AUTHENTICATING THE GEOGRAPHICAL ORIGIN OF WINE USING FLUORESCENCE SPECTROSCOPY AND MACHINE LEARNING

Ranaweera K.R. Ranaweera[1], Adam M. Gilmore[2], Dimitra L. Capone[1,3], Susan E.P. Bastian[1,3], David W. Jeffery[1,3, *]

[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, United States
[3]Australian Research Council Training Centre for Innovative Wine Production, The University of Adelaide, PMB 1, Glen Osmond, South Australia 5064, Australia

Corresponding author* david.jeffery@adelaide.edu.au

Article extracted from the presentation held during Enoforum Web Conference (23-25 February 2021)

Wine Fraud

As a luxury product within a global industry worth hundreds of billions of dollars, wine is prone to fraud because of the potential to deceptively obtain considerable amounts of money. There is perhaps no better example of wine fraud than the case of Rudy Kurniawan, who featured in the movie Sour Grapes (Rothwell & Atlas, 2016). In that case, labels, corks, etc. were being faked for bottles of wine sold at auction, but fraud can also involve refilling empty bottles from expensive wine, substitution of grapes from a different variety or region, or using non-permitted additives or processes (that can also potentially introduce a safety issue), with the intention of using something cheaper or of inferior quality than the authentic product. Table 1 provides a summary of the types of fraud that may occur, how easily detected they may be, and the common methods used to potentially detect such instances.

Region of origin and quality

Wines from different regions are viewed with varying levels of prestige and consumers have a growing awareness about the source of the goods they consume (Souza Gonzaga, Capone, Bastian, & Jeffery, 2021). They seek authentic products and are willing to pay a premium for what they perceive is a quality brand or product (Ranaweera, Souza Gonzaga, Capone, Bastian, & Jeffery, 2020). Authenticity is therefore key to brand identity, so it is important that mechanisms are in place to protect reputations. Regulations surrounding geographical indication or protected designation of origin play a role, but so too do testing capabilities such as those listed in Table 1 that help verify the authenticity and provide confidence in the product for consumers.

Region of origin and its associated terroir also affect grape and wine composition, which impacts perceived wine quality and flows through to the sensory traits that consumers experience in wine. This provides the link between product and place of origin that can be exploited through chemical testing methods on the basis of the unique profiles of grape and wine metabolites. Such intricate natural phenomenon is essentially impossible (or too costly) to artificially replicate by a potential fraudster, especially if considering multiple analytes rather than a single component that may be easily manipulated. Thus, approaches that use chemical fingerprinting to explore patterns in the chemical data are of obvious benefit (Waterhouse, Sacks, & Jeffery, 2016a). Furthermore, differences in chemical composition due to variety or vintage can also be detected.

Authentication according to chemical composition

Various techniques can be employed to authenticate wine, ranging from "omics" approaches including genomics, transcriptomics, proteomics, isotopolomics, and metabolomics, with the latter being most often applied to wine (Ranaweera, Souza Gonzaga, Capone, Bastian, & Jeffery, 2020). Targeted or non-targeted options are available, where specific analytes or

marker compounds that vary in concentration are evaluated in the case of the first approach, or chemical fingerprints are recorded and similarities or differences are detected in the second. Whichever is used, multivariate data analysis (chemometrics) provides a useful tool for interrogating the data and building classification models used for authentication.

Table 1. Types of wine fraud that can occur and approaches for its detection (adapted from Waterhouse, Sacks, and Jeffery (2016a)).*

Type of fraud	Representative examples in wine	Relative difficulty in detecting	Common literature approaches
Brand/product label (counterfeiting)	Refilling of authentic bottles or forgery of labels with inexpensive wine and/or simulated products	Easy to medium (if authentic products available, or if comparing wine to non-wine) Difficult (if authentic products unavailable)	Wide range of appropriate techniques, including GC- or HPLC-MS, FT-IR, NMR
Adulteration *Naturally occurring*	Addition of sugar, acid, or water during winemaking where forbidden	Medium	IRMS
Exogenous	Addition of diethylene glycol to Austrian wines to increase body	Easy (if expected) Difficult (if unexpected)	GC- or HPLC-MS
Process	Use of unauthorised processing aids or winemaking techniques, e.g., the use of micro-oxygenation in regions where forbidden	Medium to difficult	Not commonly evaluated
Species/variety	Substitution of less expensive Merlot and other red grapes for Pinot noir in Languedoc (France)	Easy (grapes) Medium to difficult (wine)	Wide range of appropriate techniques, including GC- or HPLC-MS, FT-IR, NMR
Place of origin	Substitution of less expensive Central Valley grapes for Napa grapes in California	Medium to difficult	Isotopic (IRMS) and elemental (ICP-MS) analyses

* FT-IR, Fourier transform infrared; GC, gas chromatography; HPLC, high performance liquid chromatography; ICP-MS, inductively coupled plasma mass spectrometry; IRMS, isotope ratio mass spectrometry; MS, mass spectrometry; NMR, nuclear magnetic resonance.

Analytical techniques tend to focus on specific compound categories such as volatiles, phenolics, amino acids, elements or isotopes (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021). This involves sophisticated and expensive instrumentation such as HPLC- or GC-MS, IRMS, or ICP-MS, which is not really implementable in a winery or supply chain setting. As such, spectroscopic techniques are appealing as they tend to be rapid, simple, robust, and cost-effective. Typically, these include UV, visible, Raman, and near- or mid-infrared spectroscopy along with chemometrics to analyse the multivariate data (Ranaweera, Souza Gonzaga, Capone, Bastian, & Jeffery, 2020). Fluorescence spectroscopy also fits this category and is attractive due to being highly sensitive, as well as having the other desirable

traits associated with spectroscopic methods. There is a range of fluorophores in wine, such as phenolics, aromatic amino acids, and vitamins that are detected according to certain excitation and emission wavelengths, producing a unique molecular fingerprint of the sample in the form of an excitation-emission matrix (EEM) (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021). Example EEMs are shown in Figure 1 for two Shiraz wines, where even a cursory glance can reveal some differences in the contours, for example at around 255, 275 and 325 nm for excitation, and between 320 and 375 nm for emission.
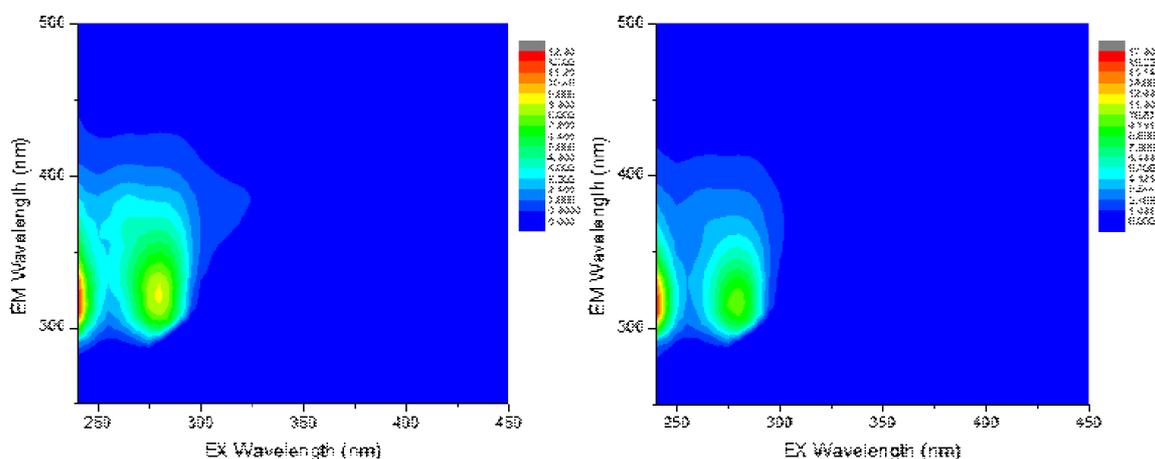

Figure 1. EEM contour plots for Shiraz wines from two geographical indications in South Australia.

Development of an authentication method

The benefits of a fluorescence approach were apparent after reviewing options in the literature, and this was the starting point for the development of a method for authentication of wine. We anticipated that EEM data in conjunction with suitable chemometric modelling could discriminate wines from different regions, and in the first instance we focused on commercial Cabernet Sauvignon wines produced in three Australian regions and in Bordeaux. We evaluated different machine learning algorithms and stepwise discriminant analysis (DA) to model the data, and compared the classification accuracy from an EEM approach against ICP-MS for the same set of wines (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021).

ICP-MS analysis of elemental composition

Of the 65 elements measured using direct dilution and ICP-MS, 25 differed significantly (p ≤ 0.035) according to wine region (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021). These elements were generally in common with a previous study that used elemental composition for regional discrimination of Australian wines (Martin, Watling, & Lee, 2012), with some variation in concentration for a given region. Principal component analysis (PCA) of significantly different elements revealed a slight separation of regions and highlighted the most characteristic elements (Figure 2a). Coonawarra wines were relatively more abundant in strontium (Sr), lithium (Li), and sodium (Na); Yarra Valley wines had caesium (Cs), rubidium (Rb) and barium (Ba); Margaret River wines had magnesium (Mg); and Bordeaux wines were characterised by elements like lead (Pb), cerium (Ce), titanium (Ti), arsenic (As), and beryllium (Be). Although mineral content can be used to relate a wine to its region of origin, it can also be affected by multiple oenological factors (Waterhouse, Sacks, & Jeffery, 2016b).

Chemometrics with ICP-MS data

PCA data for the elements was used to build a classification model using DA, which showed separation of Bordeaux samples from the others, but overlap of the Australian regions, especially Coonawarra and Margaret River (Figure 2b). DA modelling was compared to extreme gradient boosting (XGB) DA, a highly effective machine learning technique in terms of predictive performance that can cope with heterogeneous or imbalanced class distributions

(Babajide Mustapha & Saeed, 2016). Results of the modelling are shown in Table 2. DA performed similarly to other studies, with > 90% accuracy (94.2% total): there was a small number of misclassified wines for the Australian regions but 100% correct classification for wines from Bordeaux. When applying XGBDA to the element data, the classification accuracy for the Australian wines improved slightly (97.7% total, two misclassified wines), driven by ~5% increases for Coonawarra and Margaret River (to 100% in this case), whereas Yarra Valley and Bordeaux remained the same as with DA (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021).
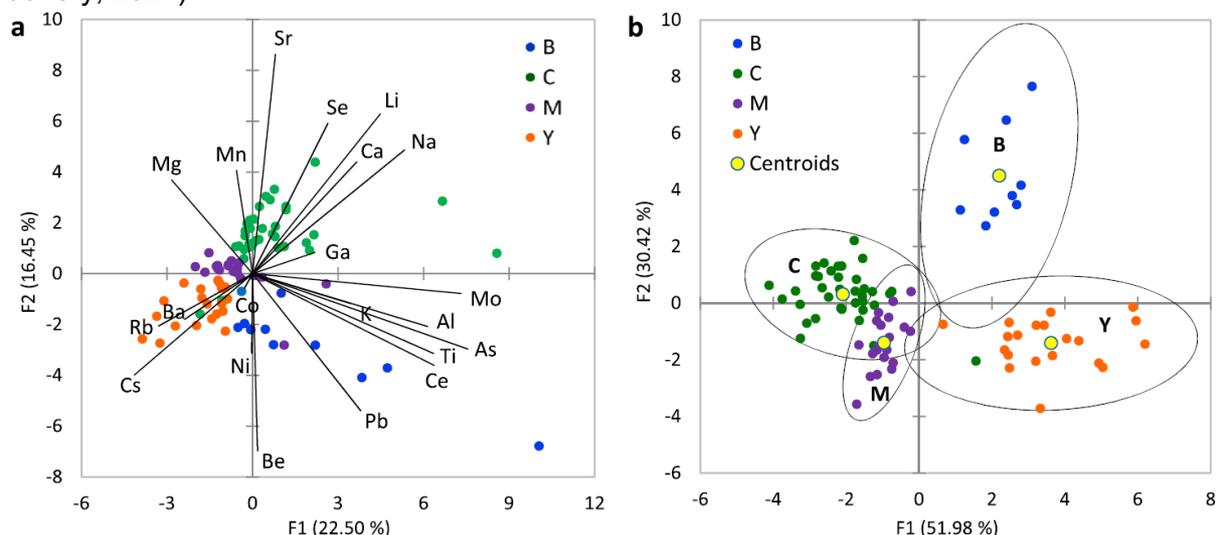

Figure 2. (a) PCA biplot of significant elements from ICP-MS for the different regions and (b) DA classification of regions (with 95% confidence ellipses) based on PCA results after varimax rotation. Reprinted from Food Chemistry, 335, Ranaweera K.R. Ranaweera, Adam M. Gilmore, Dimitra L. Capone, Susan E.P. Bastian, David W. Jeffery, Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling, 127592, Copyright (2021), with permission from Elsevier.

Table 2. Confusion matrix for cross-validated models from stepwise discriminant analysis (DA) and extreme gradient boost discriminant analysis (XGBDA) using element data from ICP-MS.*

		DA					XGBDA					
Predicted Actual	B	C	M	Y	Total	Correct %	B	C	M	Y	Total	Correct %
B	10	0	0	0	10	100.00	10	0	0	0	10	100.00
C	0	33	2	1	36	91.67	0	35	0	1	36	97.20
M	0	1	19	0	20	95.00	0	0	20	0	20	100.00
Y	0	0	1	19	20	95.00	1	0	0	19	20	95.00
Total	10	34	22	20	86	94.19	11	35	20	20	86	97.67

*Table reprinted from Food Chemistry, 335, Ranaweera K.R. Ranaweera, Adam M. Gilmore, Dimitra L. Capone, Susan E.P. Bastian, David W. Jeffery, Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling, 127592, Copyright (2021), with permission from Elsevier.

Chemometrics with EEM data

The authentication results with ICP-MS were very encouraging and attention was turned to classification modelling with EEM fingerprints. Machine learning models were the focus, with support vector machine (SVM) DA and XGBDA being evaluated (Table 3). Different pre-processing and data reduction options were used to obtain the highest classification probability: PCA compression was used for SVDMA and partial least squares regression was

used for XGBDA (Ranaweera, Gilmore, Capone, Bastian, & Jeffery, 2021). SVMDA with EEM data performed relatively poorly in comparison to using elements by ICP-MS, yielding 90% accuracy at best, for wines from Bordeaux, and a low of 72.5% for Yarra Valley. The overall correct classification of 84.7% (Table 3) compared reasonably well with other work using SVM with Cabernet Sauvignon wines (from Brazil and Chile), albeit slightly lower than the reported 89% (da Costa, Castro, & Barbosa, 2016). In contrast, XGBDA with EEM data performed exceptionally well, yielding 100% classification accuracy and highlighting the potential for using fluorescence spectroscopy as a platform for authenticity testing. The vast improvement over SVDMA may reflect the in-built parallel processing capability and tree pruning effectiveness of XGBDA. When considering EEM fingerprints versus ICP-MS analysis of elements, both the instrumentation and technical requirements for obtaining EEM are much more favourable in terms of having lower complexity and cost.

Table 3. Confusion matrix for cross-validated models from support vector machine discriminant analysis (SVMDA) and extreme gradient boost discriminant analysis (XGBDA) using fluorescence data (samples run in duplicate).*

	SVMDA						XGBDA					
Predicted / Actual	B	C	M	Y	Total[a]	Correct %	B	C	M	Y	Total[a]	Correct %
B	18	0	2	0	20	90.00	20	0	0	0	20	100.00
C	0	64	2	6	72	88.88	0	72	0	0	72	100.00
M	0	0	33	5	38	86.84	0	0	38	0	38	100.00
Y	0	7	4	29	40	72.50	0	0	0	40	40	100.00
Total	18	71	41	40	170	84.70	20	72	38	40	170	100.00

*Table reprinted from Food Chemistry, 335, Ranaweera K.R. Ranaweera, Adam M. Gilmore, Dimitra L. Capone, Susan E.P. Bastian, David W. Jeffery, Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling, 127592, Copyright (2021), with permission from Elsevier.

Conclusion

Wine fraud is a serious concern for the international wine industry, necessitating techniques that can authenticate wine based on provenance, variety, and vintage. Among the range of approaches that could be taken for regional authentication, this study focused on wine chemical composition, and more specifically, utilised fluorescence spectroscopy along with chemometrics to build classification models. Being a relatively standard approach, ICP-MS was evaluated as a reference method, yielding classification results that compared well to other studies (> 90% accuracy). As with a range of other advanced analytical techniques, however, ICP-MS cannot feasibly be introduced into the production and supply chain. In contrast, fluorescence spectroscopy offers the benefits of being simple, rapid, and cost-effective, and the use of EEM fingerprints in conjunction with XGBDA as a machine learning algorithm provided an unsurpassable level of 100% accuracy. The outcomes provide confidence in pursuing the use of EEMs with machine learning for the geographical authentication of wine. The next steps will involve expanding this approach to other varieties and regions, and exploring the impact of production practices.

Abstract

Wine is a luxury product and a global beverage steeped in history and mystery. Over time, various regions have become renowned for the quality of wines they produce, which adds considerable value to the regions and the brands. On the whole, the international wine market is worth many hundreds of billions of dollars, which attracts unscrupulous operators intent on defrauding wine consumers. Countering such fraudulent activities requires the means to test and classify wine, but the task is considerable due to the complexity of wine. However, just as wine origin influences chemical and sensory profiles, indicators of wine provenance are naturally embedded in the chemical composition of wine. A range of methods of varying intricacy are available to analyse wine for authentication of variety or geographical origin. Instruments and techniques within the domain of research laboratories are not so practical or deployable in winery or supply chain settings, however. This is where spectroscopic methods are attractive, as they can be rapid, cost-effective and simple. In the search for such a method, we identified fluorescence spectroscopy, and more specifically, the collection of an excitation-emission matrix (EEM) that acts like a molecular fingerprint. Multivariate statistical modelling is then used in conjunction with the EEM data to develop classification models for wines from various regions. We have developed such a technique, using a relatively new type of machine learning algorithm known as extreme gradient boosting discriminant analysis. This unique approach, which can routinely achieve a level of accuracy of 100% in comparison to ICP-MS at an average of 85%, is being applied to a range of studies on Shiraz and Cabernet Sauvignon wines from different regions of Australia.

Acknowledgements

References

Babajide Mustapha, I., & Saeed, F. (2016). Bioactive molecule prediction using extreme gradient boosting. *Molecules, 21*(8), 983.

da Costa, N. L., Castro, I. A., & Barbosa, R. (2016). Classification of Cabernet Sauvignon from two different countries in South America by chemical compounds and support vector machines. *Applied Artificial Intelligence, 30*(7), 679-689.

Martin, A. E., Watling, R. J., & Lee, G. S. (2012). The multi-element determination and regional discrimination of Australian wines. *Food Chemistry, 133*(3), 1081-1089.

Ranaweera, R. K. R., Gilmore, A. M., Capone, D. L., Bastian, S. E. P., & Jeffery, D. W. (2021). Authentication of the geographical origin of Australian Cabernet Sauvignon wines using spectrofluorometric and multi-element analyses with multivariate statistical modelling. *Food Chemistry, 335*, 127592.

Ranaweera, R. K. R., Souza Gonzaga, L., Capone, D. L., Bastian, S. E. P., & Jeffery, D. W. (2020). Authenticity and Traceability in the Wine Industry: From Analytical Chemistry to Consumer Perceptions. In *Reference Module in Food Science* 1st ed., (pp. 1-29): Elsevier.

Rothwell, J., & Atlas, R. (2016). Sour Grapes. In). United States: Dogwoof.

Souza Gonzaga, L., Capone, D. L., Bastian, S. E. P., & Jeffery, D. W. (2021). Defining wine typicity: sensory characterisation and consumer perspectives. *Australian Journal of Grape and Wine Research, Early View*.

Waterhouse, A. L., Sacks, G. L., & Jeffery, D. W. (2016a). Authentication. In *Understanding Wine Chemistry*, (pp. 383-391). Chichester, UK: John Wiley & Sons, Ltd.

Waterhouse, A. L., Sacks, G. L., & Jeffery, D. W. (2016b). Minerals. In *Understanding Wine Chemistry*, (pp. 34-39). Chichester, UK: John Wiley & Sons. Ltd.