Spectrofluorometric analysis to trace the molecular fingerprint of wine during the winemaking process and recognise the blending percentage of different varietal wines

Ranaweera K.R. Ranaweera¹, Adam M. Gilmore², Susan E.P. Bastian¹,³, Dimitra L. Capone¹,³ and David W. Jeffery¹,³

¹ Department of Wine Science and Waite Research Institute, The University of Adelaide, PMB 1, Glen Osmond, South Australia 5064, Australia
² HORIBA Instruments Inc., 20 Knightsbridge Rd., Piscataway, NJ 08854, USA
³ Australian Research Council Training Centre for Innovative Wine Production, The University of Adelaide, PMB 1, Glen Osmond, South Australia 5064, Australia

ABSTRACT

As a robust analytical method, spectrofluorometric analysis with machine learning modelling has recently been used to authenticate wine from different regions, vintages and varieties. This preliminary study investigated whether the molecular fingerprint obtained with this approach is maintained throughout the winemaking process, along with assessing different percentages of wine in a blend. Monovarietal wine samples were collected at different stages of the winemaking process and analysed with the absorbance-transmission and fluorescence excitation-emission matrix (A-TEEM) technique. Wines were clustered tightly according to origin for the different winemaking stages, with some clear separation of different regions and varieties based on principal component analysis. In addition, wines were classified with 100 % accuracy according to varietal origin using extreme gradient boosting (XGB) discriminant analysis. The sensitivity of the A-TEEM technique was such that it allowed for accurate modelling of wine blends containing as little as 1 % of Cabernet-Sauvignon or Grenache in Shiraz wine when employing XGB regression, which performed better than partial least squares regression. The overall results indicated the potential for applying A-TEEM and machine learning modelling to wine chemical traceability through production to guarantee the provenance of wine or identify the composition of a blend.

KEYWORDS: Authenticity, excitation-emission matrix, traceability, chemometrics, vinification
**INTRODUCTION**

Wine is an attractive target for fraud because it is a luxury product in a high-value industry worth hundreds of billion dollars globally. Wine fraud can occur in different forms, such as dilution, substitution, illegal addition and mislabelling (Ranaweera et al., 2021). To ensure the provenance of wine and to combat wine fraud, it is important to verify the origin and identity of the product by applying proper authentication and traceability techniques. Even though several analytical methods have been developed for wine authentication, it is challenging to find a technique to verify the original fingerprint of the product that has been maintained throughout production due to the complexity of the winemaking process (Aceto et al., 2013).

At the very least, winemaking involves alcoholic fermentation but can encompass other processes such as malolactic fermentation, use of permitted additives or maturation techniques and blending of different varietals. Each of these processes imparts alterations to wine composition: alcoholic fermentation produces compounds such as higher alcohols, esters, glycerol, acetaldehyde and acids (Styger et al., 2011); malolactic fermentation involves changes that impact wine aroma and flavour profiles beyond the conversion of malic acid into lactic acid (Lonvaud-Funel, 2010); and interactions of wine macromolecules such as polysaccharides with proteins, tannins and aroma compounds also affect the wine matrix (Jones-Moore et al., 2022). Some components in wine do not change significantly during the vinification process, which offers the opportunity to identify chemical markers that could be applied for authentication purposes (Catalano et al., 2016; Versari et al., 2014).

Few studies have been conducted to verify the possibility of tracing chemical markers during winemaking. Analysis of metal composition throughout the winemaking process revealed that only a few elements maintained constant concentrations (Castiñeira et al., 2004). In their study, Almeida and Vasconcelos (2004) showed that Sr/Sr isotope values were statistically identical and can be applied to the provenance of soil and respective grape juice and wine. A study of phenolic profile during winemaking using Fourier-transform infrared spectroscopy identified that the total phenolic content did not change significantly after primary and malolactic fermentation (Preserova et al., 2015). However, the blending process used to produce a finished wine affects polyphenols and colour (Li et al., 2020) and bentonite used for protein stabilisation can influence the distribution of various metals (Aceto et al., 2013). Furthermore, although blending is an important step for producing wine with appealing sensory properties (Dooley et al., 2012) that may underpin the reputation of a designated origin (DO), such as Bordeaux blends involving Cabernet and Merlot or Australian Shiraz and Cabernet blends (Souza Gonzaga et al., 2021; Wine Australia, 2017), it can introduce uncertainty for confirming authenticity. For example, there could be unauthorised blending of DO wine with a small percentage of non-DO wine to increase total volume, or there may be a need to identify blending proportions for labelling requirements, such as having 85 % or more of the variety or geographical indication stated on the bottle label in accordance with the label integrity programme in Australia (Wine Australia, 2018). Imparato et al. (2011) applied nuclear magnetic resonance (NMR) profiling to a range of red wine varieties and achieved a precision of about 10 % when differentiating wine blends. However, for authentication purposes, a robust (and preferably rapid) method with high accuracy was still required to verify the blends of different grape varieties.

Considering that fluorescence spectroscopy can offer a viable method for wine authentication (Ranaweera et al., 2021a, Ranaweera et al., 2021b), the present study used a spectrofluorometric technique (absorbance-transmission and fluorescence excitation-emission matrix, or A-TEEM) in combination with machine learning modelling to test two hypotheses for the first time: 1) the molecular fingerprint of wine as a function of origin can be traced through steps of the winemaking process and 2) the blending percentages of different wines can be detected. The effectiveness of the cross-validated models was evaluated and compared according to the score probabilities in the confusion matrix and root mean square error of cross-validation (RMSECV) along with the coefficient of determination of cross-validation (R²CV).

**MATERIALS AND METHODS**

1. **Chemicals and solvents**

HPLC gradient grade absolute ethanol and analytical grade 37 % hydrochloric acid (HCl) were purchased from Chem-Supply (Port Adelaide, SA, Australia). High purity water was obtained from a Milli-Q purification system (Millipore, North Ryde, NSW, Australia).

2. **Wine samples**

Two sets of wine samples were obtained to examine the stage of wine production and for blending experiments. For the stage of production, five different monovarietal wines (Grenache from Alverstoke vineyard and Coome vineyard at the University of Adelaide’s Waite Campus, Mataro from Coome vineyard, Shiraz from Barossa Valley, and Nebbiolo from Southern Flinders Ranges) were collected in 2021 from the research and teaching winery at the Waite Campus at three different processing stages: post-primary fermentation (PF) when glucose and fructose were less than 2 g/L; post-malolactic fermentation (MF) when malic acid concentration was less than 0.1 g/L; and pre-blending (PB) from 225 L barrels. For the blending experiments, three different commercially produced but unreleased monovarietal wines (Shiraz from Langhorne Creek, Cabernet-Sauvignon from Langhorne Creek, and Grenache from Riverland) were obtained from a local producer in 2020.

3. **Analytical procedures for basic chemical parameters**

Wine pH and titratable acidity (TA) were measured with a Mettler Toledo T50 autotitrator, and alcohol content (percentage by volume) was measured by densitometry after...
distillation by Commercial Services at the Australian Wine Research Institute. Analyses were undertaken in duplicate.

4. Sample preparation and A-TEEM analysis for winemaking stages

Samples were obtained from fermentation vessels or barrels at PF, MF and PB stages of production and stored in plastic containers in a freezer at -20 °C until required for analysis to inhibit fermentation. At the time of analysis, samples were defrosted at room temperature and prepared and analysed in duplicate as described by Ranaweera et al. (2021b), undertaking two measurements of each replicate sample. Briefly, samples (1 mL) were centrifuged (Eppendorf 5415D, Adelab Scientific, Thebarton, SA, Australia) at 9300 × g for 10 min and an aliquot (40 μL) was diluted 1:100 with 50 % aqueous ethanol that had been adjusted to pH 2 with HCl and degassed by vacuum filtration (0.45 μm PTFE membrane). The dilution factor of wine-to-solvent was determined by considering the absorbance values of samples according to Beer-Lambert law (Gilmore, 2014). Samples were mixed for 60 s using a benchtop vortex (Grant-bio, PV-1) and degassed by sonication for 10 min with a Unisonics ultrasonic cleaner (Rowe Scientific, Adelaide, SA, Australia). A-TEEM analysis was conducted with a HORIBA Scientific Aqualog spectrophotometer (version 4.2, Quark Photonics, Adelaide, SA, Australia) using the same instrument settings as reported previously (Ranaweera et al., 2021b) (i.e., an excitation wavelength range of 240–800 nm with a 5 nm increment under medium gain and 0.2 s integration time; emission wavelength range of 242–824 nm with a 4.66 nm increment as set by the instrument). Samples were analysed in a Hellma type 1FL (1 cm path length) Macro Fluorescence cuvette (Sigma-Aldrich, Castle Hill, NSW, Australia). Absorbance spectra (240–700 nm) and EEMs were recorded using Origin software for data acquisition (version 8.6, OriginLab Corporation, Massachusetts, USA). Wine colour measurements comprising CIELab, hue and intensity were also recorded. Pre-processing of excitation-emission matrix (EEM) data involved normalisation according to the water Raman scattering units for the specified emission conditions and correcting for the influence of inner filter effects (IFE), solvent background, dark detector signals and Rayleigh masking to eliminate spectral distortion (Gilmore et al., 2017).

5. Sample preparation and A-TEEM analysis for blending experiment

Wines were added into 12 mL glass vials with Teflon lined caps to prepare the blends as shown in Table 1 to obtain a final volume of 10 mL. After addition, vials were mixed thoroughly for 60 s using a benchtop vortex and samples were prepared and analysed in duplicate as described in Section 4 of Materials and methods, but using a dilution of 1:150.

6. Statistical analysis

One-way analysis of variance (ANOVA) with Tukey’s honestly significant difference (HSD) post hoc test for pairwise comparisons (α = 0.05) for basic chemical measures and wine colour parameters according to stage of winemaking and region was undertaken with XLSTAT (version 2019.03.02, Addinsoft, Boston, USA). EEM data were unfolded into a two-way array using transform unfold multiway (mode 1) in Solo software (version 8.7.1, Eigenvector Research, Inc., Manson, WA, USA). Principal component analysis (PCA) was carried out with singular value decomposition and autoscale pre-processing with four principal components to explore variations in samples at different stages of winemaking using Solo software. Samples were labelled with their variety according to the winemaking stage and classified using extreme gradient boosting discriminant analysis (XGBDA) after partial least squares (PLS) compression using five latent variables (LV), with mean centring pre-processing and decluttering with generalised least squares weighting (GLSW) at 0.2 to both calibrate and cross-validate (k = 10, Venetian blinds procedure). According to previous studies, the model was evaluated using confusion matrix score probabilities (Ranaweera et al., 2021a, Ranaweera et al., 2021b). For the blending experiment, unfolded EEM data were modelled with PLS and XGB regression algorithms (Solo software) using blending percentage as the y-block. Root mean square error of cross-validation (RMSECV) (Venetian blinds with 10 splits) and coefficients of determination for both calibration and cross-validation (R² cal, R² CV) were used to evaluate the effectiveness of the models.

RESULTS AND DISCUSSION

1. Variations according to stage of winemaking

CIELab colour parameters and basic oenological measurements of wine samples obtained during the

| TABLE 1. Percentages of wine in blends of Shiraz with Cabernet-Sauvignon or Grenache. |
|-----------------|-----------------|
| Variety         | Blending percentage (v/v) |
| Shiraz          | 100              |
|                | 99               |
|                | 95               |
|                | 90               |
|                | 85               |
|                | 60               |
|                | 50               |
|                | 0                |
| Cabernet-Sauvignon or Grenache | 0 |
|                | 1                |
|                | 5                |
|                | 10               |
|                | 15               |
|                | 40               |
|                | 50               |
|                | 100              |
The winemaking process were assessed with one-way ANOVA according to different winemaking stages as well as according to origin (for different varieties), as shown in Table S1 and Table S2, respectively, of the Supplementary data. When analysed according to the winemaking stage (Table S1), there were no significant differences ($p > 0.26$) in basic chemistry (alcohol, pH, TA) nor colour parameters (hue, intensity, $L^*$, $a^*$, $b^*$, $C^*$). Values for the chromatic characteristics at the different winemaking stages showed that the wines were relatively low in lightness ($L^*$), moderately high in red ($a^*$) and yellow ($b^*$) and high in chroma ($C^*$). These results generally aligned with variations among oenological properties and colour expression during winemaking (Arcena et al., 2020), depending on the stage/time period of sampling. According to the origin of the samples (Table S2), alcohol % v/v and all colour parameters showed significant variation ($p < 0.0001$), whereas pH and TA were not significantly different.

In the CIE 1931 $xyY$ colour space, all samples were congregated together in the red zone ($x = 0.68$ to 0.72 and $y = 0.27$ to 0.31, Figure 1A), which contrasted with the hue vs. intensity plot, where clear separation of Shiraz from Barossa Valley and Nebbiolo from Southern Flinders Ranges could be observed (Figure 1B). Furthermore, Grenache and Mataro samples from vineyards at the Waite campus (Alverstoke and Coombe) were clustered relatively close but were still somewhat differentiated. Based on this simple analysis, it appeared that unique information related particularly to the origin of samples that was not impacted by the processing stage could be expressed from absorbance data. The observations were interesting, but the stages of winemaking were seemingly overshadowed. As such, further exploratory analysis was carried out with EEM data (which can be considered as a molecular fingerprint (Gilmore et al., 2017) using PCA (Figure 2). The first three principal components explained $94.80\%$ of the total variance for the samples, which were perfectly clustered according to their origin for the different winemaking stages. Wines from the different regions and varieties were reasonably well separated along PC1 except for Mataro from Coombe vineyard and Shiraz from Barossa Valley. PC2 especially segregated Nebbiolo from Southern Flinders Ranges from the remainder and the Shiraz to an extent, whereas Waite Campus vineyard samples (Grenache and Mataro) overlapped. The Shiraz was well separated from other samples along PC3. This outcome indicated for the first time that the fluorescence molecular fingerprint according to origin could be traced (and seemingly preserved) during winemaking.

XGBDA was subsequently carried out as reported (Ranaweera et al., 2021a, Ranaweera et al., 2021b) for classification by origin. Figure 3 shows the class cross-validation (CV, Venetian blinds method) prediction probability from this machine learning approach, revealing the probability of each sample belonging to the class it most closely resembles. Class CV prediction demonstrated excellent separation of samples according to their origin, grouping all stages of winemaking (i.e., post-primary fermentation, post-malolactic fermentation, and pre-blending) together for each class. These results further emphasised the distinct possibility of tracing samples through different stages of winemaking according to their origin. Thus, EEM data from the A-TEEM technique could provide an original spectral fingerprint of

**FIGURE 1.** Analysis of colour parameters according to the stage of the winemaking process ($n = 15$, duplicate samples analysed twice) showing (A) CIE 1931 plot (inset shows clustering of samples) and (B) hue vs. colour intensity graph. PF, post-primary fermentation; MF, post-malolactic fermentation, PB, pre-blending. 1, Grenache from Alverstoke vineyard; 2, Grenache from Coombe vineyard; 3, Shiraz from Barossa Valley; 4, Mataro from Coombe vineyard; 5, Nebbiolo from Southern Flinders Ranges.
FIGURE 2. Scores from PCA of EEM data for samples of different variety/origin collected at three stages of winemaking (n = 15, duplicate samples analysed twice). Gre Alv, Grenache from Alverstoke vineyard; Gre Coo, Grenache from Coombe vineyard; Mat Coo, Mataro from Coombe vineyard; Neb SFR, Nebbiolo from Southern Flinders Ranges; Shz BV, Shiraz from Barossa Valley. PF, post-primary fermentation; MF, post-malolactic fermentation; PB, pre-blending.

FIGURE 3. Class CV predicted for wine origin from XGBDA analysis of EEM data for samples collected at three stages of winemaking (n = 15, duplicate samples analysed twice). Gre Alv, Grenache from Alverstoke vineyard; Gre Coo, Grenache from Coombe vineyard; Mat Coo, Mataro from Coombe vineyard; Neb SFR, Nebbiolo from Southern Flinders Ranges; Shz BV, Shiraz from Barossa Valley. PF, post-primary fermentation; MF, post-malolactic fermentation; PB, pre-blending.
FIGURE 4. XGB regression of measured vs. CV predicted blending percentages for Shiraz wine containing proportions of (A) Cabernet-Sauvignon or (B) Grenache. Insets show more detail of the sample separation for 0%–15% blends.
the product that can be maintained during wine production, thereby opening up avenues for this being used as a chemical signature for traceability.

2. Modelling to identify blend proportions

Testing the A-TEEM approach for sensitivity in terms of changes in matrix from introducing a blending component was another important consideration regarding potential fraud detection. To evaluate the possibility of identifying the blending percentage of each sample, regression methods were applied to EEM data for Shiraz wine containing proportions of Cabernet-Sauvignon or Grenache. As a common method, PLS regression (PLSR) was applied for the two sets of wines blended according to the amounts in Table 1. The correlation between the actual blends and predicted percentages were evaluated, with $R^2$ CV and RMSECV values for Shiraz and Cabernet-Sauvignon blends (0.996, 2.17) and Shiraz and Grenache blends (0.992, 3.12) as shown in Figure S1. The accuracy of the models was good, with $R^2$ CV values > 0.990 for both sets of blends, but the RMSECV values were slightly high, at 2–3 %. PLSR uses latent variables (components) that explain as much of the covariance as possible between a set of predictor X-variables and response Y-variables (Ghanem et al., 2015). A study by Gilmore et al. (2020) identified that XGB regression (XGBR) yielded more precise fits for the prediction of phenolic compound and anthocyanin concentrations from A-TEEM data compared to PLSR. Therefore, XGB regression was applied to the blending experiment data to seek improvements in the regression models. Figure 4 shows the results, with the XGBR models having a perfect $R^2$ CV of 1.000 and exceedingly low RMSECV of 0.00028 for both sets of Shiraz blends. XGBR can clearly predict the blend percentage for each sample, notably with a clear distinction between 0 % blend and 1 % blend for both Shiraz/Cabernet-Sauvignon and Shiraz/Grenache. This was a striking result, highlighting that XGBR modelling of EEM data could be a successful option for detecting the addition of small proportions of different varietal wines. With further development and ultimately the production of databases, it is conceivable that this approach could be applied for robust prediction of the composition of unknown sample blends. In addition, the approach is simple and rapid compared to sensitive DNA techniques (e.g., based on cultivar genotype to determine wine blends), which suffer from reproducibility problems when authenticating experimental or commercial wines (Boccacci et al., 2020).

CONCLUSIONS

The A-TEEM approach with machine learning modelling continued to show promise as an indispensable tool for wine authentication. In this preliminary work, A-TEEM was applied to monovarietal unfinished wine samples collected from different stages of the winemaking process (i.e., post-primary fermentation, post-malolactic fermentation and pre-blending) to investigate the possibility of tracing molecular fingerprints during wine production. PCA separated samples from different origins based on EEM data and subsequent XGBDA modelling could differentiate the samples with 100 % accuracy. Further highlighting the power of the A-TEEM technique, two sets of wine blends (Shiraz/Cabernet-Sauvignon and Shiraz/Grenache) were analysed to model the proportions of wine in the blend (beginning as low as 1 %). Regression models built with PLSR and XGBR were evaluated in terms of correlation coefficient and cross-validation error, with unrivalled accuracy achieved for the XGBR model with $R^2$ CV equal to 1.000 and small RMSECV for both sets of wine blends. Given the possibility of tracing a wine’s origin through production in conjunction with identifying small additions of other wine in a blend, this approach could foreseeably be developed into a robust method and applied in the industry not only for validating the origin of wine but also detecting other aspects of wine fraud.

ACKNOWLEDGEMENTS

We are grateful to Peter Leske, Director of Reiven Winemaking and Paul Grbin, Head Winemaker in the School of Agriculture, Food and Wine, the University of Adelaide, for supplying wine samples. We thank Lara Westaway, the University of Adelaide, for the collection of samples during winemaking. R.K.R.R. appreciates the financial support provided by an Australian Government Scholarship and Wine Australia supplementary scholarship (WA Ph1909). The Australian Research Council Training Centre for Innovative Wine Production (www.ARCwinecentre.org.au; project number IC170100008) is funded by the Australian Government with additional support from Wine Australia, Waite Research Institute and industry partners. The University of Adelaide is a member of the Wine Innovation Cluster.

REFERENCES


by blending wines of different phenolic profiles testified under
blending strategies for wine color modification I: Color improvement


