Received: 29 September 2020 | Accepted: 20 October 2020 | Published: 27 November 2020
DOI:10.20870/oeno-one.2020.54.4.4501

OENO One
VINE AND WINE OPEN ACCESS JOURNAL

Detection of smoke-derived compounds from bushfires in Cabernet-Sauvignon grapes, must, and wine using Near-Infrared spectroscopy and machine learning algorithms

Vasiliki Summerson[1], Claudia Gonzalez Viejo[1], Damir D. Torrico[2], Alexis Pang[1] and Sigfredo Fuentes[1,*]

[1]Digital Agriculture Food and Wine Group, School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Building 142, Parkville 3010, Victoria, Australia
[2]Department of Wine, Food and Molecular Biosciences, Faculty of Agriculture and Life Sciences, Lincoln University, Lincoln 7647, Canterbury, New Zealand

*corresponding author: sigfredo.fuentes@unimelb.edu.au

ABSTRACT

The number and intensity of wildfires are increasing worldwide, thereby raising the risk of smoke contamination of grapevine berries and the development of smoke taint in wine. This study aimed to develop five artificial neural network (ANN) models from berry, must, and wine samples obtained from grapevines exposed to different levels of smoke: (i) Control (C), i.e., no misting or smoke exposure; (ii) Control with misting (CM), i.e., in-canopy misting, but no smoke exposure; (iii) low-density smoke treatment (LS); (iv) high-density smoke treatment (HS) and (v) a high-density smoke treatment with misting (HSM). Models 1, 2, and 3 were developed using the absorbance values of near-infrared (NIR) berry spectra taken one day after smoke exposure to predict levels of 10 volatile phenols (VP) and 18 glycoconjugates in grapes at either one day after smoke exposure (Model 1: R = 0.98; R2 = 0.97; b = 1) or at harvest (Model 2: R = 0.98; R2 = 0.97; b = 0.97), as well as six VP and 17 glycoconjugates in the final wine (Model 3: R = 0.98; R2 = 0.95; b = 0.99). Models 4 and 5 were developed to predict the levels of six VP and 17 glycoconjugates in wine. Model 4 used must NIR absorbance spectra as inputs (R = 0.99; R2 = 0.99; b = 1.00), while Model 5 used wine NIR absorbance spectra (R = 0.99; R2 = 0.97; b = 0.97). All five models displayed high accuracies and could be used by grape growers and winemakers to non-destructively assess at near real-time the levels of smoke-related compounds in grapes and/or wine in order to make timely decisions about grape harvest and smoke taint mitigation techniques in the winemaking process.

KEYWORDS

remote sensing, climate change, artificial neural networks, smoke taint

Supplementary data can be downloaded through: https://oeno-one.eu/article/view/4501

INTRODUCTION

Recent climate change forecasts have predicted an increase in the number and intensity of wildfires, as well as a lengthening of the fire season in many grape-growing regions throughout the world, including Australia, Greece, California, Chile, and South Africa (CSIRO & Australian Government Bureau of Meteorology, 2018; Favell *et al*., 2019; Fuentes *et al*., 2019; Hughes & Alexander, 2017; Noestheden *et al*., 2018b; Simos, 2008). As a consequence, the incidence of grapevine smoke exposure and the subsequent development of objectional smoky aromas in wine known as smoke taint is also likely to increase, resulting in significant financial losses to the wine industry (Bell *et al.*, 2013; Kelly *et al*., 2012; Kennison *et al*., 2011; Noestheden *et al*., 2018a; Noestheden *et al*., 2017).

It has been found that the exposure of grapevines to smoke during the critical period of approximately seven days post-veraison - which also corresponds to the period at highest risk of wildfire development - leads to the exposure of fruit to volatile phenols (VP), including guaiacol, 4-methylguaiacol, syringol and cresols, which then accumulate in glycoconjugate forms in grape berries and leaves (Fuentes & Tongson, 2017; Ristic *et al*., 2017; van der Hulst *et al*., 2019). During fermentation, these glycoconjugates are hydrolysed back into their free, sensorially active forms, which express smoky aromas in the produced wine to the point of taint; however, a significant pool of glycoconjugates remains in the wine. It is believed that both the free VP and their glycoconjugates contribute to the smoky aromas (such as 'burnt', 'earthy', and 'ashy' notes) in wine; therefore, they must both be measured in order to ascertain the level of smoke taint (Hayasaka *et al*., 2013; Kennison *et al*., 2008; Kennison *et al*., 2007; Ristic *et al*., 2017; Singh *et al*., 2011; van der Hulst *et al*., 2019). While low levels of these compounds may add complexity to wine flavour and aroma, high levels above the detection threshold result in the unpleasant aromas associated with smoke taint (Singh *et al*., 2011). Some aroma detection thresholds for VP in red wines have been reported, with guaiacol, m-cresol, and 4-methylguaiacol (believed to be key contributors to smoke-taint aromas) having the lowest detection thresholds of all the VP at 23 µg/L, 20 µg/L, and 30 µg/L respectively (Härtl & Schwab, 2018; Parker *et al*., 2013). In comparison, wines matured in oak barrels reportedly contain 10–100 µg/L guaiacol and 1–20 µg/L 4-methylguaiacol (Pollnitz *et al.*, 2004). Others have also reported that a possible synergistic effect may occur when smoke-taint characteristics are perceived, despite individual concentrations of VP being below detection thresholds (De Vries *et al*., 2016; Kennison *et al*., 2009). These heavily smoke-tainted wines are unpalatable and unprofitable, costing the wine industry millions of dollars in lost wine revenue. For example, it is estimated that the 2006/2007 bushfires in Victoria, Australia, resulted in approximately 75-90 billion AU$ in lost revenue, while the 2009 Black Saturday bushfires resulted in a loss of approximately 300 million AU$ to the local Australian wine industry (Department of Primary Industries, 2009; Favell *et al*., 2019; Fudge *et al*., 2011; Kennison *et al*., 2007; Noestheden *et al*., 2017; Singh *et al*., 2011).

In order to ascertain the level of smoke taint in wine, concentrations of both free VP and bound glycoconjugates must be determined (Allen *et al.*, 2013; van der Hulst *et al*., 2019). Techniques like gas chromatography-mass spectrometry (GC-MS) and high-performance liquid chromatography (HPLC) are often used for quantifying levels of free and glycosidically bound VP in both grapes and wine (Hayasaka *et al*., 2010a; Hayasaka *et al.*, 2010b; Hayasaka *et al*., 2013; Pollnitz *et al*., 2004). Unfortunately, there are several drawbacks to these chromatographic techniques; for example, time-consuming sample preparation, reagent and instrumentation costs, need for trained personnel, destructive sampling techniques. In addition, there is often a long waiting time between taking the sample and obtaining the results (Fudge *et al*., 2012b, Fudge *et al.*, 2013; Kemps *et al.*, 2010): growers may face significant delays in receiving results for smoke contamination analysis from commercial laboratories, particularly during affected periods or the winemaking process, when there are many samples requiring analysis from affected vineyards within a region (Fudge *et al*., 2012b; Fudge *et al*., 2013). Therefore, research is required on alternative methods of detection that provide rapid results and hence allow immediate action to be taken to reduce affected grapes or to modify the winemaking processes.

The use of spectroscopic methods for quantitative and qualitative analysis has gained popularity due to their rapid results, ease of use, non-destructive nature permitting repeated measurements, and the portability of devices for in-field use (Fudge *et al*., 2012b; Fudge *et al*., 2013; Hall, 2018; Kemps *et al*., 2010; Teixeira dos Santos *et al*., 2013). Some spectroscopic techniques in the ultraviolet

(UV; 250-400 nm), visible (Vis; 400-700 nm), near-infrared (NIR; 700-2500 nm), and mid-infrared (MIR; 2500-25000 nm) regions have been used for various grape and wine assessments, including the classification of grape juice based on the grape variety, the determination of polyphenolic compounds in red wines, and the assessment of the aroma potential of Tannet grapes by measuring their glycosylated aroma compound content (Boido *et al*., 2013; Cozzolino *et al*., 2012; Martelo-Vidal & Vázquez, 2014; Pirie *et al.*, 2005). Most notably, the use of MIR spectroscopy has demonstrated great potential in classifying smoke-tainted wines; however, it does not provide details about the levels of glycoconjugates and VP in the wine, and classification rates have been found to be impacted by grape variety, oak maturation and degree of smoke taint (Fudge *et al*., 2012b).

Increasing research is being conducted to investigate the use of artificial neural networks (ANNs) for the analysis of UV-Vis-NIR spectroscopy, particularly since ANNs are better able to analyse non-linear data than conventional chemometric techniques, such as principal component analysis (PCA) (Diamantopoulou & Milios, 2010; Martelo-Vidal & Vázquez, 2015; Yu *et al.*, 2018). ANN models can predict the physical properties of food products from a learning algorithm trained using experimentally-derived data or values from validated real-life models (Dieulot & Skurtys, 2013; Martelo-Vidal & Vázquez, 2015). Coupled with UV-Vis-NIR spectroscopy, ANN has been used to develop reliable models for predicting levels of malic acid, tartaric acid, and ethanol in wine (Martelo-Vidal & Vázquez, 2015). There has also been significant research investigating the use of ANNs for the detection of grapevine smoke contamination and smoke taint compounds in wine. Using grapevine berry and leaf NIR readings as inputs, ANN models have been developed to classify the spectral readings according to smoke exposure levels with high accuracy (Summerson *et al*., 2020). Other ANN models have been developed using readings obtained from a low-cost electronic nose (E-nose) to accurately predict levels of smoke compounds in wines (Fuentes *et al*., 2020). Furthermore, other research has used NIR spectroscopy within the region of 700-1100 nm to develop a model for quantifying levels of guaiacol glycoconjugates in berries and wine and levels of guaiacol in wine (Fuentes *et al*., 2019).

This paper presents the use of NIR spectroscopy as a rapid method for assessing grapes, must, and wine with different levels of smoke contamination and with or without in-canopy misting. Five ANN regression models were developed to predict levels of smoke taint markers in grape berries or wine. All five models had high accuracy and could be used by grape growers and winemakers as a non-destructive method for assessing near real-time smoke contamination levels in grapes and wine. This would allow them to make timely decisions about which fruit to sample for further chemical analysis and/or which fruit to harvest to maintain wine quality or to apply smoke taint mitigation techniques to must/wine, such as through the use of activated carbon.

MATERIALS AND METHODS

1. Field application of smoke treatments to grapevines and winemaking

The experimental site is located at the University of Adelaide's Waite campus in Urrbrae, South Australia (34°58'S, 138°38'E). The experiment was conducted during the 2018/2019 season, as previously described by Szeto *et al*. (2020) and Summerson *et al*. (2020). The grapevines were planted at 2.0 and 3.3 m spacing between vines and rows, respectively, and trained to a bilateral cordon and vertical shoot-positioned trellis system (VSP). Grapevines were hand-pruned to a two-node spur system, with drip irrigation under the vine (twice weekly for 6 hours from fruit set to pre-harvest) with a drip distance of 0.75 m and a dripper water rate of 1.6 L/h (8.5 mm/week). Three smoke treatments with or without in-canopy misting were applied to five to six Cabernet-Sauvignon grapevines (within two adjacent panels) at approximately seven days post-veraison. Smoke treatments consisted of i) low-density smoke without misting (LS), ii) high-density smoke with misting (HSM), and high-density smoke without misting (HS). There were also two control treatments: i) control without misting (C) and control with misting (CM). Each treatment consisted of six adjacent grapevines. Smoke treatments involved the application of smoke for one hour using purpose-built tents (6 m x 2.5 m x 2 m) constructed from galvanised steel framing and greenhouse-grade Solarweave plastic (Gale Pacific, Australia) that enables plant photosynthesis under experimental conditions similar to those described previously (Kennison *et al*., 2008; Kennison *et al*., 2009, Kennison *et al*., 2011; Ristic *et al*., 2016). Vines were enclosed in the tents with openings at the

extremities to avoid damaging the vines. Low and high-density smoke exposure was achieved by burning approximately 1.5 and 5.0 kg of barley straw, respectively. Control vines were separated by at least one buffer vine so as not to be exposed to smoke. Misting treatments involved the continuous application of fine water droplets (65 µm) to the grapevine bunch zone, using a purpose-built sprinkler system (delivering water at 11 L/h), as previously described (Caravia *et al*., 2017; Szeto *et al*., 2020). Average ambient temperature and humidity were 31 °C and 36 %, respectively, with minimum cloud coverage.

The wine was produced on a small-scale using approximately 5 kg bunches per fermentation, performed in triplicate for each treatment, as set out by Szeto *et al*. (Szeto *et al*., 2020). Bunches were de-stemmed and crushed, 50 mg/L of sulphur dioxide (SO_2) was added to the must, and pH then adjusted to 3.5 with tartaric acid. The must was then inoculated with PDM yeast (150 mg/mL; Maurivin, AB Biotek, Sydney, NSW, Australia) and fermented with skins. Fermentation took place at ambient temperature (25-27 °C), and the cap was plunged twice daily until wines approached dryness; they were then pressed and held at 25 °C until fermentation was complete. The wines did not undergo malolactic fermentation. The wines were then racked from gross lees and cold stabilised before the pH, and free SO_2 were adjusted (3.5 and 30 mg/L, respectively) prior to bottling.

2. Chemical analysis of volatile phenols and their glycoconjugates in grape juice/homogenate and wine

Levels of VP and their glycoconjugates were determined in grape juice/homogenate, as well as in the final wine (Table 1) using previously published stable isotope dilution analysis (SIDA) methods (Hayasaka *et al*., 2010a; Y. Hayasaka *et al*., 2013; Pollnitz *et al*., 2004; Szeto *et al*., 2020). Volatile phenols were measured using an Agilent 6890 gas chromatograph coupled to a 5973 mass spectrometer (Agilent Technologies, Forest Hill, Vic., Australia), with isotopically labelled standards of d4-guaiacol and d3-syringol prepared by the Australian Wine Research Institute's (AWRI) Commercial Services Laboratory in Adelaide, Australia as previously reported (Dungey *et al.*, 2011; Hayasaka *et al*., 2010a; Pollnitz *et al*., 2004). The limit of quantitation for volatile phenols was 1-2 µg/L. Volatile phenol glycoconjugates were measured using liquid chromatography-tandem mass spectrometry (HPLC–MS/MS) applying previously outlined methods (Hayasaka *et al*., 2010a; Hayasaka *et al*., 2013). An Agilent 1200 high-performance liquid chromatograph (HPLC) was used. It was equipped with a 1290 binary pump, coupled to an AB SCIEX Triple QuadTM 4500 tandem mass spectrometer with a Turbo VTM ion source (Framingham, MA, USA). An isotopically labelled internal standard of d3-syringol gentiobioside was prepared according to published methods (Hayasaka *et al*., 2010a; Hayasaka *et al*., 2013). The limit of quantitation for volatile phenol glycosides was 1 µg/kg.

3. NIR absorbance patterns for grape berries, must, and wine samples

Grape berry spectra were collected 24 hours after smoke treatments were applied, as previously described by Summerson *et al*. (2020), using the microPHAZIR™ RX Analyzer (Thermo Fisher Scientific, Waltham, MA, USA), which had a spectral range of 1596 to 2396 nm at intervals of 7-9 nm. The device's calibration was carried out as required using a white background calibration standard (included with the device) prior to starting and after every ten readings. From each treatment, two vines were selected for analysis. Two bunches were selected from each vine, and nine berries were measured in triplicate (36 berries per treatment measured three times each; total n = 540). All measurements were conducted at ambient temperature between 9:00 and 18:00 with berries still attached to the bunch.

Must and wine samples were measured using a modified procedure previously described by Gonzalez Viejo *et al*. (2018). A Whatman® filter paper (Whatman plc., Maidstone, UK) of quality grade three and 7.0 cm diameter was used as the holding medium for the must and wine samples. The dry filter paper was first analysed using the microPHAZIR™ RX Analyzer by placing it directly onto the front of the 5 mm measuring region and placing the white background standard behind it to prevent signal noise inclusion due to environmental factors. The filter paper was then submerged and soaked with the specific must/wine sample and analysed immediately. The spectral readings obtained from the dry filter paper were then subtracted from the readings obtained from the filter paper soaked in the specific sample to obtain only the sample spectral reading results. Each treatment was measured three times in triplicate for both must and wine samples (total n = 45), with measurements conducted at room temperature (20-23 °C).

TABLE 1. List of volatile phenols and their glycoconjugates used as inputs.

	Smoke compound	Abbreviation	Sample tested
Volatile phenols	guaiacol	Gu	Berries/Must/Wine
	4-methylguaiacol	4MG	Berries/Must/Wine
	phenol	Ph	Berries
	syringol	Sy	Berries/Must/Wine
	4-methylsyringol	Msy	Berries
	m-cresol	m-Cr	Berries/Must/Wine
	p-cresol	p-Cr	Berries/Must/Wine
	Total m/p-cresol	m/p-Cr	Berries
	o-cresol	o-Cr	Berries/Must/Wine
	total cresols	Cr	Berries
Glycoconjugates	Guaiacol pentosylglucosides	GuPG	Berries/Must/Wine
	Guaiacol gentiobioside	GuGG	Berries/Must/Wine
	Guaiacol rutinoside	GuRG	Berries/Must/Wine
	Guaiacol monoglucoside	GuMG	Berries/Must/Wine
	Methylguaiacol pentosylglucosides	MGuPG	Berries/Must/Wine
	Methylguaiacol rutinoside	MGuRG	Berries/Must/Wine
	Phenol rutinoside	PhRG	Berries/Must/Wine
	Phenol gentiobioside	PhGG	Berries/Must/Wine
	Phenol pentosylglucosides	PhPG	Berries/Must/Wine
	Phenol monoglucoside	PhMG	Berries/Must/Wine
	Syringol gentiobioside	SyGG	Berries/Must/Wine
	Syringol monoglucoside	SyMG	Berries/Must/Wine
	Syringol pentosylglucosides	SyPG	Berries/Must/Wine
	Methylsyringol gentiobioside	MSyGG	Berries/Must/Wine
	Methylsyringol pentosylglucosides	MSyPG	Berries/Must/Wine
	Cresol glucosylpentosides	CrPG	Berries/Must/Wine
	Cresol gentiobioside	CrGG	Berries
	Cresol rutinoside	CrRG	Berries/Must/Wine

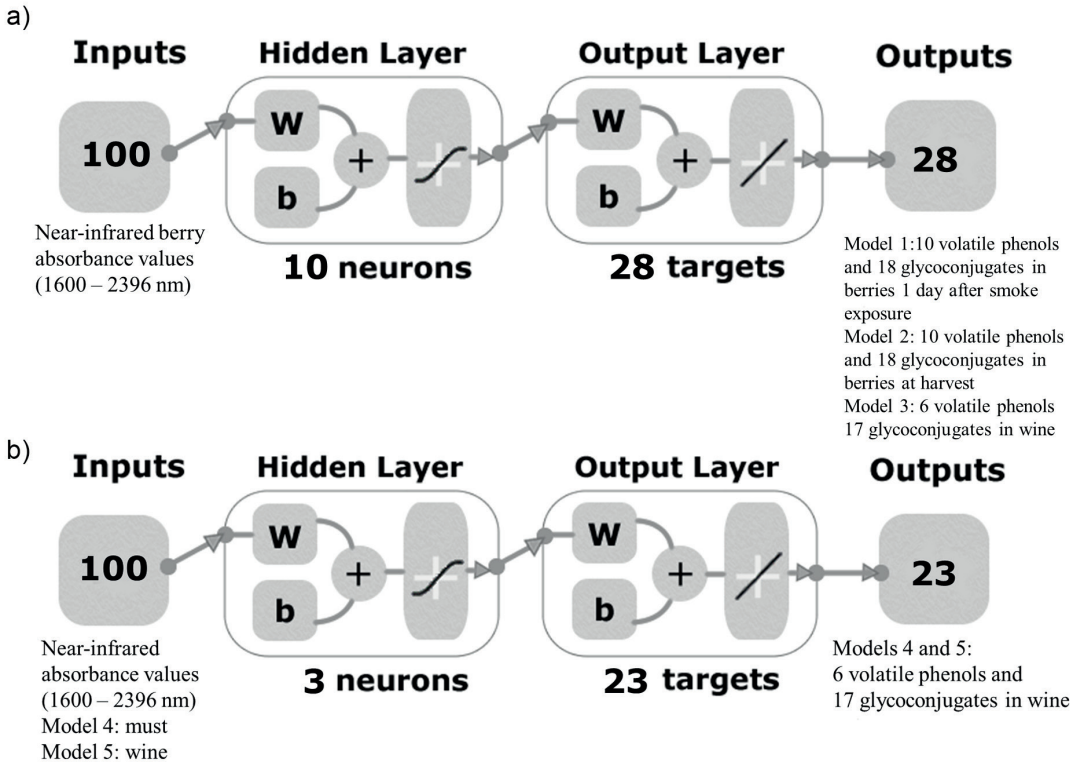

FIGURE 1. Two-layer feedforward networks for (a) Models 1 and 2 to predict levels of 10 volatile phenols and 18 glycoconjugates in grape berries and (b) Models 3, 4, and 5 to predict levels of six volatile phenols and 17 glycoconjugates in the final wine.

4. Machine learning modelling and statistical analysis

Five ANN machine learning regression models were developed using a customised code written in MATLAB® R2020b (Mathworks, Inc., Natick, MA, USA) to test 17 different training algorithms. The best results for the first three models were obtained using the Levenberg Marquardt algorithm to predict levels of volatile phenols and their glycoconjugates in grape berries one day after smoke exposure (Model 1) and at harvest (Model 2), as well as in the wine (Model 3). Berry NIR readings taken one day after smoke exposure were used as inputs for these three models, with ten volatile phenols and 18 glycoconjugates used as targets in Models 1 and 2, and six volatile phenols and 17 glycoconjugates used for Model 3, as shown in Table 1. The Bayesian regularisation algorithm was chosen for Models 4 and 5 to predict levels of 17 glycoconjugates and six volatile phenols in the wine. Model 4 used the must NIR spectra as inputs, while Model 5 used the wine NIR spectra (Figure 1b).

A random data division was used to develop all five models. Models 1, 2, and 3 used 70 % of the data for the training stage, 15 % for validation with a mean squared error (MSE) performance algorithm, and 15 % for testing, while Models 4 and 5 used 70 % for training and 30 % for testing. All ANN models consisted of a two-layer feedforward network with the hidden layer using a tan-sigmoid function and the output layer using a linear transfer function, as shown in Figure 1. Ten hidden neurons were selected for Models 1 to 3 and seven for Models 4 and 5 after conducting a trimming exercise with three, seven, and ten neurons to see which yielded the best performance. Models were assessed for over- or underfitting using statistical data consisting of the correlation coefficient (R), slope (b), MSE, and determination coefficient (R^2).

RESULTS

1. NIR absorbance spectra for berries, must, and wine

Figures 2, 3, and 4 illustrate the average berry, must, and wine NIR absorbance spectra. Clear differences in spectral readings across the entire

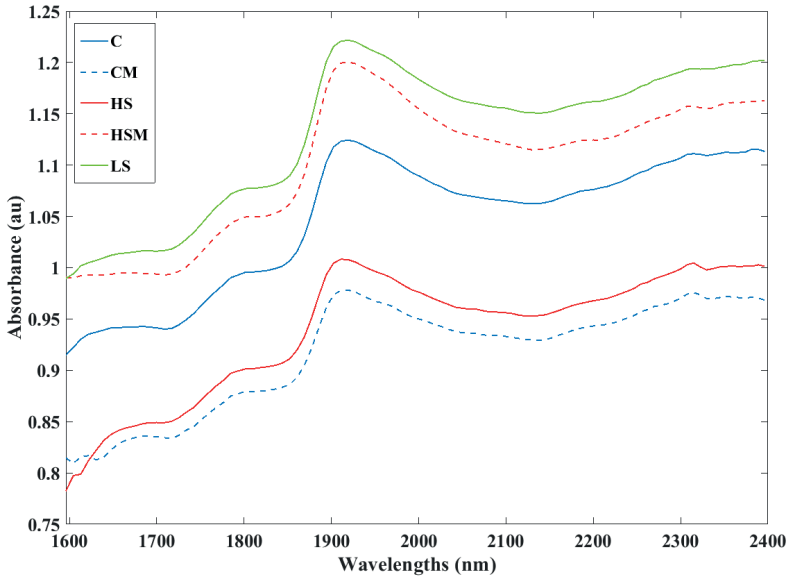

FIGURE 2. Raw berry absorbance spectra for the five smoke treatments.
Abbreviations: C = control without misting; CM = control with misting; HS = high-density smoke without misting; HSM = high-density smoke with misting; and LS = low-density smoke

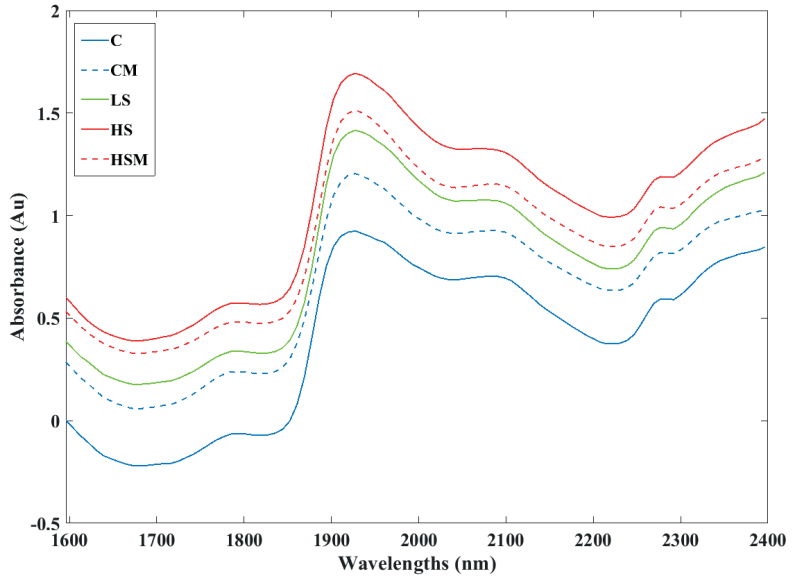

FIGURE 3. Raw must absorbance spectra for the five smoke treatments.
Abbreviations: C = control without misting; CM = control with misting; HS = high-density smoke without misting; HSM = high-density smoke with misting; and LS = low-density smoke

wavelength range for each smoke treatment was observed for the berry absorbance spectra (Figure 2a). A large peak was originally observed at approximately 1910 nm, as well as a smaller one at approximately 1790 nm. In the transformed data (Figure S1), large peaks were observed between approximately 1596-1650 nm and 1820-1950 nm.

Again, clear differences in spectral readings across the entire wavelength range for each smoke treatment were observed for the must absorbance spectra (Figure 3). A large peak was originally observed at approximately 1927 nm and two smaller peaks at approximately 1784 and 2090 nm (Figure 3a), while for the transformed data, large peaks could be seen between approximately 1818-1902 nm and 2246-23322 nm (Figure S2).

For the wine absorbance spectra (Figure 4), the largest difference in absorbance values amongst the different smoke treatments was observed at the 1927 nm peak or overtone. Further peaks were also observed at approximately 2090, 2270,

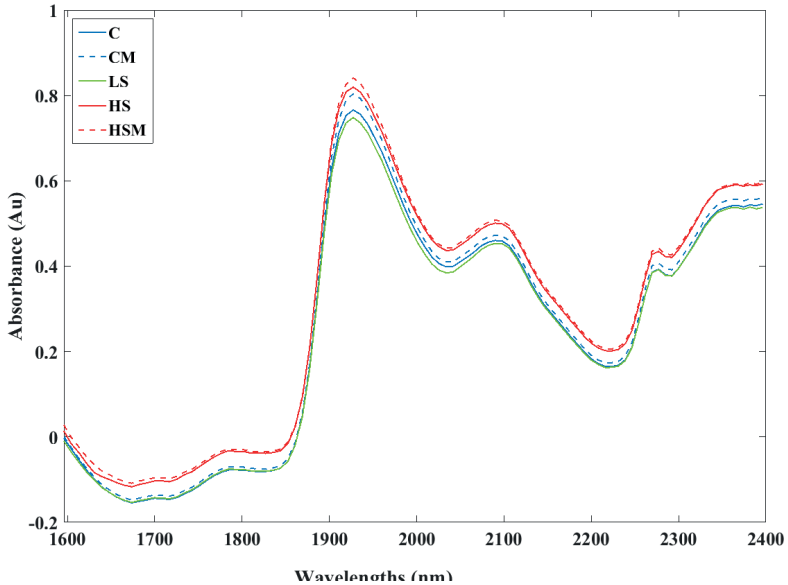

FIGURE 4. Raw wine absorbance (a) and second derivative spectra (b) for the five smoke treatments.
Abbreviations: C = control without misting; CM = control with misting; HS = high-density smoke without misting; HSM = high-density smoke with misting; and LS = low-density smoke

and 2340 nm, while in the transformed data (Figure S3), large peaks were observed between approximately 1835–1950 nm and 2220–2300 nm.

2. Levels of smoke taint compounds in grape juice/homogenate and wine

Differences in volatile phenol and glycoconjugate levels amongst the different smoke treatments could be seen at both one day after smoke exposure (Figure S1) and at harvest (Figure S2), except for syringol and 4-methylsyringol, in which there were no significant differences ($p > 0.05$) amongst the five smoke treatments at harvest. Significant differences ($p < 0.05$) in volatile phenol and glycoconjugate levels were also observed amongst the wine samples produced from grapes exposed to the different smoke treatments (Table S3).

3. Machine learning modelling

Statistical data for the five regression models are shown in Table 2. Model 1 had high overall correlation and determination coefficients ($R = 0.98$; $R^2 = 0.97$; Figure 5a.), with values of the validation correlation coefficient ($R = 0.97$) being close to the training correlation coefficients ($R = 0.95$). In addition to this, performance values for validation (MSE = 80.28) and testing (MSE = 83.14) were similar among them and higher than that of the training stage (MSE = 16.96), which further indicates no sign of under- or overfitting. Models 2 and 3 also had high correlation and determination coefficients ($R = 0.98$ for each; $R^2 = 0.97$ for Model 2 and 0.95 for Model 3; Figures 5b and 5c). Again, there were no signs of under- or overfitting as values for the validation and training coefficients were close, and the values for the performance of the training stage were lower than those of the validation and testing stages, with the latter values being similar. Models 4 and 5 displayed very high correlation and determination coefficients ($R = 0.99$ for each; $R^2 = 0.99$ for Model 4 and $R^2 = 0.97$ for Model 5; Figures 5d and 5e), with performance values of the training stage lower than the testing stage, which does not indicate any over- or underfitting.

DISCUSSION

The NIR wavelength range used to develop the five models was between 1596 and 2396 nm (Figures 2, 3, and 4), which includes several key overtones. The C-H stretch first overtone associated with aromatic compounds can be seen between 1680 and 1690 nm, while O-H stretching associated with glucose, cellulose, water, and alcohol can be observed at approximately 1930, 2090, 2270, and 2330 nm. Lastly, the region between 1900 and 1910 nm corresponds to C=O stretching associated with carboxylic acids and water (Boido *et al.*, 2013; Burns & Ciurczak, 2007; Gonzalez Viejo *et al.*, 2018). Thus, this region was found to be effective for assessing levels of smoke contamination. For the berry absorption spectra (Figure 2), the observed peaks (1790 and 1910 nm)

TABLE 2. Statistical results from the five developed artificial neural network regression models, which estimate the levels of volatile phenols and their glycoconjugates in grapes one day after smoking (Model 1), at harvest (Model 2), and wine (Models 3-5) showing the correlation coefficient (R), determination of coefficient (R^2), slope (b) and performance based on mean squared error (MSE) for each stage.

Stage	Samples	Observations	R	R^2	b	Performance (MSE)
Model 1						
Training	378	10584	0.99	0.99	1.00	16.96
Validation	81	2268	0.97	0.93	1.00	80.28
Testing	81	2268	0.95	0.91	0.99	83.14
Overall	540	15120	0.98	0.97	1.00	
Model 2						
Training	378	10584	0.99	0.99	0.96	332.26
Validation	81	2268	0.96	0.93	0.95	1533.75
Testing	81	2268	0.96	0.93	1.00	1716.87
Overall	540	15120	0.98	0.97	0.97	
Model 3						
Training	378	8694	0.99	0.97	1.00	79.07
Validation	81	1863	0.95	0.91	0.98	299.77
Testing	81	1863	0.94	0.89	0.93	290.90
Overall	540	12420	0.98	0.95	0.99	
Model 4						
Training	31	713	0.99	0.99	1	0.43
Testing	14	322	0.99	0.98	1	88.14
Overall	45	1035	0.99	0.99	1	
Model 5						
Training	31	713	0.99	0.99	1	8.03
Testing	14	322	0.96	0.92	0.92	258.70
Overall	45	1035	0.99	0.97	0.97	

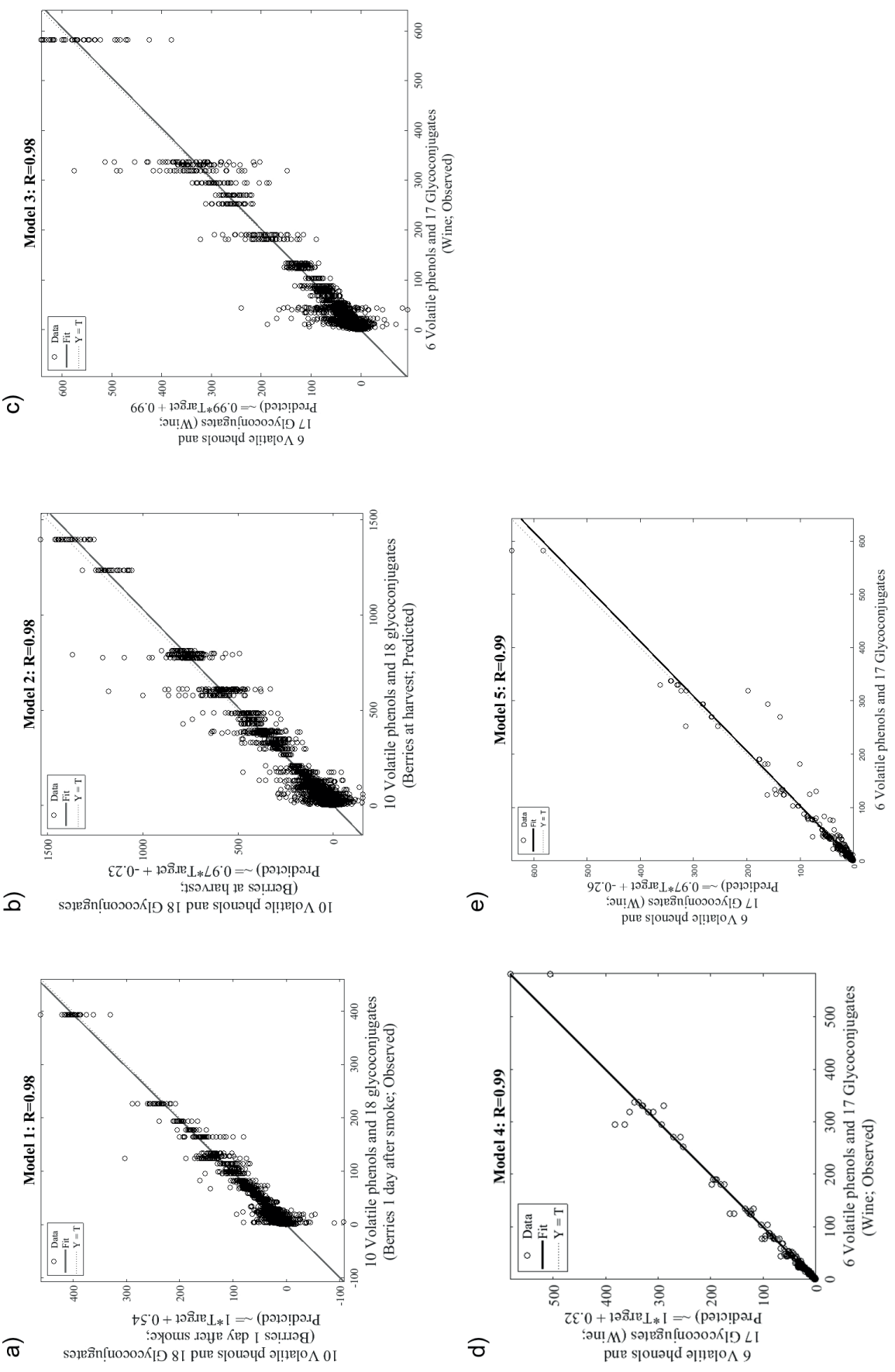

FIGURE 5. Overall correlation of the models to predict ten volatile phenols and 18 glycoconjugates (Table 1) of (a) Model 1: berries one day after smoking, (b) Model 2: berries at harvest, six volatile phenols, and 17 glycoconjugates of (c) Model 3: wine, (d) Model 4: wine and (e) Model 5: wine.

were in the NIR regions associated with C = O and O-H overtones in carboxylic acid or water, while for the must and wine absorption spectra (Figures 3 and 4) the observed peaks (1784, 1927, 2090 and 2270 nm) were in the regions associated with C-H and O-H stretching of starch and alcohol (Boido *et al*., 2013; Burns & Ciurczak, 2007).

Only the HS treatment berries one day after smoke exposure contained average free guaiacol levels above the aroma detection threshold (Table S1); m-cresol and 4-methylguaiacol average levels were below the aroma detection thresholds for all smoke treatments at this time period. Berries at harvest, however, did not contain any VP above the aroma detection thresholds for all smoke treatments (Figure S2), showing that most VP had formed glycoconjugates. In the final wine, both the HS and HSM treatments had average free guaiacol levels above or at the aroma detection threshold (Figure S3), with m-cresol and 4-methylguauaicol levels below their aroma detection thresholds. This highlights the importance of assessing VP and glycoconjugates levels in both grape berries and wine to obtain an idea of the level of smoke contamination and smoke taint in the final wine. Furthermore, in addition to a possible synergistic effect of smoke compounds to the overall smoky aroma, it has been reported that the structure/body of the juice or wine strongly influences the level and detection of smoke taint, with medium-bodied red wines having a lower guaiacol detection threshold (15-25 µg/L) than full-bodied style wines such as Shiraz (30-40 µg/L) (Simos, 2008). It is, therefore, important to consider the style of wine when assessing levels of smoke compounds. At present, the only available means that grape growers and winemakers have for determining the levels of volatile phenols and their glycoconjugates in grapes and wine are to send samples to a commercial laboratory, which is time-consuming and requires destructive sample preparation (Fudge *et al*., 2012b; Fudge *et al*., 2013; Kemps *et al*., 2010; Summerson *et al*., 2020). Models 1-3 may, therefore, offer grape growers and winemakers a rapid and non-destructive in-field measurement technique for assessing levels of smoke compounds in grapes and potential wine produced from them, with a high level of accuracy and precision. Growers would then be able to make timely decisions, such as avoiding heavily contaminated grapes for winemaking or sending a smaller selection of berry samples for further chemical analysis, depending on the results obtained. Furthermore, Models 4 and 5 offer winemakers near-real-time measurements of the levels of volatile phenols and their glycoconjugates in must and wine, which indicate the level of smoke taint in wine. Winemakers can then decide to apply smoke taint mitigation techniques, such as through treatment with activated carbon (Fudge *et al*., 2012a). Research by Fudge *et al*. (2012a) found that treating Cabernet-Sauvignon smoke-affected wines with activated carbon reduced all volatile phenols' concentration by 56 to 71 %. Specifically, levels of guaiacol were reduced from an average of 18 µg/L in the control wine to 8 µg/L following treatment with activated carbon, while 4-methylguaiacol was reduced from an average of 3 µg/L to 1 µg/L, and total cresols from an average of 7 µg/L to 2 µg/L. It can, therefore, be concluded that the treatment with activated carbon is effective at ameliorating smoke taint. Moreover, as this assessment method is non-destructive, repeated measurements can be made, which is particularly useful as bottle aging has been demonstrated to increase levels of both naturally occurring and smoke-derived VP (Ristic *et al*., 2017). Hence, repeated measurements over time could be performed to assess levels of both free volatile phenols and their glycoconjugates in wine and must. There is also the possibility of developing models that can assess wine non-destructively while in the bottle (Cozzolino *et al*., 2007). Research by Cozzolino *et al*. (2007) found great promise in the use of Vis-NIR spectroscopy to assess wine composition in the bottle, as well as to detect problems in the wine that may have occurred during or after bottling and before selling.

While the five models developed in this study accurately predicted levels of smoke compounds in berries and wine, further research is required to assess whether these models can be used for other grape and wine varieties. Differences in berry and wine composition may affect the accuracy in predicting volatile phenols' levels and their glycoconjugates in other varieties (Fudge *et al*., 2012b). Research by Fudge *et al*. (2012b) found that compositional differences due to grape variety had greater influence in classifying smoke-affected wine over wine exposed to low levels of smoke when using MIR spectroscopy as an assessment technique. However, in another study, Fuentes *et al*. (2019) developed a model that accurately predicts levels of guaiacol glycoconjugates in berries, as well as guaiacol and guaiacol glycoconjugates for seven different grapevine cultivars, using NIR berry measurements at between 700 and 1100 nm. Therefore, there is great potential for developing future models that can predict levels of smoke

compounds in multiple grape varieties and wines using the NIR regions between 1680 and 1690 nm, as was done in this study. Further research by Fuentes *et al*. (2020) also found the use of a low-cost E-nose to be effective in assessing levels of smoke compounds in wine, as well as the intensity of 12 wine descriptors, via consumer sensory testing. Therefore, NIR spectroscopy, together with an E-nose, could potentially be used to assess levels of smoke compounds non-destructively. Lastly, different winemaking techniques may affect the levels of smoke compounds in the wine; for example, both the type of yeast used and the duration of skin contact time during fermentation can influence levels of volatile phenols and their glycoconjugates in the final wine (Kennison *et al*., 2008; Ristic *et al*., 2011; Simos, 2008). Models 3 and 4 would thus need to be adjusted for different winemaking techniques, including the use of different yeasts, length of fermentation on skins, and the addition of malolactic fermentation. In this experiment, wines did not undergo malolactic fermentation, which reduces the pH of the wine and may, therefore, affect the hydrolysis of glycoconjugates into free VP and hence the level of smoke taint.

CONCLUSIONS

The use of NIR spectroscopy, coupled with machine learning, has demonstrated great potential as a non-destructive tool for measuring levels of smoke compounds in Cabernet-Sauvignon grapes and wine. The models developed could be used by grape growers and winemakers either in-field to take repeated assessments of grapes or in the winery to assess must and wine. This could assist them in making informed decisions about berry sampling and winemaking and application of smoke taint amelioration techniques, such as treatment with activated carbon to minimise levels of volatile phenols in wine. Further research is required to assess whether the models developed in this study could be used for other grape and wine varieties and winemaking techniques, such as using different yeast strains and duration of skin contact time during fermentation.

Acknowledgements: This research was supported by the Australian Government Research Training Program Scholarship, as well as the Digital Viticulture program funded by the University of Melbourne's Networked Society Institute, Australia. The authors gratefully acknowledge the Digital Agriculture, Food, and Wine Group. They also gratefully acknowledge Kerry Wilkinson and Colleen Szeto for the opportunity to collaborate in the field trials, providing data on levels of volatile phenols and their glycoconjugates and supplying wine samples. Colleen Szeto was supported by the Australian Research Council Training Centre for Innovative Wine Production (www.arcwinecentre.org.au) funded as part of the ARC's Industrial Transformation Research Program (Project No. ICI70100008), with support from Wine Australia and industry partners.

REFERENCES

Allen, D., Bui, A. D., Cain, N., Rose, G., & Downey, M. (2013). Analysis of free and bound phenolics in wine and grapes by GC-MS after automated SPE. *Analytical and Bioanalytical Chemistry, 405*(30), 9869-9877. https://doi.org/10.1007/s00216-013-7405-0

Bell, T., Stephens, S., & Moritz, M. (2013). Short-term physiological effects of smoke on grapevine leaves. *International Journal of Wildland Fire, 22*(7), 933-946. https://doi.org/10.1071/WF12140

Boido, E., Fariña, L., Carrau, F., Dellacassa, E., & Cozzolino, D. (2013). Characterization of glycosylated aroma compounds in Tannat grapes and feasibility of the near infrared spectroscopy application for their prediction. *Food Analytical Methods, 6*(1), 100-111. https://doi.org/10.1007/s12161-012-9423-5

Burns, D., & Ciurczak, E. (2007). *Handbook of Near-infrared Analysis*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781420007374

Caravia, L., Pagay, V., Collins, C., & Tyerman, S. D. (2017). Application of sprinkler cooling within the bunch zone during ripening of Cabernet-Sauvignon berries to reduce the impact of high temperature. *Australian Journal of Grape and Wine Research, 23*(1), 48-57. https://doi.org/10.1111/ajgw.12255

Cozzolino, D., Cynkar, W., Shah, N., & Smith, P. (2012). Varietal differentiation of grape juice based on the analysis of near-and mid-infrared spectral data. *Food Analytical Methods, 5*(3), 381-387. https://doi.org/10.1007/s12161-011-9249-6

Cozzolino, D., Kwiatkowski, M. J., Waters, E. J., & Gishen, M. (2007). A feasibility study on the use of visible and short wavelengths in the near-infrared region for the non-destructive measurement of wine composition. *Analytical and Bioanalytical Chemistry, 387*(6), 2289-2295. https://doi.org/10.1007/s00216-006-1031-z

CSIRO and Australian Government Bureau of Meteorology. (2018). *State of the Climate 2018* Retrieved from

De Vries, C., Mokwena, L., Buica, A., & McKay, M. (2016). Determination of volatile phenol in Cabernet-Sauvignon wines, made from smoke-affected grapes, by using HS-SPME GC-MS. *South African Journal of Enology and Viticulture, 37*(1), 15-21. https://doi.org/10.21548/37-1-754

Department of Primary Industries. (2009). Impacts of smoke on grapes and wine in Victoria. Retrieved from http://wine.wsu.edu/research-extension/files/2012/10/DPI-fact-sheet_Impacts-of-smoke-on-grapes-and-wine-in-Victoria_final.pdf

Diamantopoulou, M. J., & Milios, E. (2010). Modelling total volume of dominant pine trees in reforestations via multivariate analysis and artificial neural network models. *Biosystems Engineering, 105*(3), 306-315. https://doi.org/10.1016/j.biosystemseng.2009.11.010

Dieulot, J. Y., & Skurtys, O. (2013). Classification, modeling and prediction of the mechanical behavior of starch-based films. *Journal of Food Engineering, 119*(2), 188-195. https://doi.org/10.1016/j.jfoodeng.2013.05.028

Dungey, K. A., Hayasaka, Y., & Wilkinson, K. L. (2011). Quantitative analysis of glycoconjugate precursors of guaiacol in smoke-affected grapes using liquid chromatography–tandem mass spectrometry based stable isotope dilution analysis. *Food Chemistry, 126*(2), 801-806. https://doi.org/10.1016/j.foodchem.2010.11.094

Favell, J. W., Noestheden, M., Lyons, S.-M., & Zandberg, W. F. (2019). Development and evaluation of a vineyard-based strategy to mitigate smoke-taint in wine grapes. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.9b05859

Fudge, A., Ristic, R., Wollan, D., & Wilkinson, K. (2011). Amelioration of smoke taint in wine by reverse osmosis and solid phase adsorption. *Australian Journal of Grape and Wine Research, 17*(2), S41-S48. https://doi.org/10.1111/j.1755-0238.2011.00148.x

Fudge, A., Schiettecatte, M., Ristic, R., Hayasaka, Y., & Wilkinson, K. (2012a). Amelioration of smoke taint in wine by treatment with commercial fining agents. *Australian Journal of Grape and Wine Research, 18*(3), 302-307. https://doi.org/10.1111/j.1755-0238.2012.00200.x

Fudge, A., Wilkinson, K., Ristic, R., & Cozzolino, D. (2012b). Classification of smoke tainted wines using mid-infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry, 60*(1), 52-59. https://doi.org/10.1021/jf203849h

Fudge, A., Wilkinson, K., Ristic, R., & Cozzolino, D. (2013). Synchronous two-dimensional MIR correlation spectroscopy (2D-COS) as a novel method for screening smoke tainted wine. *Food Chemistry, 139*(1–4), 115-119. https://doi.org/10.1016/j.foodchem.2013.01.090

Fuentes, S., Summerson, V., Gonzalez Viejo, C., Tongson, E., Lipovetzky, N., Wilkinson, K. L., ... Unnithan, R. R. (2020). Assessment of Smoke Contamination in Grapevine Berries and Taint in Wines Due to Bushfires Using a Low-Cost E-Nose and an Artificial Intelligence Approach. *Sensors, 20*(18), 5108. https://doi.org/10.3390/s20185108

Fuentes, S., & Tongson, E. (2017). Advances in smoke contamination detection systems for grapevine canopies and berries. *Wine & Viticulture Journal, 32*(3), 36.

Fuentes, S., Tongson, E. J., De Bei, R., Gonzalez Viejo, C., Ristic, R., Tyerman, S., & Wilkinson, K. (2019). Non-Invasive Tools to Detect Smoke Contamination in Grapevine Canopies, Berries and Wine: A Remote Sensing and Machine Learning Modeling Approach. *Sensors, 19*(15), 3335. https://doi.org/10.3390/s19153335

Gonzalez Viejo, C., Fuentes, S., Torrico, D., Howell, K., & Dunshea, F. R. (2018). Assessment of beer quality based on foamability and chemical composition using computer vision algorithms, near infrared spectroscopy and machine learning algorithms. *Journal of the Science of Food and Agriculture, 98*(2), 618-627. https://doi.org/10.1002/jsfa.8506

Hall, A. (2018). Remote Sensing Applications for Viticultural Terroir Analysis. *Elements, 14*(3), 185-190. https://doi:10.2138/gselements.14.3.185

Härtl, K., & Schwab, W. (2018). Smoke Taint in Wine-How smoke-derived volatiles accumulate in grapevines. *Wines & Vines*(99/02), 62-64.

Hayasaka, Y., Baldock, G., Parker, M., Pardon, K., Black, C., Herderich, M., & Jeffery, D. (2010a). Glycosylation of smoke-derived volatile phenols in grapes as a consequence of grapevine exposure to bushfire smoke. *Journal of Agricultural and Food Chemistry, 58*(20), 10989-10998. https://doi.org/10.1021/jf103045t

Hayasaka, Y., Dungey, K., Baldock, G., Kennison, K., & Wilkinson, K. (2010b). Identification of a β-d-glucopyranoside precursor to guaiacol in grape juice following grapevine exposure to smoke. *Analytica Chimica Acta, 660*(1–2), 143-148. http://dx.doi.org/10.1016/j.aca.2009.10.039

Hayasaka, Y., Parker, M., Baldock, G. A., Pardon, K. H., Black, C. A., Jeffery, D. W., & Herderich, M. J. (2013). Assessing the impact of smoke exposure in grapes: development and validation of a HPLC-MS/MS method for the quantitative analysis of smoke-derived phenolic glycosides in grapes and wine. *Journal of Agricultural and Food Chemistry, 61*(1), 25-33. https://doi.org/10.1021/jf305025j

Hughes, L., & Alexander, D. (2017). Climate Change and the Victoria Bushfire Threat: Update 2017. *Climate Council Report.* Retrieved from https://www.climatecouncil.org.au/uploads/98c26db6af45080a32377f3ef4800102.pdf

Kelly, D., Zerihun, A., Singh, D., Vitzthum von Eckstaedt, C., Gibberd, M., Grice, K., & Downey, M. (2012). Exposure of grapes to smoke of vegetation with varying lignin composition and accretion of lignin derived putative smoke taint compounds in wine. *Food Chemistry, 135*(2), 787-798. https://doi.org/10.1016/j.foodchem.2012.05.036

Kemps, B., Leon, L., Best, S., De Baerdemaeker, J., & De Ketelaere, B. (2010). Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. *Biosystems Engineering, 105*(4), 507-513. https://doi.org/10.1016/j.biosystemseng.2010.02.002

Kennison, K., Gibberd, M., Pollnitz, A., & Wilkinson, K. (2008). Smoke-Derived Taint in Wine: The Release of Smoke-Derived Volatile Phenols during Fermentation of Merlot Juice following Grapevine Exposure to Smoke. *Journal of Agricultural and Food Chemistry, 56*(16), 7379-7383. https://doi:10.1021/jf800927e

Kennison, K., Wilkinson, K., Pollnitz, A., Williams, H., & Gibberd, M. (2009). Effect of timing and duration of grapevine exposure to smoke on the composition and sensory properties of wine. *Australian Journal of Grape and Wine Research, 15*(3), 228-237. https://doi.org/10.1111/j.1755-0238.2009.00056.x

Kennison, K., Wilkinson, K., Pollnitz, A., Williams, H., & Gibberd, M. (2011). Effect of smoke application to field-grown Merlot grapevines at key phenological growth stages on wine sensory and chemical properties. *Australian Journal of Grape and Wine Research, 17*(2), 5-12. https://doi.org/10.1111/j.1755-0238.2011.00137.x

Kennison, K., Wilkinson, K., Williams, H., Smith, J., & Gibberd, M. (2007). Smoke-derived taint in wine: Effect of postharvest smoke exposure of grapes on the chemical composition and sensory characteristics of wine. *Journal of Agricultural and Food Chemistry, 55*(26), 10897-10901. https://doi.org/10.1021/jf072509k

Martelo-Vidal, M., & Vázquez, M. (2014). Determination of polyphenolic compounds of red wines by UV–VIS–NIR spectroscopy and chemometrics tools. *Food Chemistry, 158*, 28-34. https://doi.org/10.1016/j.foodchem.2014.02.080

Martelo-Vidal, M. J., & Vázquez, M. (2015). Application of artificial neural networks coupled to UV–VIS–NIR spectroscopy for the rapid quantification of wine compounds in aqueous mixtures. *CyTA-Journal of Food, 13*(1), 32-39. https://doi.org/10.1016/j.foodchem.2014.02.080

Noestheden, M., Dennis, E. G., & Zandberg, W. F. (2018a). Quantitating Volatile Phenols in Cabernet franc Berries and Wine after On-Vine Exposure to Smoke from a Simulated Forest Fire. *Journal of Agricultural and Food Chemistry, 66*(3), 695-703. https://doi:10.1021/acs.jafc.7b04946

Noestheden, M., Noyovitz, B., Riordan-Short, S., Dennis, E. G., & Zandberg, W. F. (2018b). Smoke from simulated forest fire alters secondary metabolites in Vitis vinifera L. berries and wine. *Planta*. https://doi:10.1007/s00425-018-2994-7

Noestheden, M., Thiessen, K., Dennis, E. G., Tiet, B., & Zandberg, W. F. (2017). Quantitating Organoleptic Volatile Phenols in Smoke-Exposed Vitis vinifera Berries. *Journal of Agricultural and Food Chemistry, 65*(38), 8418-8425. https://doi:10.1021/acs.jafc.7b03225

Parker, M., Baldock, G., Hayasaka, Y., Mayr, C., Williamson, P., Francis, I. L., . . . Johnson, D. (2013). Seeing through smoke. *Wine Vitic. J, 28*, 42-46.

Pirie, A., Singh, B., & Islam, K. (2005). Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Soil Research, 43*(6), 713-721. https://doi.org/10.1071/SR04182

Pollnitz, A. P., Pardon, K. H., Sykes, M., & Sefton, M. A. (2004). The effects of sample preparation and gas chromatograph injection techniques on the accuracy of measuring guaiacol, 4-methylguaiacol and other volatile oak compounds in oak extracts by stable isotope dilution analyses. *Journal of Agricultural and Food Chemistry, 52*(11), 3244-3252. https://doi.org/10.1021/jf035380x

Ristic, R., Fudge, A., Pinchbeck, K., De Bei, R., Fuentes, S., Hayasaka, Y., . . . Wilkinson, K. (2016). Impact of grapevine exposure to smoke on vine physiology and the composition and sensory properties of wine. *Theoretical and Experimental Plant Physiology, 28*(1), 67-83. https://doi:10.1007/s40626-016-0054-x

Ristic, R., Osidacz, P., Pinchbeck, K., Hayasaka, Y., Fudge, A., & Wilkinson, K. (2011). The effect of winemaking techniques on the intensity of smoke taint in wine. *Australian Journal of Grape and Wine Research, 17*(2), S29-S40. https://doi.org/10.1111/j.1755-0238.2011.00146.x

Ristic, R., van der Hulst, L., Capone, D., & Wilkinson, K. (2017). Impact of Bottle Aging on Smoke-Tainted Wines from Different Grape Cultivars. *Journal of Agricultural and Food Chemistry, 65*(20), 4146-4152. https://doi:10.1021/acs.jafc.7b01233

Simos, C. (2008). The implications of smoke taint and management practices. *Australian Viticulture Jan/Feb*, 77-80.

Singh, D., Chong, H., Pitt, K., Cleary, M., Dokoozlian, N., & Downey, M. (2011). Guaiacol and 4-methylguaiacol accumulate in wines made from smoke-affected fruit because of hydrolysis of their conjugates. *Australian Journal of Grape and Wine Research, 17*(2), S13-S21. https://doi.org/10.1111/j.1755-0238.2011.00128.x

Summerson, V., Gonzalez Viejo, C., Szeto, C., Wilkinson, K. L., Torrico, D. D., Pang, A., ... Fuentes, S. (2020). Classification of SmokeContaminated Cabernet-Sauvignon Berries and Leaves Based on Chemical Fingerprinting and Machine Learning Algorithms. *Sensors, 20*(18), 5099. https://doi.org/10.3390/s20185099

Szeto, C., Ristic, R., Capone, D., Puglisi, C., Pagay, V., Culbert, J., . . . Wilkinson, K. (2020). Uptake and Glycosylation of Smoke-Derived Volatile Phenols by

Cabernet-Sauvignon Grapes and Their Subsequent Fate during Winemaking. *Molecules, 25*(16), 3720. https://doi.org/10.3390/molecules25163720

Teixeira dos Santos, c., Lopo, M., Ricardo, N., & Lopes, J. (2013). A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry. *Applied Spectroscopy, 67*(11), 1215-1233. https://doi:10.1366/13-07228

van der Hulst, L., Munguia, P., Culbert, J. A., Ford, C. M., Burton, R. A., & Wilkinson, K. L. (2019). Accumulation of volatile phenol glycoconjugates in grapes following grapevine exposure to smoke and potential mitigation of smoke taint by foliar application of kaolin. *Planta, 249*(3), 941-952. https://doi.org/10.1007/s00425-018-03079-x

Yu, J., Wang, H., Zhan, J., & Huang, W. (2018). Review of recent UV–Vis and infrared spectroscopy researches on wine detection and discrimination. *Applied Spectroscopy Reviews, 53*(1), 65-86. https://doi:10.1080/05704928.2017.1352511

