T.; Tep, Y.; Aerny, J. Natural factors of isotope fractionation and the characterization of wines. *J. Agric. Food Chem.* **1988**, *36*, 316–322. [[CrossRef](#)]
35. Bonello, F.; Cravero, M.; Dell'Oro, V.; Tsolakis, C.; Ciambotti, A. Wine traceability using chemical analysis, isotopic parameters, and sensory profiles. *Beverages* **2018**, *4*, 54. [[CrossRef](#)]
36. Collins, C.; Petrie, P.; Bastian, S.; Bindon, K.; Bonada, M.; Boss, P.; Bramley, R.; Cavagnaro, T.; Danner, L.; Francis, L.; et al. Understanding the drivers of terroir in the Barossa. In *Final Report to Wine Australia*; 2022; pp. 1–279. Available online: [https://www.wineaustralia.com/getmedia/07afbc68-8596-466b-95fd-b05899d673c7/UA-1602-Final-Report\\_Final.pdf](https://www.wineaustralia.com/getmedia/07afbc68-8596-466b-95fd-b05899d673c7/UA-1602-Final-Report_Final.pdf) (accessed on 17 March 2024).
37. Ranaweera, R.K.R.; Gilmore, A.M.; Jeffery, D.W. Fluorescence spectroscopy for red wine authentication. In *Wine Analysis and Testing Techniques*; Pozo-Bayón, M.Á., Muñoz González, C., Eds.; Springer: New York, NY, USA, 2024; pp. 23–38.
38. Lenhardt Acković, L.; Zeković, I.; Dramićanin, T.; Bro, R.; Dramićanin, M.D. Modeling food fluorescence with PARAFAC. In *Reviews in Fluorescence 2017*; Geddes, C.D., Ed.; Springer International Publishing AG: Cham, Switzerland, 2019; pp. 161–197.
39. Bramley, R.G.V.; Ouzman, J.; Trought, M.C.T. Making sense of a sense of place: Precision viticulture approaches to the analysis of terroir at different scales. *OENO One* **2020**, *54*, 903–917. [[CrossRef](#)]
40. Schueuermann, C.; Silcock, P.; Bremer, P. Front-face fluorescence spectroscopy in combination with parallel factor analysis for profiling of clonal and vineyard site differences in commercially produced Pinot Noir grape juices and wines. *J. Food Compos. Anal.* **2018**, *66*, 30–38. [[CrossRef](#)]
41. Airado-Rodríguez, D.; Galeano-Díaz, T.; Durán-Merás, I.; Wold, J.P. Usefulness of fluorescence excitation–emission matrices in combination with PARAFAC, as fingerprints of red wines. *J. Agric. Food Chem.* **2009**, *57*, 1711–1720. [[CrossRef](#)] [[PubMed](#)]

42. Díaz, T.G.; Durán Merás, I.; Airado Rodríguez, D. Determination of resveratrol in wine by photochemically induced second-derivative fluorescence coupled with liquid-liquid extraction. *Anal. Bioanal. Chem.* **2007**, *387*, 1999–2007. [[CrossRef](#)]
43. Szukay, B.; Gałęcki, K.; Kowalska-Baron, A.; Przybyt, M.; Saletnik, L.; Budzyński, J.; Fisz, J.J. Application of steady-state and time-resolved fluorescence spectroscopy in identification of cold-pressed vegetal oils. *Food Anal. Methods* **2023**, *16*, 1571–1582. [[CrossRef](#)]
44. Barossa Australia. Barossa Vintage Reports. Available online: <https://barossawine.com/barossa-vintages/barossa-vintage-reports/> (accessed on 17 March 2024).

**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.